STOCHASTIC SYSTEMS AND STATE ESTIMATION

Terrence P. McGarty
Massachusetts Institute of Technology
Cambridge, Massachusetts

A Wiley-Interscience Publication
JOHN WILEY & SONS
New York · London · Sydney · Toronto

Library of Congress Cataloging in Publication Data:

McGarty, Terrence P 1943–

Stochastic systems and state estimation.

Bibliography: p.

1. System analysis. 2. Stochastic processes.
3. Estimation theory. I. Title.

QA402.M28 519.2 73–18294

ISBN 0–471–58400–2

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Comp: Insert Dedication

PREFACE

To anyone familiar with measurements it should be quite obvious that the world is filled with uncertainty. Through ingenuity and insight scientists and engineers have over the past several centuries found ways to combat these uncertainties. Methods of smoothing and interpolation have evolved significantly. Statistical techniques play indispensable roles in the fields of meteorology, medicine, biology, and physiology, as well as all the engineering disciplines. This book is directed to those who have struggled, or will eventually struggle, with the problems of estimation found so frequently in these fields. The techniques described in this book assume a certain maturity in the science to which they are to be applied. Namely, they require that one have a model of the process whose values are to be estimated. In many instances this is not the case, so the techniques developed will be of no use. Yet, in those situations where a model exists, it has been found that these techniques significantly improve the estimation results obtained by any other method.

The approach of this book has been more pedagogical than that of Bucy and Joseph and more theoretical than that of Jaswinski. We have included much of the peripheral theory but in so doing have increased the length proportionately. The purpose of this approach was to broaden the size of the potential audience and to make the material more widely available to those whose theoretical backgrounds may not encompass such works as Doob and Halmos but who would like to bring themselves up to that level in this area. Chapter 3, 5, and 6 have been taught in special courses by the author at Massachusetts Institute of Technology. They generally encompass one semester's worth of material. The material in Chapter 2 is considered as a review, while that in Chapter 4 provides the basis for a deeper analysis of the theory of estimation.

/ the

The audience for this course has usually consisted of physicists, mathematicians (pure), mechanical and electrical engineers, and meteorologists. Thus, the greatest problem is that of showing relevancy. With such an audience the profusion of examples and extensions can provide a book in

itself, so that self-motivation in the text has been assumed. It is hoped this will not be too severe a drawback for the readers. To help in this matter, we have provided an extensive number of problems. They are a mix of practice type problems and those that actually extend the theory. A solution manual for these problems will be available shortly.

The classes to whom this material has been taught have had in general only a probability background equivalent to Davenport and Root's text and systems theory equivalent to that of Brockett's or Ogata's text. Thus understanding of convergence in the mean and in probability is to be understood at that level. An analysis course commensurate with Rudin [1] has been found necessary to provide adequate mathematical sophistication. An in-depth knowledge of measure theory, as in Halmos [2], has been available to many students.

The approach to the material has been made at an abstract level, but at the same time, an attempt has been made to make it presentable to those with the above background. To do so, we may at times seem to emphasize theory too much and, at other times, seem to develop a proof too briefly. To those who object to this, I merely say that what I have presented is my "view from the bridge."

Many people helped to develop this book into its present form. The initial encouragement of Dick Harlow and Carl Gray of the M.I.T. Charles S. Draper Laboratory led to the first draft of the document. The support of this manuscript and research in its early stages by Professor Draper's Laboratory is gratefully acknowledged. The comments of Professor Sanjoy Mitter of M.I.T. and the opportunity to teach this material with him led to many useful changes in presentation. Particular thanks goes to both Professor Tom Kailath of Stanford, who read and reread the entire manuscript and whose suggestions, criticisms, and mastery of the field proved to be invaluable, and to Professor John Clark of the University of Colorado, who provided continued comments and encouragement. Discussion with Professor Richard Dudley of M.I.T. led to the presentation of Chapter 4, and his assistance is gratefully acknowledged. Considerable assistance was also provided by the comments and suggestions made by Drs. Paul Frost, Al Gilman, Ken Senne, and John Morissey. And finally, but most important, I would like to express my deepest appreciation to my typist, Robin Schneider, who did the first draft in record time; to my wife Winnie, who has made endless corrections in endless further drafts; and to my children Terry and Krissy, who provided constant support and all their love.

Terrence McGarty

Acton, Massachusetts
June 1973

CONTENTS

CHAPTER 1

INTRODUCTION

The recovery of information from measurements corrupted by uncertainty has long been a struggle endured by many an investigator. Limited either by his choice of measurements or the nature of the variables he is interested in, or both, many techniques have evolved to combat these inefficiencies and obtain the best possible estimate of the desired variables. The techniques employed may be merely a simple method of data-smoothing or regression techniques, least-squares fit, polynomial approximations, or just plain educated guesswork. No matter what technique was employed the investigator always sought better and improved methods of data estimation.

With the advent of high-speed data-processing more advanced techniques of data analysis have been developed. One of these techniques is the subject of this book, namely, the estimation of random processes that are Markov and are observed through highly nonlinear and complex systems. The type of estimates are minimum mean square error (MMSE) estimates and are related closely to the least-squares approach taken by Gauss in his estimation of planetary trajectories. Gauss's theory was later extended by Kolmogorov and Wiener for the filtering of stationary random sequences and processes, respectively. The development of the linear filtering theory using the state-space approach was carried out by Kalman, and its extensions to arbitrary Markov processes with measurements of varying types by Stratonovich, Kushner, and Snyder. The theory presented in this book reflects the contributions made by these investigators in their attempts to understand and extend the knowledge of estimating stochastic systems.

To understand the extent to which the theory of state estimation of Markov processes can be applied, it is necessary to develop an adequate facility with both modern control theory, vis-à-vis state-space techniques and the ideas of probability theory and stochastic processes. It is the union of these two areas of knowledge that has made the contribution of nonlinear estimation so all-encompassing. The understanding of the state-space techniques allows for the development of a very robust model development, which

allows the techniques of estimation to be applied to a wide class of problems. The understanding gained in the use of both probability theory and stochastic-process theory permits the elucidation of eloquent results. The combination of the two yields a highly viable and worthwhile theory as well as an indispensable technique.

In this chapter we first present a general outline of the class of problems that are to be investigated in the book. The purpose of this outline is twofold. First, it is to show the reader that the theory has many worthwhile practical applications. This is performed by presenting two specific examples in which it is extensively used. The second purpose is to point out those areas of analysis that require some in-depth treatment in order for the theory to stand on its merits of mathematical consistency. The second section of this chapter presents a chapter-by-chapter preview of what will be covered. The purpose of this is to delineate those areas of particular interest and to show how they relate to the whole book.

1.1 THE PROBLEM

A system is, in a rather general sense, some organized dynamic object that can be influenced externally and whose behavior can in some fashion be monitored. More precise definitions are available (see Kalman, Falb, and Arbib) but for a preliminary presentation this should suffice. More simply we can consider a system as a mathematical embodiment of some natural phenomenon. For example, the human body is a system, albeit a very complex one, and the clock pendulum is also a system, one of preise mathematical description. To each system we ascribe quantities called states, and these quantities are used to describe the evolution of the system as some set of independent variables (usually time) change.

The state of a system is represented by some quantity called $\mathbf{x}(t)$, where t is the independent variable (s). For our purposes t represents the single variable time, although such things as position coordinates are also possible. The systems in which we shall find most interest are those in which the states form a finite vector so that $\mathbf{x}(t)$ is represented by an $n \times 1$ vector, namely,

$$\mathbf{x}(t)=\begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix} \tag{1.1}$$

These are called *finite dimensional systems* and are quite common. The dynamics of each system is assumed to be governed by a differential equation of the form

$$\frac{d\mathbf{x}(t)}{dt}=\mathbf{f}(\mathbf{x}(t),\ t)+\frac{d\mathbf{n}(t)}{dt} \tag{1.2}$$

This equation is called the *state equation*. The quantity $d\mathbf{n}(t)/dt$ represents any all external disturbances. These disturbances may be either deterministic or stochastic or some combination of both. The basic fact is that the complete temporal behavior of the system, and thus its complete behavior, is given by the state equation. Furthermore, it is the behavior of this system that one is usually seeking to ascertain. For example, the physiological or pathological state of the renal system may be described adequately by such a system of equations, with the input being the ingestion of a glucose solution driving the state of the kidney in some manner.

All the systems that we are interested in are observed by means of some measurement system. Furthermore, all the measurements that we obtain are perturbed by some form of measurement noise, which in turn introduces uncertainty into our knowledge of the system itself. As in our example of the renal system, a measurement may be that of the sugar content of the blood and the amount of uric acid. But the measurement techniques are not perfect, and thus, errors are made. For example, a linearly perturbed measurement may be given by the $m \times 1$ vector $\mathbf{z}(t)$, where

$$\mathbf{z}(t) = \mathbf{h}(\mathbf{x}(t), t) + \frac{d\mathbf{w}(t)}{dt} \tag{1.3}$$

This equation is called the *measurement equation*. The quantity $d\mathbf{w}(t)/dt$ is a noise disturbance and $\mathbf{h}(\mathbf{x}(t), t)$ represents the $m \times 1$ vector transformation from the state to the measurement. Another example of a measurement may be the particle count rate of some nuclear tracer. In this count the rate may be the function, which depends upon the state. Namely, if $\mathbf{N}(t)$ is the number of counts observed from, say, a Poisson process from $(0, t)$ and the average number in an interval dt is $\lambda(\mathbf{x}(t), t)dt$, then we may want to determine $\mathbf{x}(t)$ from knowledge of $\mathbf{N}(t)$, the measurement transformation.

The problem of state estimation, then, is that of taking the noisy measurements, where the noise has been suitably defined, and processing them in

Figure 1.1 General problem.

some fashion so that we can obtain a good guess of the state. This guess or estimate is termed $\hat{\mathbf{x}}(t)$. This general scheme is depicted in Figure 1.1. The specific structure of the estimator is the object of this book. It will depend upon how we define the driving forces on the system, the disturbances affecting the measurements, and in what way our guess is considered to be a good guess. Throughout the analysis we inherently assume that there is a clearly defined structure to both the state and the measurement system and that this structure is known.

To determine a specific choice for the estimator it is first necessary to describe what we mean by a good estimate. This must be done in a quantitative fashion. To do so, consider the error in the estimate of the ith state at a given instant of time t. Let this be denoted by $\tilde{x}_i(t)$, which is

$$\tilde{x}_i(t) = x_i(t) - \hat{x}_i(t) \tag{1.4}$$

Figure 1.2(a) shows a possible sample path for a given $x_i(t)$ and a given estimator $\hat{x}_i(t)$. In Figure 1.2(b) the error is plotted as a function of time.

Figure 1.2 Comparison of state, estimate, and error time behavior. (a) State and estimate; (b) error.

For different estimator structures, different errors, $\tilde{x}_i(t)$, will be obtained. To obtain an optimum estimator we would intuitively like to have one that minimizes the errors in the system. Namely, we would want one to minimize $\tilde{x}_i(t)$. However, this may be equivalent to minimizing some appropriate function of $\tilde{x}_i(t)$ also, for example $g(\tilde{x}_i(t))$. Also, since $\tilde{x}_i(t)$ is a stochastic process and is merely one sample path of that process, we would like to minimize this error over the entire ensemble. Namely, we would like to minimize the function

$$E[g(\tilde{x}_i(t))] \quad \forall t \tag{1.5}$$

where $x_i(t)$ is a stochastic process and $\hat{x}_i(t)$ is also a stochastic process that depends on the measurements $\mathbf{z}(s)$ for all $t_0 \leqq s \leqq t$. That is $\hat{x}_i(t)$ depends on the entire record of past measurements. t_0 is the initial time that the measurements were made.

The choice of the weighting function $g(\ \ )$ is rather arbitrary, but there are certain analytical properties that it must satisfy. For example, there must exist an $\hat{x}_i(t)$ that minimizes the function, and furthermore, this estimate should be unique. A form of $g(\ \ )$ that insures this is the minimum mean square error (MMSE) estimate form, specifically,

$$E[g(\tilde{x}_i(t))] = E[(x_i(t) - \hat{x}_i(t))^2] \tag{1.6}$$

Given this as a weighting function and optimization criterion, it can be shown that there exists a unique estimate $\hat{x}_i(t)$ that minimizes this expression for each t and depends on the past measurements $\mathcal{O}_{t_0,t}$, where $\mathcal{O}_{t_0,t}$ is the observation record $\{\mathbf{z}(s): t_0 \leqq s \leqq t\}$. This estimate is the conditional mean or MMSE estimate:

$$\hat{x}_i(t) = E[x_i(t) | \mathcal{O}_{t_0,t}] = \int u_i p_i(\mathbf{u},t | \mathcal{O}_{t_0,t}) d\mathbf{u} \tag{1.7}$$

where $p_{\mathbf{x}}(\mathbf{u},t|\mathcal{O}_{t_0,t})$ is the conditional probability density of $x_i(t)$, given the observations $\mathcal{O}_{t_0,t}$. Thus, it is equivalent to know either the conditional mean directly or the conditional probability density.

It is the purpose of this book to develop the theory necessary to understand the model, the basis of MMSE estimation, and the evaluation of conditional expectations and probability densities. With this knowledge we can then develop the estimator structures for the Markov models we have discussed. As for the application of these techniques, we shall briefly outline two specific ones. The first represents an example of a class of problems where there is a well-defined deterministic system driven by random disturbances and where the measurement system is also clearly defined. In both the measurement and the system there are nonlinearities. This is representative of many problems with a fundamental physical embodiment. The second example represents an application to the communication field where the state system is not an

actual physical reality but a model that represents the stochastic nature of a singnal source.

For the first example, consider the motion of a point mass (m_1) about another mass (m_2). Let $\mathbf{r}$ be the coordinate vector (x, y, z). Then the law of gravitation and Newton's law yields (see Battin, p. 9)

$$m_1 \frac{d^2\mathbf{r}}{dt^2} = -\frac{Gm_1m_2\,\mathbf{r}}{|\mathbf{r}|^3}$$

for the dynamic behavior of the position vector $\mathbf{r}$. If, however, the mass m_2 is a distributed mass and if m_1 is a satellite and m_2 the earth and there are other planets, then they are acting as forces that we have not accounted for. These anomalies represent the unknown and, in some very real sense, random driving forces. Let these forces be represented by $d\mathbf{n}(t)/dt$. Also by defining $x_1 = x$, $x_2 = y$, $x_3 = z$, $x_4 = \dot{x}_1$, $x_5 = \dot{x}_2$, and $x_6 = \dot{x}_3$ we can write the equations of motion as

$$\frac{dx_i(t)}{dt} = x_{i+3}(t) \qquad (i = 1, \cdots, 3)$$

$$\frac{dx_{i+3}(t)}{dt} = f_i(\mathbf{x}(t), t) + \frac{dn_i(t)}{dt} \qquad (i = 1, \cdots, 3)$$

where $f_i(\mathbf{x}(t), t)$ is given by $Gm_2\, x_i/|\mathbf{r}|^3$. Thus we let the state $\mathbf{x}(t)$ be a 6×1 vector. Now if we measure the position of the satellite with a radar that gives angle and range, then we can develop a set of equations for the measurement system. Namely, if we let z_1 be the range, then

$$z_1(t) = [x_1^2(t) + x_2^2(t) + x_3^2(t)]^{1/2} + w_1(t)$$

where $w_1(t)$ represents measurement noise. If we let z_2 represent a longitude angle (θ; see Figure 1.3), then

$$z_2(t) = \frac{x_3(t)}{[x_1^2(t) + x_2^2(t) + x_3^2(t)]^{1/2}} + w_2(t)$$

Likewise $z_3(t)$ is a latitude angle and is given by

$$z_3(t) = \frac{x_1(t)}{[x_1^2(t) + x_2^2(t)]^{1/2}} + w_3(t)$$

Thus, using the theory developed, we should be able to obtain an estimate of the state of the satellite, $\hat{\mathbf{x}}(t)$, given $\mathbf{z}(t)$. In this example the system is the set of equations representing the dynamical behavior of the position of the satellite. The measurement system is given by the three radar signals. The estimator structure then is determined, using this as a model and the MMSE criterion as a standard of performance. Historically, this is one of the first and most important uses of the linear Kalman-Bucy filtering equations, which are a special case of the nonlinear estimators to be developed.

In the previous example the state equation is given based upon some

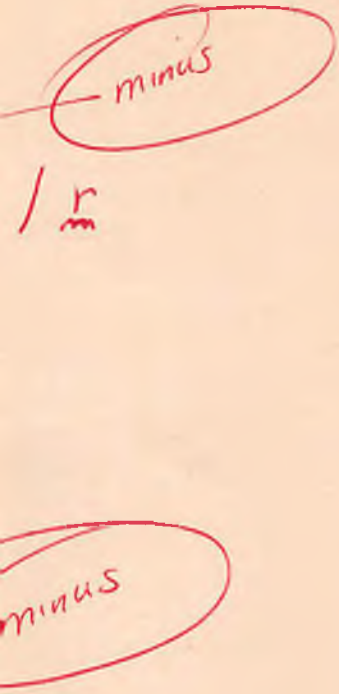

Figure 1.3 Example of two-body motion.

well-known physical phenomenon. An alternative approach that is useful for modeling communications systems is to let the state equation be such that it has given second-order statistics. Specifically, by choosing a linear time-invariant system of the form

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{A}\mathbf{x}(t) + \frac{d\mathbf{n}(t)}{dt}$$

where $\mathbf{A}$ is an $n \times n$ matrix and $d\mathbf{n}(t)/dt$ is a white noise process, $x_i(t)$ can have a given spectrum by choosing $\mathbf{A}$ accordingly. Then we can pose the classical communication problem of how we estimate a signal of known spectral characteristics that is sent over a possibly nonlinear and noisy channel. A specific example is frequency modulation where the received signal or the measurement is of the form

$$z(t) = \cos\left[\omega_0 t + \int h(t-\tau)x_i(\tau)d\tau\right] + w(t)$$

The function $h(t)$ is a possible preemphasis filter. This system is shown schematically in Figure 1.4. The estimator structure then takes $z(s)$ and uses it to generate $x_i(t)$, the estimate of the signal.

In both of these examples we noted the presence of both the state model

Figure 1.4 Example of F-M transmission.

and a measurement model. Both were perturbed by noise of some form. By properly defining this noise, we shall see that an estimation structure will evolve that will satisfy the desired MMSE constraint.

Thus, the object of this book is fourfold:

1. To develop a clear and consistent understanding of the nature of the stochastic processes that define both the state and measurement systems, specifically, to model the disturbances so that they are mathematically consistent and at the same time yield systems amenable to further analysis.
2. To study the nature and structure of the optimization problem and to observe what types of cost criteria or optimization structures yield the best results. This will entail a deep understanding of the interrelationships between the abstract probability spaces and the optimum solutions.
3. To develop models that depict analytically the state of the stochastic estimation problem. This requires the development of a system of equations that will allow us to evaluate conditional probability densities, $p_{\mathbf{x}}(\mathbf{u},t|\mathcal{Y}_{t_0 t})$.
4. To use the results of the propagation analysis to obtain equations for the optimum estimate and the performance of that estimate and, furthermore, to look at special cases, particularly linear systems, to see what simplifications can be obtained.

In the next section we shall briefly preview each chapter highlighting the important topics.

1.2. OUTLINE OF THE BOOK

The book can be divided into two parts: definition, and solution and implementation. Part I represents the definition phase. It presents us with those tools necessary to clearly define the model, noise and performance. Part II is divided into two sections; solution and implementation. The first section, that of solution, provides us with answers to the problem of estimation. In general, these answers are too complex to state in a closed form, so that little of a specific nature can be said. Thus the second section, implementation, provides us with the tools to solve the estimation problem. In general, these tools are based on simplifying assumptions, which when applied, yield tractable computational algorithms. What we shall do now is to review the six following chapters, which deal with the development and discuss the salient issues.

Chapter 2 provides a general deterministic context for the discussion of dynamic systems. The state-space approach is used for several purposes. First, it is a time-domain approach and the incorporation of time-varying system dynamics or nonlinearities is quite simple. The more classical method

of using transfer functions, although computationally simpler at times, was only useful for linear time-invariant systems. Second, simulations on digital computers become possible with the state-space representation. A third reason is that many results in both optimal control (Athans and Falb) and estimation theory are formulated in state-space terms.

We first present the concepts of a dynamical system and the state of a system. The definition presented is quite formal, but via several examples the concrete nature of a dynamic system is presented.

In this chapter we briefly introduce the transition matrix and the adjoint system. Both concepts become essential in our latter discussion of filtering. Also, we discuss the problem of linearization and the useful structure of additive disturbances.

The next section discusses controllability and observability. An excellent complementary reference to this material is Brockett. These concepts are essential to the proper workings of any estimator. For example, if we do not have an observable system, then the estimation of some state may be impossible. This is part of what is called the inversion problem. The inversion problem is defined as the inability to estimate the state of a dynamical system based upon some prescribed set of measurements. This often arises in the system identification problem discussed in the previous section. There is a parallelism between deterministic and stochastic observability, which will be brought out in more complete detail in Chapter 6.

The last topic discussed in Chapter 2 concerns stability. Its importance is demonstrated in Chapter 6 and Appendix C. Further discussion of stability is contained in Brockett and in Ogata. Brockett's discussion is much more abstract, while that in Ogata follows the one presented here.

Chapter 3 considers a more abstract set of problems. The first important concept introduced here is the definition of a probability space and of a Markov process. This is important because many of the systems in common use are Markov in nature. Furthermore, all of the models that we consider are Markov.

The second major topic is that of independent increment processes. A special class of these processes is the Wiener process, also called the Brownian motion process. We spend the remainder of the chapter discussing the structure of such processes and their effect on dynamical systems. This leads us to introduce the Ito integral and the Fisk-Stratonovich integral.

An important observation made concerning the Wiener process was that in a loose sense its derivative is a white noise process. We show that actually the Wiener process is not of bounded variation and thus its derivative does not exist (Spiegel, p. 97), but for some practical applications we retain this formalism. Conceptually such a process is quite useful, since because it contains all frequencies with equal weight, it can excite every mode of a

dynamic system. The ramifications of this fact have been used extensively in communication theory (see Wozencraft and Jacobs or Van Trees [1]).

Much of Chapter 3 assumes a familiarity with abstract probability theory. This material is found in Doob [2]; Breiman; Feller [2]; or Loeve. The advanced concepts of diffusion processes are discussed in Ito and McKean and in Ito [2]. In McKean [2] the stochastic integral (Ito integral) is discussed in detail. Further ramifications of its use in random process theory are discussed. An introduction to semigroup theory is in Feller [2]. These extensions are not necessary for an understanding of the present theory, but extensions to infinite dimensional systems will require careful consideration of them. These extensions are contained in Falb. These previous references are in general more advanced than required for an understanding of the material in this chapter. They do, however, provide the completeness necessary for future research.

Chapter 4 presents results concerning the optimization criterion. In the past there have been many other criteria used, several of which are discussed in the books by Newton, Gould, and Kaiser; and Van Trees [1]. For our purposes the mean-square-error criterion is adequate. In order to show the existence and uniqueness of an estimate satisfying this criterion, it is useful to develop the structure of a Hilbert space. The concept of the Hilbert space is also discussed in Rudin [2]; Halmos [3]; Taylor; and Schmeidler.

The Hilbert space is a complete space: that is, every Cauchy sequency converges in that space. A second important property of Hilbert spaces is that the norm comes from an inner product. It is this fact that allows us to obtain the existence and uniqueness properties of optimum estimates. If the norm were not derived from an inner product, then we would have a Banach space. For a Banach space, the existence and uniqueness of orthogonal projections cannot be obtained.

The cost criterion is also called an optimization criterion. Thus, the extension from estimation to general optimization in Hilbert spaces is possible. This is carried out by Luenberger, who discusses the MMSE estimator as a special case of a more general set of constrained optimization criteria.

If we have a finite dimensional Hilbert space, we have the familiar finite dimensional vectors dealt with in Chapter 2 and in Halmos [4]. If we had started with this supposition, then existence and uniqueness could have been easily shown. This was the path initially chosen by Kalman [1]. The approach is used by Meditch [2] in a general exposition of linear filtering and by Meditch [1] for the problem of smoothing.

In the second section of Chapter 4 we discuss the problem of obtaining the MMSE estimate of a random variable, given a random process over some finite time interval. In order to present this adequately, we employ several results from measure theory, specifically, the Radon-Nikodym derivative.

Using an extended version of the definition of the martingale, we show the existence of a function defined on the probability space with the desired characteristics of a conditional expectation. We show that the MMSE estimate is $E[\mathbf{x}|\mathcal{O}_{t_0,t}]$, where $\mathcal{O}_{t_0,t}$ is the minimum σ-field generated by the observation process.

We conclude Chapter 4 with a derivation of the discrete-time version of the Kalman filter by means of the orthogonal projection lemma. This is used to demonstrate the power of the concept of orthogonal projection and also to obtain a basis for later comparisons. A computational framework for use of the discrete-time version of the Kalman filter is also discussed.

The first section of the second part of the book develops the solution to the MMSE problem. We say that it develops the solution because the prime interest is in obtaining the conditional density function and not the actual estimate. From the Fokker-Planck equations through the representation theorem the object is to obtain varying structures for the conditional density. We obtain two different methods. The first is a propagation equation for the conditional density, which is called the Kushner-Stratonovich equation. The second method uses the representation theorem of Bucy to obtain a function space representation of the conditional density.

In Chapter 5 we first discuss the Fokker-Planck equation (FPE) and the Feller-Kolmogorov equation (FKE). They are partial differential equations similar in form to the diffusion equation. It provides us with the transition density of the state of a dynamical system excited by white noise. This equation is quite useful in several areas:

1. When obtaining optimum expansion points for the implementation of estimators, the FPE can be used.
2. If we were to estimate the state of a dynamical system based upon very noisy measurements, then we would find that the FPE or FKE would provide that estimate. This is called a priori estimation or prediction. When the measurement noise is not excessively large, we use the measurements, and this is called a posteriori estimation.
3. The FPE derivation provides us with a technique that will be used to obtain the a posteriori estimate equations.

When the system is linear we find that the FPE is a classical diffusion equation whose solution is a Gaussian density. This should have been obvious from the discussion of the state transition matrix. That is, for a linear system the noise adds linearly. Such a superposition of Gaussian random variables will also be Gaussian.

The main fact used in obtaining the FPE is the Markov nature of the process. When such a supposition no longer holds, the equations must be altered to account for this. A solution to this question was given by the generalized

FPE. Unfortunately, its solution is quite difficult.

The FPE is also used in statistical mechanics to study fluctuation phenomena. It is possible to derive it from the Master equation, which is used in the study of nonequilibrium quantum statistical mechanics. This is discussed in Kac [2]. Applications to fluctuation phenomena are discussed in Reif.

The second section of Chapter 5 presents the solution to the estimation problem. It develops the propagation equation for the density function of the state variable conditioned on the measurement sets. Again the Markovian property plays a crucial role. We first derive the propagation equation for the conditional density in the case where the measurements are explicitly in terms of the state and are additively disturbed by white Gaussian noise of a given covariance. The resulting set of equations are called the Kushner-Stratonovich equations (KSE). They are nonlinear partial-differential integral equations. In general, then, solutions are unobtainable analytically. The second results are for Poisson measurements wherein the measurements implicitly reflect the effects of the state through an arrival rate for the Poisson process. Again we obtain a propagation equation for the conditional density of the state x at time t, given measurements from t_0 to t. The resulting equation is called Snyder's equation (SE), Snyder having first obtained it in 1970. It is also a nonlinear partial-differential integral equation.

We conclude this section with several numerically obtained results to show that the conditional density may be a multimodal density of quite complex shape.

The last section of Chapter 5 discusses Bucy's representation theorem, which is a function space representation of the conditional density. It has been shown by Kallianpur and Striebel [1]–[3] that the representation theorem approach is equivalent to the results obtained by the propagation methods. The use of the representation theorem is that it provides an alternate view of the estimation problem.

Chapter 6 considers the problem of implementation. The first part develops approximate estimation equations for the case of continuous time measurements. In this approach we follow Snyder [1], who considers expanding the nonlinearities in Taylor series and using quasi-Gaussian assumptions. Similar results were obtained by Bass, Norum, and Schwartz [1], [2] using differing assumptions. We obtain equations for both the estimate $\hat{x}$ and the covariance of the estimate. It is shown in the case of linear systems and measurements that these equations are exact. The result is the classical Kalman-Bucy equations.

The second section of Chapter 6 deals with the case of continuous systems and discrete measurements. The technique developed by Athans, Wishner, and Bertolini for an optimum driving function is developed. The use of this technique is in expanding the nonlinearities in a possibly more optimum

fashion than merely about the last estimate. An example is presented to compare the results.

The next section discusses a discrete-time measurement and system technique called *maximum a posteriori (MAP) estimation.* Instead of finding the mean of the conditional density, we obtain that value of the state that maximizes it. The results, originating with Cox, present a technique suitable for the use of dynamic programming. A linearized result is presented that is shown to be equivalent to the extended Kalman filter.

The final section of Chapter 6 discusses the concepts of divergence and stability on linear discrete-time filters. The divergence issues present an analysis of the problems encountered when certain inaccuracies arise in the implementation. The stability of the estimator equation is also discussed via the use of the Lyapanov theory developed in Chapter 2.

The three appendices develop side issues that relate to the general theory. Appendix A discusses the issue of the existence and uniqueness of the solutions to differential equations. In a similar fashion the existence and uniqueness questions for stochastic integral equations are presented in Appendix B. Finally, the stability results for the discrete-time estimator are developed in Appendix C.

CHAPTER 2

DYNAMICAL SYSTEMS

The nature of our world is such that most things depend upon time, and viewed in that fashion, their progression in time can possibly be both observed and influenced. There were certain nineteenth-century mathematicians and physicists who described nature as a deterministic system with many facets, albeit quite complex, yet in theory amenable to complete deterministic analyses. The position and motion of each particle and its interaction with every other particle could in principle be expressed, and the present, past, and future state of existence predicted. Such a grandiose scheme is a useful introduction to the ideas of dynamical systems, for in this chapter we are to discuss the idea of a system, propose a definition of the state of such a system, and delineate its structure.

The systems to be discussed in this chapter are completely deterministic in that there exist no random or uncertain portions. We first discuss such systems in a formal way by defining the state concept and then structuring the idea of a dynamic system. Such a system will be composed of inputs, outputs, and states as well as restrictions on how the states progress in time. We then present the canonical models of nonlinear state variable models of dynamic systems and measurements. Considerable advances have been made in understanding the consequences of such a definition of a dynamic system, but we shall not discuss them at length (see Kalman, Falb, and Arbib). Our main purpose in this chapter is to review to some extent concepts from dynamic-system theory that the reader may or may not have formally observed previously.

In the second section we develop the transition matrix and relate it to the transition function used in the definition of a dynamic system. The properties of the transition matrix are explored and the concept of the adjoint system is developed. The transition matrix is used as a basis for the study of discrete-time systems. Up to this point the set of times on which our states are defined is some interval of the real line. For discrete-time systems the domain becomes the integers; that is, the state progresses at discrete instances of time. Both linear and nonlinear discrete-time systems are developed.

Section three develops the ideas of controllability and observability for linear discrete- and continuous-time systems. These ideas are essential to the understanding of deterministic systems, and they have certain counterparts in the analysis of stochastic systems. Basically controllability is the property of a system that allows us to drive it to a given state by a suitable manipulation of the input variables over some finite set of time. Similarly, observability implies that by observing the output over a finite time, we can determine the state of the system. It will be noted that these properties will be representative of the representation of the specific system (see Desoer, Chapter 7).

Finally, we discuss the issue of stability of systems from a Lyapanov viewpoint. There are other viewpoints (see Willems) of importance that deal with such issues as passivity, but the Lyapanov theory as developed herein also provides us with more insight into the systems behavior as well as the basis for studies of the stability of estimation vis-à-vis stochastic controllability and observability concepts. In this section we first present several definitions of stability and then develop the idea of a Lyapanov function. This is directly related to the "energy" in a system and its ability to dissipate as a function of time. We conclude by proving results for linear discrete-time systems.

2.1. THE SYSTEM MODEL

Models of systems include dependent variables, those which we are interested in, and independent variables, those whose change produces changes in the variables of interest. The water level of a lake may be a variable of interest, and its behavior as a function of space and time may be sought. The independent variables in this context would be the space-time coordinates. In this book we shall concentrate on systems whose variables of interest depend solely on time, and our interest will focus on the values those variables take on as time changes. Systems of this sort are termed dynamic systems.

To fully describe the system at any one time, we may need information not only of the variable of interest but of other variables that are time dependent. Consider, for example, the position of a particle undergoing a time-varying force. Thus, to know the position of the particle, we can use Newton's laws of motion and observe that they yield information on the time rate of change of the velocity. Therefore, to know position, we must know the velocity, which in turn is determined by the force balance. In this simple case, knowledge of two quantities is sufficient to describe the complete motion of the single particle system. In other cases, as we shall see, several such variables will be necessary. As long as the number of such variables is finite, we have a finite dimensional system and these variables are called the state of the system.

In this section we shall introduce several formal concepts from system theory with the purpose of developing a solid foundation for an understanding of stochastic systems. We shall first introduce the concept of state and develop the more general structure of dynamic systems. For further discussions of these ideas the reader is referred to Kalman, Falb, and Arbib (Chapter 1); Desoer (Chapter 2); Athans and Falb (pp. 159–172); and Zadeh and Desoer (Chapter 1).

DEFINITION 1.1. The set $\mathbf{X}$ with elements $\mathbf{x} \in \mathbf{X}$, where $\mathbf{x}: T \to \mathbf{R}^n$ is called the state set of the system if knowledge of any state $\mathbf{x}$ at any time $t \in T$ is sufficient to fully describe the system at time t.

For example, the motion of a particle in a three-dimensional coordinate system is fully described by the three position and momentum coordinates. These six quantities, then, at any time t represent the state of this particular system.

Now a system can be affected by external disturbances or forcing functions that tend to change the state of the system. Also, there are observations made of the system that are given in terms of certain transformations of the state. This leads us to the definition of a dynamic system.

The state of a system is but one of the essential elements. It represents the internal workings of the system. There are two other concepts that are fundamental. The first is that of inputs to a system, or how we can by some set of external methods influence the state of the system. The second is the observations made on the system. For example, we may not observe all or even any of the states directly. We are usually presented with certain transformations of the states. All three of the above—state, input, and observations—are related to the independent parameter set, time. We shall let T represent the set of times of interest. Thus, all of the three are mappings or transformations in part from this basic parameter space into some other suitable spaces or sets.

The input then can be considered as a mapping from T into an appropriate space $\mathbf{U}$. The possible mappings are usually restricted to a certain class of functions denoted by $\mathcal{U}$. Similarly, the set of outputs can then be considered as mappings from T to an output space $\mathbf{Z}$. Again the set of output transformations is restricted to a given class $\mathcal{Z}$. In our case of interest the space for the inputs is $\mathbf{R}^k$, the k dimensional vector space. For example, a specific $\mathbf{u}(t)$ is a $k \times 1$ vector. Similarly for $\mathbf{X}$ the set of states is $\mathbf{R}^n$ and for $\mathbf{Z}$ the set of outputs is $\mathbf{R}^m$. All of these concepts can be formalized into a structural definition of a dynamic system.

DEFINITION 1.2. S is called a dynamic system and it is represented by the quintuple $\{\mathcal{U}, \mathbf{X}, \mathcal{Z}, \phi, h\}$ such that

(a) $\mathcal{U}$ is a set of input functions that map $T \to \mathbf{U}$ such that

$$\mathcal{U} = \{\mathbf{u}: \mathbf{u}: T \to \mathbf{U}\}$$

(b) $\mathbf{X}$ is the set of states.

(c) $\mathcal{Z}$ is a set of output functions which map $T \to \mathbf{Z}$ such that

$$\mathcal{Z} = \{\mathbf{z}: \mathbf{z}: T \to \mathbf{Z}\}$$

(d) $\phi(t_1\ \mathbf{x}_0, t_0, \mathbf{u})$ is called the state trasition function defined for all $t_1 \geqq t_0$ such that

$$\mathbf{x}(t_1) = \phi(t_1; \mathbf{x}_0, t_0, \mathbf{u})$$

where $\mathbf{u}$ represents $\mathbf{u}(s)$, $s \in [t_0, t_1]$. The state trasition function also satisfies the two axioms:

(i) For any interval $[t_0, t_1]$ and any functions $\mathbf{u}_1, \mathbf{u}_2 \in \mathcal{U}$ such that

$$\mathbf{u}_1(t) = \mathbf{u}_2(t); \quad \forall t \in [t_0, t_1]$$

then,

$$\phi(t_1; \mathbf{x}_0, t_0, \mathbf{u}_1) = \phi(t_1; \mathbf{x}_0, t_0, \mathbf{u}_2)$$

(ii) (Semigroup property) Let $t_0 \leqq t_1 \leqq t_2$. Then for all $\mathbf{x}_0$.

$$\phi(t_2; \mathbf{x}_0, t_0, \mathbf{u}) = \phi(t_2; \phi(t_1; \mathbf{x}_0, t_0, \mathbf{u}), t_1, \mathbf{u})$$

(e) There exists a function $\mathbf{h}$ such that

$$\mathbf{h}: T \times \mathbf{X} \times \mathbf{U} \to \mathbf{R}^m$$

and for all $t \in T$, $\mathbf{x}(t) \in \mathbf{X}$, $\mathbf{u}(t) \in \mathcal{U}$,

$$\mathbf{z}(t) = \mathbf{h}(\mathbf{x}(t), t, \mathbf{u}(t))$$

This definition is quite broad and provides a sufficiently large base to develop most systems of interest. The specific nature of the system dynamics is given by the trasition function $\phi(t; \mathbf{x}_0, t_0, \mathbf{u})$. The most important property of this function is the semigroup property, which states that the state at any time t, given the state at t_0, can also be obtained from the state at some intermediary time based upon knowledge of the initial state. Thus, knowledge of the state trasition function ϕ and the state at any previous time is sufficient to obtain the state at time t. A second point of the model is that of the nature of the output. That is, $\mathbf{h}(\ )$ is a zero-memory output in that what is observed at time t depends only on t, the state at time t, and the input at time t.

The systems we are most interested in are finite dimensional continuous- or discrete-time dynamic systems. They are defined as follows:

DEFINITION 1.3. A dynamic system S is finite dimensional if $\mathbf{X}$ is a finite dimensional linear space. It is a continuous-time system if T is the set of real numbers and discrete if T is the set of integers.

A special class of finite dimensional dynamical systems is those whose states are defined in terms of a differential equation of the form

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{f}(\mathbf{x}, t; \mathbf{u}(t)) \tag{1.1}$$

where $\mathbf{x}(t)$ and $\mathbf{f}(\mathbf{x}, t; \mathbf{u}(t))$ are $n \times 1$ vectors, $\mathbf{u}(t)$ is a $k \times 1$ vector, and $\mathbf{x}(t_0)$ is known. By direct integration we have

$$\mathbf{x}(t) = \int_{t_0}^{t} \mathbf{f}(\mathbf{x}, \xi; \mathbf{u}(\xi))\, d\xi \tag{1.2}$$

Then clearly the integral represents the transition function for a dynamic system. The function $\mathbf{u}(t)$ is the input to this system and is a $k \times 1$ vector.

The measurement $\mathbf{z}(t)$ is also considered to be a finite dimensional vector given by:

$$\mathbf{z}(t) = \mathbf{h}(\mathbf{x}(t), t) \tag{1.3}$$

The functions $\mathbf{z}(t)$ and $\mathbf{h}(\mathbf{x}(t), \mathbf{t})$ are $m \times 1$ vectors.

In general, systems of the form of (1.1) are too difficult to analyze, whereas systems of the form

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{f}(\mathbf{x}, t) + \mathbf{B}(\mathbf{x}, t)\, \mathbf{u}(t) \tag{1.4}$$

are more amenable to analysis. Systems of this form are called linearly driven dynamic systems. The matrix $\mathbf{B}(\mathbf{x}, t)$ is an $n \times k$ matrix transformation.

The system described by (1.4) and (1.3) is defined as the finite dimensional linearly driven dynamic system. This is frequently diagramed in terms of block diagrams as shown in Figure 2.1. In that figure we represent $\dot{\mathbf{x}}(t)$, the

Figure 2.1 Block diagram of system and measurement. (*a*) System; (*b*) measurement.

time derivative of the state as the sum of the nonlinear function $\mathbf{f}(\mathbf{x}(t), t)$ and the forcing function. The integral of $\dot{\mathbf{x}}(t)$ yields $\mathbf{x}(t)$.

The following example considers a specific implementation of a dynamic system represented in a state formulation.

Example. A mass m is suspended vertically on a spring that has a nonlinear restoring force f_s given by

$$f_s(x)=k_0x(1-a_0x) \tag{1.5}$$

where x is the instantaneous displacement of the mass from equilibrium. The mass is suspended in a tank that provides a viscous friction force $f_v(x)$, which depends on v the instantaneous velocity of the mass. It is given by

$$f_s(v)=B_0v(1+c_0v) \tag{1.6}$$

and v is the velocity of the mass given by

$$v=\frac{dx}{dt} \tag{1.7}$$

Writing a force balance on the system yields

$$m\frac{dv}{dt}=f_s(x)+f_s(v) \tag{1.8}$$

To reduce this to state variable form we let

$$x_1(t)=x(t) \tag{1.9}$$

and

$$x_2(t)=v(t)=\dot{x}(t)=\frac{dx_1(t)}{dt} \tag{1.10}$$

Then the force balance can be written as

$$\frac{dx_2}{dt}=\frac{B_0}{m}x_2(1+c_0x_2)+\frac{k_0}{m}x_1(1-a_0x_1) \tag{1.11}$$

The state equation becomes

$$\frac{d\mathbf{x}(t)}{dt}=\mathbf{f}(\mathbf{x},\ t) \tag{1.12}$$

where the components of the vector $\mathbf{f}(\mathbf{x},\ t)$ are

$$f_1(\mathbf{x},\ t)=x_2(t) \tag{1.13}$$

$$f_2(\mathbf{x},\ t)=\frac{B_0}{m}x_2(1+c_0x_2)+\frac{k_0}{m}x_1(1-a_0x_1) \tag{1.14}$$

We can now consider a special case of (1.4). Assume that the nonlinear function $\mathbf{f}(\mathbf{x}(t), t)$ is to be expanded about some arbitrary point $\mathbf{x}^*(t)$. This will be the multidimensional Taylor series formulation. Thus,

$$\mathbf{f}(\mathbf{x}(t),\ t)=\mathbf{f}(\mathbf{x}^*(t),t)+\mathbf{A}(\mathbf{x}^*(t),\ t)[\mathbf{x}(t)-\mathbf{x}^*(t)]$$
$$+\sum_{i=1}\gamma_i(\mathbf{x}(t)-\mathbf{x}^*(t))^T\,\mathbf{F}_i(\mathbf{x}^*(t),\ t)(\mathbf{x}(t)-\mathbf{x}^*(t))$$
$$+\cdots \tag{1.15}$$

where

$$\mathbf{A}(\mathbf{x}^*(t),\ t)=\left[\begin{array}{ccc}\dfrac{\partial f_1(\mathbf{x},t)}{\partial x_1} & & \dfrac{\partial f_1(\mathbf{x},\ t)}{\partial x_n}\\ \vdots & & \vdots\\ \dfrac{\partial f_n(\mathbf{x},\ t)}{\partial x_1} & & \dfrac{\partial f_n(\mathbf{x},t)}{\partial x_n}\end{array}\right]_{\mathbf{x}=\mathbf{x}^*(t)} \tag{1.16}$$

and

$$\mathbf{F}_i(\mathbf{x}^*(t),\ t)=\left[\begin{array}{ccc}\dfrac{\partial^2}{\partial x_1\,\partial x_1}f_i(\mathbf{x},\ t) & & \dfrac{\partial^2}{\partial x_1\,\partial x_n}f_i(\mathbf{x},\ t)\\ \dfrac{\partial^2}{\partial x_n\,\partial x_1}f_i(\mathbf{x},\ t) & & \dfrac{\partial^2}{\partial x_n\,\partial x_n}f_i(\mathbf{x},\ t)\end{array}\right]_{\mathbf{x}=\mathbf{x}^*(t)} \tag{1.17}$$

and γ_i is an $n\times 1$ vector with 1 in the ith row and zero elsewhere. Such expansions will be used in later chapters. A special case is that where

$$\mathbf{f}(\mathbf{x}(t),\ t)=\mathbf{A}(t)\,\mathbf{x}(t) \tag{1.18}$$

In this case the system is linear but time variant. Furthermore, if we have

$$\mathbf{B}(\mathbf{x}(t),\ t)\ =\ \mathbf{B}(t) \tag{1.19}$$

we have a linear time-varying dynamic system. In a similar fashion the measurement can be expanded, with a special case being

$$\mathbf{h}(\mathbf{x}(t),\ t)\ =\ \mathbf{C}(t)\mathbf{x}(t) \tag{1.20}$$

where $\mathbf{C}(t)$ is an $m\times n$ matrix.

DEFINITION 1.4. A linear time-varying dynamic system with linear measurements is given by the equation pair

$$\frac{d\mathbf{x}(t)}{dt}\ =\ \mathbf{A}(t)\mathbf{x}(t)\ +\ \mathbf{B}(t)\,\mathbf{u}(t) \tag{1.21}$$

$$\mathbf{z}(t)\ =\ \mathbf{C}(t)\,\mathbf{x}(t) \tag{1.22}$$

If $\mathbf{A}(t)$, $\mathbf{B}(t)$ $\mathbf{C}(t)$ are time invariant, the system is called a time-invariant linear dynamic system.

The block diagrams for the above systems are shown in Figure 2.2.

Example. A phase modulated signal is one of the form

$$z(t)\ =\ \cos(2\pi f_0 t\ +\ s(t)) \tag{1.23}$$

where $s(t)$ is the phase modulation term. A possible variation in phase may be a parabolic dependence on time, that is,

Figure 2.2 Block diagram of linear dynamic system. (*a*) System; *b*) measurement.

$$s(t) = \alpha t^2 \tag{1.24}$$

Now consider the term t^2 by itself. If we define $\dot{x}_1(t)$ through

$$\dot{x}_1(t) = 0 \tag{1.25}$$

Then clearly $x_1(t)$ is some constant. Now, if we let

$$\dot{x}_1(t) = 0 \tag{1.26a}$$
$$\dot{x}_2(t) = x_1(t) \tag{1.26b}$$
$$\dot{x}_3(t) = x_2(t) \tag{1.26c}$$

we can easily show that $x_3(t)$ has the form

$$x_3(t) = C_1 \frac{t^2}{2} + C_2 t + C_3 \tag{1.27}$$

Then by choosing $x_3(0) = 0$, $x_2(0) = 0$, $x_1(0) = 2$, we can show that

$$x_3(t) = t^2 \tag{1.28}$$

Thus $s(t)$ can be given by

$$s(t) = \mathbf{C}\,\mathbf{x}(t) \tag{1.29}$$

where $\mathbf{C}$ is the 1×3 vector

$$\mathbf{C} = [0 \;\; 0 \;\; \alpha] \tag{1.30}$$

The state $\mathbf{x}(t)$ satisfies the equation

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) \tag{1.31}$$

with

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \tag{1.32}$$

Alternate realizations of this are possible by using a forcing function $\mathbf{u}(t)$.

This previous example also indicates an interesting fact that an n dimensional state equation can generate an $n-1$ dimensional polynomial. This is useful in circumstances where an arbitrary function defined on a closed interval $[t_1, t_2]$ can be suitably modeled by an $n-1$ dimensional polynomial via a least-squares fit or an osculating polynomial technique.

Example. Consider the following nth-order differential equation:

$$y^n + a_{n-1}y^{n-1} + \cdots + a_1\dot{y} + a_0 = u_0(t) \tag{1.33}$$

with initial conditions $y(t_0), \dot{y}(t_0), \cdots, y^{n-1}(t_0)$. Now define the following variables:

$$x_1(t) = y(t) \tag{1.34a}$$
$$x_2(t) = \dot{y}(t) \tag{1.34b}$$
$$\vdots$$
$$x_{n-1}(t) = \frac{d^{n-2}}{dt^{n-2}}y(t) \tag{1.34c}$$
$$x_n(t) = \frac{d^{n-1}}{dt^{n-1}}y(t) \tag{1.34d}$$

Then, clearly,

$$\dot{x}_1(t) = x_2(t) \tag{1.35a}$$
$$\dot{x}_2(t) = x_3(t) \tag{1.35b}$$
$$\vdots$$
$$\dot{x}_{n-1}(t) = x_n(t) \tag{1.35c}$$

And the differential equation can be written as

$$\dot{x}_n(t) + a_{n-1}x_n(t) + \cdots + a_1x_1(t) + a_0 = u_0(t) \tag{1.36}$$

If we now define the vector $\mathbf{x}(t)$ as

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix} \tag{1.37}$$

we then have the state equation

$$\dot{\mathbf{x}}(t) = \mathbf{A}\,\mathbf{x}(t) + \mathbf{u}(t) \tag{1.38}$$

where

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \cdots & 0 & 0 \\ 0 & 0 & 1 \cdots & 0 & 0 \\ 0 & 0 & 0 \cdots & 0 & 0 \\ 0 & 0 & 0 \cdots & 0 & 1 \\ -a_1 & -a_2 & -a_3 & -a_{n-1} & -a_n \end{bmatrix} \tag{1.39}$$

and

$$\mathbf{u}(t) = \begin{bmatrix} 0 \\ \vdots \\ -a_0 + u_0(t) \end{bmatrix} \tag{1.40}$$

with the initial condition vector $\mathbf{x}(t_0)$.

The above example shows that an nth-order differential equation can be written as a first-order differential equation in terms n—1 state variables. This holds in general (see Ince, p. 14) and is one of the fundamental uses of the state variable formulation.

Example. Consider the case of a simple L-R circuit with a voltage input $u(t)$. Let the state be $x(t)$, the current through the devices. The state equations are

$$\dot{x} = -x(t) + u(t) \tag{1.41}$$

where $u(t)$ is the input voltage and $x(t)$ is the current through the devices and $R/L = 1$.

Figure 2.3 Example of R-L circuit with input voltage hand state.

This is a linearly driven system.

Example. Now let R be time varying:

$$R = 1 + u_1(t) \tag{1.42}$$

Then

$$L\dot{x}(t) = -(1 + u_1(t))x(t) + u_2(t) \tag{1.43}$$

where

$$u_2(t) = u(t) \tag{1.44}$$

In this case, the system is *not* a linearly driven system if we consider $u_1(t)$ a control.

There are many other examples of systems structures that fall within the context of the structures developed in this chapter. We develop several of these in the problems. However, the basic issue is that of the concept of a state and the propagation of the value of the state as time changes. Extentions to arbitrary systems can be found in many places. Pindyck has used the state variable formulation in a macroeconomic model, McGarty [2] for modeling atmospheric structure, and Snyder [4] for modeling biomedical systems.

2.2 THE TRANSITION MATRIX AND THE DISCRETE-TIME SYSTEM

Having discussed at length the more general formulation of a dynamical system, we now turn our attention to the most studied, the linear time-varying dynamical system. It is described by the equation

$$\dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t) \tag{2.1}$$

We intend to do three specific things in regard to (2.1). First, we shall study the undriven solution. Second, a formulation for the driven system will be presented. These two discussions will introduce the transition matrix concept. We shall not show how to calculate the transition matrix; this is adequately covered in Athans and Falb; Ogata; Zadeh and Desoer; Brockett; and DeRusso, Roy, and Close. It may not always be easily calculated, yet its use in discussing some general topics is invaluable.

For the system in (2.1) we know that for the case of no forcing function, $\mathbf{u}(t) = \mathbf{0}$, the state at time t depends linearly upon the state at time t_0. That is, there exists a linear transformation of the form $\Phi(t, t_0)$, where this is an $n \times n$ matrix, such that

$$\mathbf{x}(t) = \Phi(t, t_0)\,\mathbf{x}(t_0) \tag{2.2}$$

The transformation $\Phi(t, t_0)$ relates the transformation of the state from one time to another. Differentiate (2.2) with respect to time to yield

$$\dot{\mathbf{x}}(t) = \dot{\Phi}(t, t_0)\,\mathbf{x}(t_0) = \mathbf{A}(t)\mathbf{x}(t) \tag{2.3}$$

But this implies

$$\dot{\Phi}(t, t_0)\,\mathbf{x}(t_0) = \mathbf{A}(t)\,\Phi(t, t_0)\,\mathbf{x}(t_0)$$

However, this must hold for all possible $\mathbf{x}(t_0)$. Now the set of $\mathbf{x}(t_0)$ can be chosen such that they span $\mathbf{R}^n$, so that

$$\dot{\Phi}(t, t_0) = \mathbf{A}(t)\,\Phi(t, t_0) \tag{2.4}$$

The matrix $\Phi(t, t_0)$ is called the transition matrix. It has the following properties:

$$\Phi(t, t_0)\,\Phi(t_0, t) = \Phi(t, t) = \mathbf{I} \tag{2.5}$$

where $\mathbf{I}$ is the $n \times n$ identity matrix; also,

$$\Phi(t_1, t_2)\,\Phi(t_2, t_3) = \Phi(t_1, t_3) \tag{2.6}$$

and

$$\Phi^{-1}(t, t_0) = \Phi(t_0, t) \tag{2.7}$$

Example. Consider the linear time-invariant system

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t);\ \mathbf{x}(t_0) = \mathbf{x}_0 \tag{2.8}$$

where $\mathbf{A}$ is a constant $n \times n$ matrix. Then we can show by substitution that

$$\Phi(t, t_0) = e^{\mathbf{A}(t - t_0)} \tag{2.9}$$

where

$$e^{\mathbf{A}(t - t_0)} = \sum_{k=0}^{\infty} \frac{\mathbf{A}^k (t - t_0)^k}{k!} \tag{2.10}$$

satisfies the differential equation for the transition matrix. The matrix exponential in (2.10) can be shown to exist (converge). Then the solution to (2.8) is

$$\mathbf{x}(t) = e^{\mathbf{A}(t - t_0)}\mathbf{x}_0 \tag{2.11}$$

Consider now the system driven by $\mathbf{B}(t)\mathbf{u}(t)$, as given in (2.1). We will show that

$$\mathbf{x}(t) = \Phi(t, t_0)\mathbf{x}_0 + \int_{t_0}^{t} \Phi(t, \xi)\,\mathbf{B}(\xi)\,\mathbf{u}(\xi)\,d\xi \tag{2.12}$$

is the solution to this equation. This can easily be done by showing that (2.12) substituted into (2.1) yields an identity. Differentiating (2.12) with respect to time yields

$$\dot{\mathbf{x}}(t) = \dot{\Phi}(t, t_0)\,\mathbf{x}_0 + \int_{t_0}^{t} \dot{\Phi}(t, \xi)\,\mathbf{B}(\xi)\,\mathbf{u}(\xi)\,d\xi + \Phi(t, t)\,\mathbf{B}(t)\,\mathbf{u}(t) \tag{2.13}$$

Now, using (2.2) and (2.3), we have

$$\dot{\mathbf{x}}(t) = \mathbf{A}(t)\Phi(t, t_0)\,\mathbf{x}_0 + \int_{t_0}^{t} \mathbf{A}(t)\Phi(t, \xi)\,\mathbf{B}(\xi)\,\mathbf{u}(\xi)\,d\xi + \mathbf{B}(t)\,\mathbf{u}(t) \tag{2.14}$$

But

$$\dot{\mathbf{x}}(t) = \mathbf{A}(t)\,\mathbf{x}(t) + \mathbf{B}(t)\,\mathbf{u}(t) \tag{2.15}$$

Thus, using (2.12) on the right of (2.15), we have

$$\dot{\mathbf{x}}(t) = \mathbf{A}(t)\int_{t_0}^{t} \Phi(t, \xi)\,\mathbf{B}(\xi)\,\mathbf{u}(\xi)\,d\xi + \mathbf{A}(t)\Phi(t, t_0)\,\mathbf{x}_0 + \mathbf{B}(t)\,\mathbf{u}(t) \tag{2.16}$$

which satisfies the identity and shows that indeed (2.12) is the solution to the forced system.

The structure of the solution to equation (2.1) was through the matrix $\mathbf{A}(t)$ and the resulting transition matrix. In a similar fashion we can generate another system using the same $\mathbf{A}(t)$, called the adjoint system. Let $\mathbf{p}(t)$satisfy

$$\dot{\mathbf{p}}(t) = -\mathbf{A}^T(t)\,\mathbf{p}(t) \tag{2.17}$$

If $\Phi(t, t_0)$ is the solution of (2.4) then $\Phi^T(t_0, t)$ is the transition matrix of (2.17). This is easy to prove. Recall that

$$\Phi(t, t_0)\,\Phi(t_0, t) = \mathbf{I} \tag{2.18}$$

Now differentiate with respect to t:

$$\frac{d}{dt}\left[\Phi(t, t_0)\,\Phi(t_0, t)\right] = \dot{\Phi}(t, t_0)\,\Phi(t_0, t) + \Phi(t, t_0)\,\dot{\Phi}(t_0, t) = 0 \tag{2.19}$$

But, using (2.4)

$$0 = \mathbf{A}(t)\,\Phi(t, t_0)\,\Phi(t_0, t) + \Phi(t, t_0)\,\dot{\Phi}(t_0, t) \tag{2.20}$$

which yields

$$\Phi(t, t_0)\,\dot{\Phi}(t_0, t) = -\mathbf{A}(t) \tag{2.21}$$

or

$$\dot{\Phi}(t_0, t) = -\Phi^{-1}(t, t_0)\,\mathbf{A}(t) \tag{2.22}$$

But using (2.7)

$$\dot{\Phi}(t_0, t) = -\Phi(t_0, t)\,\mathbf{A}(t) \tag{2.23}$$

and taking the transpose yields

$$\dot{\Phi}^T(t_0, t) = -\mathbf{A}^T(t)\,\Phi^T(t_0, t) \tag{2.24}$$

We generally think of systems of the form of (2.1), (2.8), and (2.17) in block-diagram form as is shown in Figure 2.4. This is a useful concept and will be exploited throughout our discussions on both linear and nonlinear systems.

This then completes the discussion of the transition matrix. The reader should refer to some of the references to see how it is used for some varied systems. We shall close this section with a simple example.

Example. Consider a simple one-dimensional harmonic oscillator governed by the following equation:

$$m\ddot{x} + kx = w'(t) \tag{2.25}$$

where m is its mass, k is the spring constant, and $w'(t)$ is the force externally applied. In this case $w'(t)$ is positive if applied in $+x$ direction and negative

Figure 2.4 Block diagrams of dynamic systems. (*a*) Driven system; (*b*) undriven system; (*c*) adjoint system.

in the $-x$ direction. This may be thought of as the system diagramed in Figure 2.5, which shows a mass attached to a spring sliding along a frictionless shaft. Small particles are transferring energy to the mass by momentum transfer in both positive and negative x directions. Let us assume that k/m is unity and define $w(t)$ as $w'(t)/m$. The resulting equation is

Figure 2.5 One-dimensional mass on a spring with forcing functions.

$$\ddot{x} + x = w(t) \tag{2.26}$$

We can reduce this to state form. Let us define $x_1(t) = x(t)$ and $x_2(t) = \dot{x}(t)$. Then the state equations are

$$\dot{x}_1 = x_2 \tag{2.27}$$

$$\dot{x}_2 = -x_1 + w(t) \tag{2.28}$$

In matrix form this becomes

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{w} \tag{2.29}$$

where now $\mathbf{x}$ is a 2×1 vector not to be confused with the initial displacement. Written out completely, it gives us for (2.29)

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ w \end{bmatrix} \tag{2.30}$$

We shall assume that the system has zero initial conditions. This implies that $w(t)$ is the sole cause of motion.

One method for calculating the transition matrix for this system is by means of the Laplace transform. We first take the Laplace transform of both sides of the state equation:

$$s\mathbf{x}(s) = \mathbf{A}\mathbf{x}(s) + \mathbf{w}(s) \tag{2.31}$$

Rearranging this yields

$$\mathbf{x}(s) = (\mathbf{I}s - \mathbf{A})^{-1}\mathbf{w}(s) \tag{2.32}$$

The inverse is calculated by taking the inverse transform of

$$[\mathbf{I}s - \mathbf{A}]^{-1} = \begin{bmatrix} s & -1 \\ 1 & s \end{bmatrix}^{-1} = \frac{\begin{bmatrix} s & +1 \\ -1 & s \end{bmatrix}}{s^2+1} \tag{2.33}$$

which yields

$$\Phi(t, t_0) = \begin{bmatrix} \cos(t - t_0) & \sin(t - t_0) \\ -\sin(t - t_0) & \cos(t - t_0) \end{bmatrix} \tag{2.34}$$

Then for some prescribed set of initial conditions the state equation has the solution

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} \cos(t - t_0) & \sin(t - t_0) \\ -\sin(t - t_0) & \cos(t - t_0) \end{bmatrix} \begin{bmatrix} x_1(t_0) \\ x_2(t_0) \end{bmatrix} + \int_{t_0}^{t} \begin{bmatrix} \cos(t - \xi) & \sin(t - \xi) \\ -\sin(t - \xi) & \cos(t - \xi) \end{bmatrix} \begin{bmatrix} 0 \\ w(\xi) \end{bmatrix} d\xi \tag{2.35}$$

Thus, the system undergoes harmonic motion as would be suspected for this harmonic oscillator.

The transition matrix is also useful for describing discrete-time systems. Consider now the system given by

Figure 2.7 Block diagrams of linear dynamic systems.
(*a*) Continuous; (*b*) discrete.

$$\mathbf{z}(k+1) = \mathbf{C}(k+1)\,\mathbf{x}(k+1) \tag{2.41}$$

A comparison of these two distinct formulations is shown in Figure 2.7.

Example. Let $x(t)$ be a dynamic system given by

$$\dot{x}(t) = -\alpha x(t) + u(t) \tag{2.42}$$

Then the transition matrix is given by

$$\Phi(t, t') = \exp[-\alpha(t - t')] \tag{2.43}$$

Then the discrete form is given by

$$\Phi(k+1, k) = \exp(-\alpha T) \tag{2.44}$$

where again T is the sample time.

Now assume that $u(k) = u_0$ for all k. Then

$$x(k+1) = \Phi(k+1, k)\, x(k) + u(k) \tag{2.45}$$

Now, given $x(0)$, $x(k+1)$ can be solved directly. That is.

$$x(1) = \Phi(1, 0)\, x(0) + u(0) \tag{2.46}$$

$$x(2) = \Phi(2, 1)\, x(1) + u(1) \tag{2.47}$$

which is equivalent to

$$x(2) = \Phi(2, 0)\, x(0) + u(1) + \Phi(1, 0)\, u(0) \tag{2.48}$$

or in general

$$\begin{aligned} x(t+1) &= \Phi(k+1, 0)\, x(0) + \sum_{m=0}^{k} \Phi(k, m)\, u(m) \\ &= \Phi(k+1, 0)\, x(0) + \left(\sum_{n=0}^{k} \Phi(k, n)\right) u_0 \end{aligned} \tag{2.49}$$

By letting $\delta = \exp(-\alpha T)$, we can easily show that

$$x(k+1) = \delta^{k+1} x(0) + u_0 \frac{1 - \delta^k}{1 - \delta} \tag{2.50}$$

The parallelism between continuous systems and discrete systems is quite straightforward. The previous example was for a simple scalar linear time-invariant discrete-time system. In a similar fashion, we can extend the model to that of the vector nonlinear time-varying dynamical system introduced in the previous section. In that section we introduced in Definition 1.2 the transition function. This was the function that gave the value of the state at any time t, given the state at some time τ and the input over the prescribed interval. Specifically,

$$\mathbf{x}(t) = \phi(t; \mathbf{x}, \tau, \mathbf{u}) \tag{2.51}$$

In this section we saw that for linear systems

$$\mathbf{x}(t) = \Phi(t, \tau)\, \mathbf{x}(\tau) + \int_{\tau}^{t} \Phi(t, \xi)\, \mathbf{u}(\xi)\, d\xi \tag{2.52}$$

Thus the operator defined in the above expression is the transition function. Note that it is an operator because it operates on the entire trajectory of the input. For the nonlinear system over an incremental interval, we have

$$\mathbf{x}(t + dt) = \mathbf{x}(t) + \mathbf{f}(\mathbf{x}(t), t)\, dt + \mathbf{B}(\mathbf{x}(t), t)\, \mathbf{u}(t) dt \tag{2.53}$$

We can now use this to develop a formalism for the discrete-time approach. That is, let $\mathbf{x}(k+1)$ be the state at $(k+1)\,T$. Then

$$\mathbf{x}(k+1) = \mathbf{f}(\mathbf{x}(k), k) + \mathbf{B}(\mathbf{x}(k), k)\mathbf{u}(k) \tag{2.54}$$

It should be clear that this is a formalism and that the function $f(\ )$ in the discrete case is not the same as that in the continuous-time case. Furthermore, if $\mathbf{u}(k)$ equals zero for all k, then we note that $\mathbf{f}(\mathbf{x}(k), k)$ equals $\phi((k+1)T; \mathbf{x}, kT, 0)$. In most other cases this direct parallelism breaks down.

In a similar fashion, the output for the nonlinear case can be written directly as

$$\mathbf{z}(k+1) = \mathbf{h}(\mathbf{x}(k+1), k+1) \tag{2.55}$$

Here the parallelism is direct and $\mathbf{h}(\mathbf{x}(k+1), k+1)$ equals $\mathbf{h}(\mathbf{x}(t), t)$ at t equal to $(k+1)T$.

We shall use the linear and nonlinear discrete-time structure extensively in the following chapters. Of use will be both the direct relationship provided by the transition function for linear systems and the more formalistic parallelism for the nonlinear case.

2.3. CONTROLLABILITY AND OBSERVABILITY

We now wish to discuss two vital concepts in modern linear control theory that are also central issues in the theory of optimum filtering. These concepts were formalized in context of the control problem by Kalman, and it was due to his work in filtering and optimal control that we have the strong connection. In order to discuss these issues, we must now introduce the concept of the measurement system. Let $\mathbf{z}(t)$ be an $m \times 1$ vector that we have available:

$$\mathbf{z}(t) = \mathbf{C}(t)\mathbf{x}(t) + \mathbf{v}(t) \tag{3.1}$$

Now $\mathbf{C}(t)$ is an $m \times n$ matrix, and it relates the system states to the measurements. The $m \times 1$ vector $\mathbf{v}(t)$ is just some known (it may be random) bias term. The more general formulations of (3.1) is a nonlinear formulation given as

$$\mathbf{z}(t) = \mathbf{h}(\mathbf{x}(t), t) + \mathbf{v}(t) \tag{3.2}$$

Here the measurement is embedded in a nonlinear function of the state. But again $\mathbf{v}(t)$ is additive. The most general formulation given as

$$\mathbf{z}(t) = \mathbf{h}(\mathbf{x}(t), \mathbf{v}(t), t) \tag{3.3}$$

In our usage only (3.1) and (3.2) will be used. The expression (3.3) is too complex to make any useful remarks about at the present time.

The concepts of controllability and observability are best expressed in terms of the linear time-varying system. This system is defined by the set of equations

$$\dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t) \tag{3.4}$$
$$\mathbf{z}(t) = \mathbf{C}(t)\mathbf{x}(t) \tag{3.5}$$
$$\mathbf{x}(t_0) = \mathbf{x}_0 \tag{3.6}$$

The concept of controllability is exemplified by asking the following question: Under what conditions on $\mathbf{u}(t)$ is it possible to transfer the system at t_0, and $\mathbf{x}_0$ to the origin at t_1, or for that matter to any state $\mathbf{x}_1$? Here t_1 is a finite time. This is an input-oriented concept.

In contrast, observability is an output-oriented concept. It asks the question, Under what conditions is it possible in a finite time to establish the past states of the system given the measurements $\mathbf{z}(t)$? In the context of the filtering or estimation problem, it means that, given the measurements $\mathbf{z}(t)$, we can actually infer something about all the states and that there are no states whose behavior cannot be inferred by observation.

We shall follow Meditch [2] in the presentation of both controllability and observability arguments for continuous-and discrete-time systems.

Let us now define observability and obtain a set of necessary and sufficient condtiions for the case of linear time-varying systems.

DEFINITION 3.1. A system (discrete or continuous), is said to be *observable* if for some time $t_1 > t_0$ the state $\mathbf{x}(t_0)$ can be fully determined from the set of measurements $\{\mathbf{z}\}$ over the interval $[t_0, t_1]$. If this is true for any t_0, then the system is termed completely observable.

We can now prove the following theorem on continuous-time observability.

THEOREM 3.1.

The continuous system given by (3.4) and (3.5) is completely observable if and only if the symmetric $n \times n$ matrix

$$\mathbf{M}_c(t_0, t_1) = \int_{t_0}^{t_1} \boldsymbol{\Phi}^T(t, t_0)\, \mathbf{C}^T(t)\, \mathbf{C}(t)\, \boldsymbol{\Phi}(t, t_0)\, dt \tag{3.7}$$

is positive definite for some $t_1 > t_0$.

Proof. Since we assume that we know $\mathbf{u}(t)$, let it be zero. The system is

$$\dot{\mathbf{x}}(t) = \mathbf{A}(t)\, \mathbf{x}(t) \tag{3.8}$$
$$\mathbf{z}(t) = \mathbf{C}(t)\, \mathbf{x}(t) \tag{3.9}$$

Let us first prove sufficiency, that is, if (3.11) is true then from $\mathbf{z}(t)$ we can get $\mathbf{x}(t_0)$. Now recall that

$$\mathbf{x}(t) = \boldsymbol{\Phi}(t, t_0)\, \mathbf{x}(t_0) \tag{3.10}$$

Using (3.10) in (3.9), we have

$$\mathbf{z}(t) = \mathbf{C}(t)\, \boldsymbol{\Phi}(t, t_0)\mathbf{x}(t_0) \tag{3.11}$$

Now premultiply both sides by

$$\Phi^T(t, t_0)\, \mathbf{C}^T(t) \tag{3.12}$$

and integrate over (t_0, t_1)

$$\int_{t_0}^{t_1} \Phi^T(t, t_0)\, \mathbf{C}^T(t)\, \mathbf{z}(t)\, dt = \int_{t_0}^{t_1} \Phi^T(t, t_0)\, \mathbf{C}^T(t)\, \mathbf{C}(t)\, \Phi(t, t_0)\, dt\, \mathbf{x}(t_0) \tag{3.17}$$

Using the definition of $\mathbf{M}_c(t_0, t)$, we have

$$\mathbf{M}_c(t_0, t_1)\mathbf{x}(t_0) = \int_{t_0}^{t_1} \Phi^T(t, t_0)\mathbf{C}^T(t)\mathbf{z}(t)\, dt \tag{3.18}$$

Now, since $\mathbf{M}_c(t_0, t_1)$ is positive definite, its inverse exists; therefore,

$$\mathbf{x}(t_0) = \mathbf{M}_c^{-1}(t_0, t_1) \int_{t_0}^{t_1} \Phi^T(t, t_0)\mathbf{C}^T(t)\mathbf{z}(t)\, dt \tag{3.19}$$

Now let us prove necessity. This is quite simple if we do so by contradiction. Using (3.17) and the fact that $\mathbf{M}_c(t_0, t_1)$ is not positive definite, we obtain the following inequality;

$$\mathbf{x}^T(t_0)\mathbf{M}_c(t_0, t_1)\mathbf{x}(t_0) \leqq 0 \tag{3.20}$$

yields

$$\mathbf{x}^T(t_0) \int_{t_0}^{t_1} \Phi^T(t, t_0)\mathbf{C}^T(t)\, \mathbf{z}(t)dt \leqq 0 \tag{3.21}$$

or

$$\int_{t_0}^{t_1} [\mathbf{C}(t)\, \Phi(t,t_0)\mathbf{x}(t_0)]^T\, \mathbf{z}(t)dt \leqq 0 \tag{3.22}$$

yields

$$\int_{t_0}^{t_1} \mathbf{z}^T(t)\mathbf{z}(t)\, dt \leqq 0 \tag{3.23}$$

But it can never be less than zero. The case of $\mathbf{z}(t)$ being zero implies $\mathbf{x}(t_0)$ is zero. But this yields a contradiction; thus, for anynonzero $\mathbf{x}(t_0)$, $\mathbf{M}_c(t_0, t_1)$ must be positive definite. ∎

The following corollary presents a simpler condition for testing observability for linear time-invariant systems.

COROLLARY 3.1. Let $\mathbf{x}(t)$ be a linear time-invariant system given by

$$\dot{\mathbf{x}}(t) = \mathbf{A}\, \mathbf{x}(t) \tag{3.24}$$

and $\mathbf{z}(t)$ by

$$\mathbf{z}(t) = \mathbf{C}\, \mathbf{x}(t) \tag{3.25}$$

Then the system is completely observable if and only if the observability matrix $\mathbf{M}_c$

$$\mathbf{M}_c = [\mathbf{C}^T \mid \mathbf{A}^T\mathbf{C}^T \mid \cdots \mid (\mathbf{A}^{n-1})^T\mathbf{C}^T] \tag{3.26}$$

is of rank n.

Proof. See Polak and Wong (pp. 39–41). ∎

The following example discusses the implications of this theorem in a physically meaningful system.

Example. Let x_1 and x_2 represent the x and y position coordinates of an incoming missile. Assume that the projectile can be governed by linear dynamics (see Athans, Wishner, and Bertolini). Thus, the x and y velocity coordinates are x_3 and x_4, respectively. This means that:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \tag{3.27}$$

is the state vector and

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{w}(t) \tag{3.28}$$

is the state description. The measurement may just be the range. Then,

$$z(t) = [x_1^2(t) + x_2^2(t)]^{1/2} + v(t) \tag{3.29}$$

Figure 2.8 Radar tracking example.

The physical parameters and the ballistic trajectory are shown in Figure 2.8. Using an assumed trajectory, one can obtain a linearized approximation of the measurement:

$$z(t) = [c_1 \mid c_2 \mid 0 \mid 0]\,\mathbf{x} + v(t) \tag{3.30}$$

where c_i is the partial derivative at $[x_1^2(t) + x_2^2(t)]^{1/2}$ with respect to $x_i(t)$ evaluated along the reference trajectory.

We know that this system is observable if and only if the matrix

$$\mathbf{P} = [\mathbf{C}^T \mid \mathbf{A}^T\mathbf{C}^T \mid (\mathbf{A}^T)^2\,\mathbf{C}^T \mid (\mathbf{A}^T)^3\,\mathbf{C}^T] \tag{3.31}$$

is of rank n.

Let us assume that the x and y motion of the craft is decoupled. Then $\mathbf{A}$ has the form

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ a_{31} & a_{32} & 0 & a_{43} \\ 0 & 0 & a_{43} & a_{44} \end{bmatrix} \tag{3.32}$$

The observability matrix becomes

$$\mathbf{P} = \begin{bmatrix} c_1 & 0 & a_{31}c_1 & a_{31}a_{43}c_2 \\ c_2 & 0 & a_{32}c_1 & a_{32}a_{43}c_2 \\ 0 & c_1 & a_{43}c_2 & a_{31}c_1 + a_{43}^2 c_2 \\ 0 & c_2 & a_{44}c_2 & a_{31}c_1 + a_{44}^2 c_2 \end{bmatrix} \tag{3.33}$$

Clearly the fourth column is a linear combination of the other three. Thus, the matrix is not of rank n, so the system is *not* observable. Therefore, with range alone we cannot estimate position. This should have been obvious from the start. Our reason for mentioning it is to indicate that what is so obvious in this instance may be obscured by the complexity of a large-scale system.

The concept of observability is extremely important in estimation, where it is classically called *invertability*. We shall discuss this in depth in Chapter 7 when we consider the conditions necessary for the convergence of estimates.

We shall now consider the discrete case that is an analogue to the continuous version. We shall only present the theorem; if the reader is interested, the proof is in Meditch [2].

THEOREM 3.2.

For the discrete time system given by (2.40) and (2.41) the system is completely observable if and only if the symmetric $n \times n$ matrix

$$\mathbf{M}_d(0, N) = \sum_{i=1}^{N} \boldsymbol{\Phi}^T(i, 0)\,\mathbf{C}^T(i)\,\mathbf{C}(i)\,\boldsymbol{\Phi}(i, 0) \tag{3.34}$$

is positive definite for some $N > 0$.

This matrix is also called the *observability matrix* of state $\mathbf{x}(0)$ given N measurements. We can generalize this for the observability matrix of state $\mathbf{x}(k)$ given N measurements, $\mathbf{M}_d(k, N)$. This is discussed in Problem 2.9. In Chapter 7 we evaluate a stochastic observability matrix with similar properties.

There are many extensions of the above two systems that give considerable simplification and provide simple checks on observability. The reader may consult Ogata (pp. 370–436) and Athans and Falb (pp. 200–211). Moreover, Kalman [3. pp. 337–348] presents an excellent discussion of the $M(t_0, t_1)$ matrix for the Gaussian noise case and shows it to be the Fisher information matrix of statistics. He also provides the simplifications just outlined.

We shall now discuss controllability.

DEFINITION 3.2. A linear system is said to be *controllable* at time t_0 if there exists a control set $\{\mathbf{u}(t)\}$ where

$$\{\mathbf{u}(t)\} = \{\mathbf{u}(t): t \in [t_0, t_1]\} \tag{3.35}$$

depending on $\mathbf{x}(t_0)$ for which $\mathbf{x}(t_1) = 0$. If this is true for all $\mathbf{x}(t_0)$ and t_0, the system is said to be *completely controllable*.

Figure 2.9 Example of controllability conditions.

This definition means that if a system is completely controllable, there exists some $u(t)$ that can drive the system from t_0, $\mathbf{x}_0$ to the origin in some finite time t_1. An example of such a control is shown in Figure 2.9. The choice of the origin is completely arbitrary. Any other point would suffice.

THEOREM 3.3.

The continuous linear system is completely controllable if and only if the symmetric $n \times n$ matrix

$$\mathbf{W}_c(t_0, t_1) = \int_{t_0}^{t_1} \Phi(t_0, t)\mathbf{B}(t)\mathbf{B}^T(t)\Phi^T(t_0, t)\, dt \tag{3.36}$$

is positive definite for some $t_1 > t_0$.

Proof. We shall first prove the sufficiency.
Recall that

$$\mathbf{x}(t_1) = \Phi(t_1, t_0)\mathbf{x}(t_0) + \int_{t_0}^{t_1} \Phi(t_1, \xi)\mathbf{B}(\xi)\mathbf{u}(\xi)\, d\xi \tag{3.37}$$

Now let us choose an arbitrary control

$$\mathbf{u}(t) = -\mathbf{B}^T(t)\Phi^T(t_0, t)\mathbf{W}_c^{-1}(t_0, t_1)\mathbf{x}(t_0) \tag{3.38}$$

Here we have used the hypothesis of $\mathbf{W}_c(t_0, t_1)$ being positive definite. Now substitute (3.38) into (3.37):

$$\mathbf{x}(t_1) = \Phi(t_1, t_0)\mathbf{x}(t_0) - \int_{t_0}^{t_1} \Phi(t_1, \xi)\mathbf{B}(\xi)\mathbf{B}^T(\xi)\Phi^T(t_0, \xi)\, d\xi \cdot \mathbf{W}_c^{-1}(t_0, t_1)\mathbf{x}(t_0) \tag{3.39}$$

But recall that

$$\Phi(t_1, \xi) = \Phi(t_1, t_0)\,\Phi(t_0, \xi) \tag{3.40}$$

Then (3.39) becomes

$$\mathbf{x}(t_1) = \Phi(t_1, t_0)\,\mathbf{x}(t_0) - \Phi(t_1, t_0)\int_{t_0}^{t_1} \Phi(t_0, \xi)\mathbf{B}(\xi)\mathbf{B}^T(\xi)\Phi^T(t_0, \xi)\, d\xi \mathbf{W}_c^{-1}(t_0, t_1)\mathbf{x}(t_0) \tag{3.41}$$

Using (4.36), we obtain

$$\mathbf{x}(t_1) = \mathbf{0} \tag{3.42}$$

which proves that if $\mathbf{W}_c^{-1}(t_0, t_1)$ is positive definite, a control can be chosen and such a control drives the system to the origin.

We not want to prove necessity, and again we shall do so by contradiction. Assume that $\mathbf{W}_c(t_0, t_1)$ is not positive definite. That is.

$$\mathbf{x}^T(t_0)\mathbf{W}_c(t_0, t_1)\mathbf{x}(t_0) \leqq 0 \tag{3.43}$$

Let a control $\mathbf{u}^*(t)$ be defined as

$$\mathbf{u}^*(t) = -\mathbf{B}^T(t)\boldsymbol{\Phi}^T(t_0, t)\mathbf{x}(t_0) \tag{3.44}$$

Then

$$\mathbf{u}^T(t)\mathbf{u}^*(t) = \mathbf{x}^T(t_0)\boldsymbol{\Phi}(t_0, t)\mathbf{B}(t)\mathbf{B}^T(t)\boldsymbol{\Phi}^T(t_0, t)\mathbf{x}(t) \tag{3.45}$$

Integrate the above over t_0, t_1 and use the definition in (3.36):

$$\int_{t_0}^{t_1}\mathbf{u}^T(t)\mathbf{u}^*(t) = \mathbf{x}^T(t_0)\mathbf{W}_c(t_0, t_1)\mathbf{x}(t_0) \tag{3.46}$$

Replace this in assumption (3.43):

$$\int_{t_0}^{t_1}\mathbf{u}^T(t)\mathbf{u}^*(t)\ dt \leqq 0 \tag{3.47}$$

But the only possible result of (3.47) since $\mathbf{u}^T\mathbf{u}$ is always greater than, or equal to, zero is

$$\int_{t_0}^{t_1}\mathbf{u}^T(t)\mathbf{u}^*(t)\ dt > 0 \tag{3.48}$$

which implies

$$\mathbf{u}^*(t) = \mathbf{0} \tag{3.49}$$

But we assumed that the system is controllable and that there exists a $\mathbf{u}(t)$ such that

$$\boldsymbol{\Phi}(t_1, t_0)\mathbf{x}(t_0) = -\int_{t_0}^{t_1}\boldsymbol{\Phi}(t_1, \xi)\mathbf{B}(\xi)\mathbf{u}(\xi)\ d\xi \tag{3.50}$$

or, using the transition matrix property,

$$\mathbf{x}(t_0) = -\int_{t_0}^{t_1}\boldsymbol{\Phi}(t_0, \xi)\mathbf{B}(\xi)\mathbf{u}(\xi)\ d\xi \tag{3.51}$$

Now

$$\begin{aligned}\mathbf{x}^T(t_0)\mathbf{x}(t_0) &= -\mathbf{x}^T(t_0)\int_{t_0}^{t_1}\boldsymbol{\Phi}(t_0, \xi)\mathbf{B}(\xi)\mathbf{u}(\xi)\ d\xi \\ &= -\int_{t_0}^{t_1}\mathbf{x}^T(t_0)\boldsymbol{\Phi}(t_0, \xi)\mathbf{B}(\xi)\mathbf{u}(\xi)\,d\xi = -\int_{t_0}^{t_1}\mathbf{u}^T(\xi)\mathbf{u}(\xi)\ d\xi\end{aligned} \tag{3.52}$$

But $\mathbf{u}^*(\xi) = \mathbf{0}$ for all ξ. Therefore,

$$\mathbf{x}^T(t_0)\mathbf{x}(t_0) = 0 \tag{3.53}$$

or

$$\mathbf{x}(t_0) = \mathbf{0} \tag{3.54}$$

which contradicts our hypothesis, so that indeed it is necessary for $\mathbf{W}_c(t_0, t_1)$ to be positive definite. ∎

As with observability there is a simple criterion for checking controllability of time-invariant systems.

COROLLARY 3.2. Let $\mathbf{x}(t)$ be given by

$$\mathbf{x}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t)$$

where $\mathbf{A}$ is $n \times n$ and $\mathbf{B}$ is $n \times m$. The system is completely controllable if and only if the controllability matrix W_c

$$\mathbf{W}_c = [\mathbf{B} \quad \mathbf{AB} \quad \cdots \quad \mathbf{A}^{n-1}\mathbf{B}] \tag{3.55}$$

is of rank n.

Proof. See Polak and Wong (pp. 33–36).

Now as before a similar theorem holds for discrete-time systems and we will state it for completeness.

Theorem 3.4.

The discrete system is completely controllable if and only if the symmetric $n \times n$ matrix

$$\mathbf{W}_d(0, N) = \sum_{i=1}^{N} \Phi(0, i)\Gamma(i-1)\Gamma^T(i-1)\Phi^T(0, i) \tag{3.56}$$

is positive definite for some $N > 0$.

This then completes our discussion of controllability and observability. These concepts will become useful when we study the existence of solutions and their stability for the optimum filtering problem. Aside from that, the concepts by themselves provide great insight into what is necessary for a good model. Bucy and Joseph (pp. 39–42) discuss just this problem of stochastic modeling.

For those cases where an a priori model exists, we should always check our system against these concepts to see if we have a viable model.

2.4. STABILITY

The final topic that we want to discuss is that of stability. Again Kalman stands out as having been a major contributor and we shall reference our work to his (Kalman and Bertram [1, 2]). We shall discuss two points: first, the meaning of stability and, second, the application of the second method of Lyapanov to the determination of stability. We shall only discuss stability of discrete systems, since we use these results for determining the stability, sensitivity, and convergence of the estimation algorithms in Chapter 6.

We shall begin by presenting definitions of stability and then present two theorems and their proofs that employ the Lyapanov theory. Before presenting the theorems we need to discuss some of the basic notations that will be used.

DEFINITION 4.1. A positive definite matrix $\mathbf{A}$ is one such that

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i a_{ij} x_j > 0 \tag{4.1}$$

for all $\mathbf{x} \neq 0$.

DEFINITION 4.2. The euclidean norm $\|\mathbf{x}\|$ is given by

$$\|\mathbf{x}\| = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2} \tag{4.2}$$

DEFINITION 4.3. The generalized euclidean norm is given by

$$\|\mathbf{x}\|_{\mathbf{A}} = (\mathbf{x}^T\mathbf{A}\mathbf{x})^{1/2} \tag{4.3}$$

where $\mathbf{A}$ is a positive definite symmetric $n \times n$ matrix

LEMMA 4.1. For any $n \times 1$ vector, the following bounds on the euclidean norm hold:

$$\frac{\sum_{i=1}^{n} |x_i|}{\sqrt{n}} \leqq (\mathbf{x}^T\mathbf{x})^{1/2} \leqq \sum_{i=1}^{n} |x_i| \tag{4.4}$$

where x_i are the scalar components of the vector.

Proof. By definition, we have

$$(\mathbf{x}^T\mathbf{x}) = \sum_{i=1}^{n} x_i^2 \tag{4.5}$$

But

$$\left[\sum_{i=1}^{n} |x_i|\right]^2 = \sum_{i=1}^{n} |x_i|^2 + \sum_{i=1}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{n} |x_i|\,|x_j| \tag{4.6}$$

Since the other sums are all positive, it is obvious that

$$(\mathbf{x}^T\mathbf{x}) \leqq \left[\sum_{i=1}^{n} |x_i|\right]^2 \tag{4.7}$$

Now let us prove the lower bound of the inequality. Recall that

$$\left[\sum_{i=1}^{n} |x_i|\right]^2 = \sum_{i=1}^{n} |x_i|^2 + \sum_{i=1}^{n}\sum_{\substack{j=1 \\ i\neq j}}^{n} |x_i|\,|x_j| \tag{4.8}$$

Now define

$$M = \sum_{i=1}^{n} |x_i| \tag{4.9}$$

and define

$$y_i = \frac{|x_i|}{M} \tag{4.10}$$

Then divide both sides of the inequality by $1/M^2$. This yields

$$\frac{\sum_i \sum_j |x_i|\,|x_j|}{M^2} = 1 \tag{4.11}$$

and

$$\sum_i (x_i)^2 = \sum_i (y_i)^2 M^2 \tag{4.12}$$

Now the normalized variables are such that

$$\sum_{i=1}^{n} y_i = 1 \tag{4.13}$$

Now (4.13) is the equation for a plane in n dimensions intersecting the axis at the points (1,0,0,$\cdots$) (0, 1, $\cdots$, 0) and (0, 0, $\cdots$, 1). Also (4.12) defines the distance from the plane to the origin. The normal to this plane is the vector

$$\mathbf{n} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & n \end{bmatrix} \tag{4.14}$$

The minimum distance to the plane from the origin is then in this direction. This is shown in Figure 2.10. At the point of contact, we must satisfy the equation of the plane with all y_i equal. Thus, the minimum distance is

Figure 2.10 Example of inequality.

$$\frac{1}{n} = \min \sum_i (y_i)^2 \tag{4.15}$$

or, in general,

$$\frac{1}{n} \leqq \sum_i (y_i)^2 \tag{4.16}$$

By remultiplying by M^2 we obtain the left-hand side of the inequality. ∎

DEFINITION 4.4. The Euclidean norm of a matrix is

$$\|\mathbf{A}\|^2 = \max_{\mathbf{x}} \{\mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x};\ \mathbf{x}^T\mathbf{x} = 1\} \tag{4.17}$$

LEMMA 4.2. The following inequalities hold true:

$$\|\mathbf{AB}\| \leqq \|\mathbf{A}\| \cdot \|\mathbf{B}\| \tag{4.18}$$

$$\|\mathbf{A}\| \leqq \sum_{ij} |a_{ij}| \tag{4.19}$$

$$\|\mathbf{A}\| \leqq (\sum a^2_{ij})^{1/2} \tag{4.20}$$

$$\|\mathbf{A}\| \leqq \max_j (\sum_i |a_{ij}|) \tag{4.21}$$

The proof of these is a trivial extension of the above definition and the preceding lemma.

The next step is to define what we mean by stability. We shall present definitions of three different types of stability. In order to do so we must first define what we mean by an "equilibrium state."

DEFINITION 4.5. A state $\mathbf{x}_e$ of an undriven dynamic system whose state equation is

$$\mathbf{x}(k+1) = \mathbf{f}(\mathbf{x}(k), k) \tag{4.22}$$

is called an *equilibrium state* if

$$\mathbf{x}_e = \mathbf{x}(k); \qquad \forall k \tag{4.23}$$

or

$$\mathbf{x}_e = \mathbf{x}(k) = \mathbf{f}(\mathbf{x}(k), k); \qquad \forall k \tag{4.24}$$

We will find it useful to introduce another definition. Let us assume that we have a dynamical system governed by

$$\mathbf{x}(k+1) = \mathbf{f}(\mathbf{x}(k), k) \tag{4.25}$$

with an initial condition $\mathbf{x}(t_0)$. Here t_0 represents any arbitrary initial time. We shall then let

$$\mathbf{x}(k) = \phi(t_k; \mathbf{x}_0, t_0) \tag{4.26}$$

represent the state $\mathbf{x}$ at time k, given $\mathbf{x}(0) = \mathbf{x}_0$ at time t_0. Care should be

taken not to confuse this with the transition matrix. For a linear system, recall from Section 2.1 that

$$\phi(t_k; \mathbf{x}_0, t_0) = \Phi(k, 0)\mathbf{x}(0) \tag{4.27}$$

DEFINITION 4.6. An equilibrium state $\mathbf{x}_e$ of a dynamic system is *uniformly stable* if for any $\varepsilon > 0$ and for all t_0 there corresponds a number $\delta(\varepsilon) > 0$ such that if

$$\|\mathbf{x}_0 - \mathbf{x}_e\| \leqq \delta(\varepsilon) \tag{4.28}$$

then

$$\|\phi(t_k; \mathbf{x}(t_0), t_0) - \mathbf{x}_e\| \leqq \varepsilon; \qquad \forall t_k \geqq t_0 \tag{4.29}$$

This concept is shown in Figure 2.11. It should also be obvious that this concept of stability is analogous to that of continuity. This analogy will be further reinforced when we discuss uniform asymptotic stability in the large.

DEFINITION 4.7. An equilibrium state $\mathbf{x}_e$ of a dynamic system is *uniformly asymptotically stable* if

(1) it is uniformly stable; and (2) for all $\mu > 0$ and for all t_0 there exists a number $T(\mu)$ such that

$$\|\phi(t_k; \mathbf{x}_0, t) - \mathbf{x}_e\| \leqq \mu \tag{4.30}$$

Figure 2.11 Stable system.

for all

$$t_k \geqq t + T(\mu) \tag{4.31}$$

whenever

$$\|\mathbf{x}_0 - \mathbf{x}_e\| \leqq r; r > 0 \tag{4.32}$$

where r is some fixed constant that does not depend on $\mathbf{x}_0$ or μ.

Note here that asymptotic stability is a local concept in that we do not know how small r should be for this to hold. Also note that μ may be as small as possible. The concept of asymptotic stability is shown in Figure 2.12. We must now present an auxiliary definition that is needed in the final concept of stability.

Figure 2.12 Asymptotic stability.

DEFINITION 4.7. A motion is said to be uniformly bounded if for all t and for all $\hat{\delta} > 0$ there exists an $\varepsilon(\hat{\delta})$ such that

$$\|\boldsymbol{\phi}(t_k; \mathbf{x}, t) - \mathbf{x}_e\| < \varepsilon(\hat{\delta}); \qquad \forall t_k \geqq t \tag{4.33}$$

whenever

$$\|\mathbf{x} - \mathbf{x}_e\| < \hat{\delta} \tag{4.34}$$

We should immediately note how uniform boundedness differs from uniform stability. Uniform stability says that the system can be made arbitrarily

close to the equilibrium point if we start close enough to it. Uniform boundedness says that no matter how far initially we are from equilibrium, we can always find a value that will bound the solution at some future time from the equilibrium point. The reason for introducing uniform boundedness is that we now want to consider a stability concept that will insure that the solution will converge to its equilibrium value no matter how far from equilibrium we may initially be.

This then leads to the following definition.

DEFINITION 4.8. An equilibrium state $\mathbf{x}_e$ of a dynamic system is *uniformly asymptotically stable in the large* (u.a.s.i.l.) if (1) it is *uniformly stable*; (2) all motions are *uniformly bounded*; (3) all motions $\phi(t_k; \mathbf{x}_0, t_0)$, with $\mathbf{x}_0$ and t_0 being arbitrary, *converge uniformly* in $\|\mathbf{x}_0\| \leqq r$, with r being arbitrarily large, to $\mathbf{x}_e$ with increasing k.

This definition extends the previous definition to an arbitrary starting point. Obviously a necessary condition for u.a.s.i.l. is that there be a unique equilibrium state in the entire state space. In general, this is the most desirable form of stability.

We now want to prove two theorems relating to the stability of dynamical systems. The first theorem introduces the Lyapanov function and discusses the stability of nonlinear discrete-time systems. The second theorem takes a linear discrete-time system and shows a useful technique for obtaining a Lyapanov function and thus obtaining stability results.

THEOREM 4.1

Consider a discrete-time free dynamic system

$$\mathbf{x}(k+1) = \mathbf{f}(\mathbf{x}(k), k) \tag{4.35}$$

where

$$\mathbf{f}(\mathbf{0}, k) = \mathbf{0} \tag{4.36}$$

for all k. Suppose there exists a scalar function $V(\mathbf{x}, k)$ such that

$$V(\mathbf{0}, k) = 0 \tag{4.37}$$

for all k and that

1. $V(\mathbf{x}, k)$ is positive definite, that is, there exists a continuous nondecreasing scalar function α such that $\alpha(0) = 0$, and for all k and all $\mathbf{x}(k) \neq 0$,

$$0 < \alpha(\|\mathbf{x}\|) \leqq V(\mathbf{x}, k) \tag{4.38}$$

2. there exists a continuous scalar function γ such that $\gamma(0) = 0$, and for all k and $x \neq 0$,

$$\begin{aligned} &[V(\phi(k+1; \mathbf{x}(k), k), k+1) - V(\mathbf{x}(k), k)]T \\ &\triangleq \Delta V(\mathbf{x}(k), k) \leqq -\gamma(\|\mathbf{x}(k)\|) < 0 \end{aligned} \tag{4.39}$$

and

$$\gamma(\|\mathbf{x}(k)\|) > 0; \qquad \forall \mathbf{x}(k) \tag{4.40}$$

This describes the rate of increase of V along the path of motion starting at $\mathbf{x}(k)$. The sampling time T may be taken as unity:

3. there exists a continuous nondecreasing scalar function β such that $\beta(0) = 0$, and for all k and $\mathbf{x}(k) \neq 0$,

$$V(\mathbf{x}(k), k) \leqq \beta(\|\mathbf{x}\|) \tag{4.41}$$

4. $\alpha(\|\mathbf{x}\|) \to \infty \qquad$ as $\qquad \|\mathbf{x}(k)\| \to \infty$

Then the equilibrium state $\mathbf{x}_e(k) = 0$ is *u.a.s.i.l.* and $V(\mathbf{x}(k), k)$ *is a Lyapanov function* of the system.

Proof. Now by (4.40) we see that V is decreasing along any path of motion. We now wish to prove uniform stability. Consider any $\varepsilon > 0$. Take $\delta(\varepsilon) > 0$ such that

$$\beta(\delta) < \alpha(\varepsilon) \tag{4.42}$$

But recall that for any η

$$0 < \alpha(\eta) \leqq V(\eta) \leqq \beta(\eta) \tag{4.43}$$

Furthermore, recall that β was nondecreasing so that for (4.42) to hold $\varepsilon > \delta(\varepsilon)$. Thus, δ depends on ε. This relationship between (4.43) and (4.42) is shown in Figure 2.13. Then if $\|\mathbf{x}_0\| \leqq \delta$, where t_0 is arbitrary we have the following inequalities:

1. Since $\delta \geqq \|\mathbf{x}_0\|$ and $\beta(x)$ is nondecreasing.

Figure 2.13 Relationships between V, α, β.

$$\beta(\delta) \geqq \beta(\mathbf{x}_0) \geqq V(\mathbf{x}_0, t) \tag{4.44}$$

2. Since the rate of change of V is negative (4.40)

$$V(\mathbf{x}_0, t_0) > V(\boldsymbol{\phi}(t_k; \mathbf{x}_0, t_0), k) \tag{4.45}$$

3. Since V is bounded below by α, we have

$$V(\boldsymbol{\phi}(t_k; \mathbf{x}_0, t_0), k) \geqq \alpha(\|\boldsymbol{\phi}(k; \mathbf{x}_0, t_0)\|) \tag{4.46}$$

Now, using (4.42), (4.44), (4.45), and (4.46), we have

$$\alpha(\varepsilon) > \beta(\delta) \geqq V(\mathbf{x}, t) > V(\boldsymbol{\phi}(t_k; \mathbf{x}_0, t_0)) \geqq \alpha(\|\boldsymbol{\phi}(t_k; \mathbf{x}_0, t_0)\|) \tag{4.47}$$

But α is *nondecreasing* also and is *positive*. Thus,

$$\alpha(\varepsilon) > \alpha(\|\boldsymbol{\phi}(t_k; \mathbf{x}_0, t_0)\|) \tag{4.48}$$

which implies

$$\varepsilon \geqq \|\boldsymbol{\phi}(t_k; \mathbf{x}_0, t_0)\| \tag{4.49}$$

for any $\mathbf{x}_0$, satisfying

$$\delta(\varepsilon) \geqq \|\mathbf{x}(0)\| \tag{4.50}$$

But this is nothing more than the conditions for a uniformly stable system. Note also that $\mathbf{x}_e$, the equilibrium state is the origin.

We now wish to prove a.s.i.l. This can be done if we can show that $\|\boldsymbol{\phi}(t_k; \mathbf{x}(0), t_0)\| \to 0$ as k gets very large and for any $\|\mathbf{x}(0)\| \leqq r$.

Take any positive constant C_1 and find $r > 0$ satisfying

$$\beta(r) < \alpha(C_1) \tag{4.51}$$

Now choose an initial state $\|\mathbf{x}_0\| \leqq r$. Then from the first part of the proof,

$$\|\boldsymbol{\phi}(t_k; \mathbf{x}(0), t_0)\| \leqq C_1 \tag{4.52}$$

for all $k \geqq 0$. Now choose any μ such that

$$0 < \mu \leqq \|\mathbf{x}_0\| \tag{4.53}$$

Find an $\upsilon(\mu) > 0$ such that

$$\beta(\mu) < \alpha(\upsilon) \tag{4.54}$$

This is shown in Figure 2.14. Denote by $C_2(\mu, r) > 0$ the minimum of the continuous function $\gamma(\|\mathbf{x}\|)$ on the set

$$S = \{\mathbf{x}: \upsilon(\mu) \leqq \|\mathbf{x}\| \leqq C_1(r)\} \tag{4.55}$$

Define

$$T(\mu, r) = \frac{\beta(r)}{C_2(\mu, r)} > 0 \tag{4.56}$$

Now suppose that

Figure 2.14 Values for a.s.i.l.

$$\|\phi(t_k; \mathbf{x}(0), t_0)\| > \nu \tag{4.57}$$

over the interval $[t_0, t_1]$, $t_1 = t_0 + T$. By the argument presented in the first part of the proof,

$$0 < \alpha(\nu) \leqq V(\phi(t_k; \mathbf{x}(0), t_0)) \leqq V(\mathbf{x}_0, t_0) - TC_2 \leqq \beta(r) - TC_2 = 0 \tag{4.58}$$

but this implies that

$$0 < V(\phi(t_k, \mathbf{x}(0), t_0) - V(\mathbf{x}_0, t_0) \leqq -TC_2 \tag{4.59}$$

which is a contradiction. This implies that for some $t \in [t_0, t_1]$, say $t = j$, we have

$$\|\mathbf{x}(j)\| = \|\phi(k; \mathbf{x}(t_0), t)\| = \nu \tag{4.60}$$

Therefore,

$$\alpha(\|\phi(k; \mathbf{x}(j))\|) \leqq V(\phi(t_k; \mathbf{x}(j); j)) \leqq V(\mathbf{x}(j), j) \leqq \beta(\nu) < \alpha(\mu) \tag{4.61}$$

Therefore,

$$\|\phi(k; \mathbf{x}(t_0), t_0)\| < \mu \tag{4.62}$$

for all $t > t + T(\mu, r) \geqq t_2$, which proves asymptotic stability. To prove u.a. s.i.l. We observe that for any r a constant $C_1(r)$ exists such that $\beta(r) < \alpha(C_1)$ and furthermore for all t_0. ∎

We have just shown that if there exists a Lyapanov function then we can determine the stability of a discrete-time nonlinear system. What we shall now do is to assume that we have a linear discrete-time system. This added structure allows us to say much more about stability. Notably, we can actually provide one with the specific form of the Lyapanov function from which

$$\dot{\mathbf{x}}(t) = \mathbf{A}(t)\,\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \tag{2.36}$$

where $\mathbf{u}(t)$, a $k \times 1$ vector, has the form

$$\mathbf{u}(t) = \mathbf{u}(kT);\ kT \leqq t < (k+1)T \tag{2.37}$$

where $u(k)$ is constant. The interval T is the sample interval. This form implies that $u(t)$ is constant over the interval $[kT, (k+1)T)$. A sample function is shown in Figure 2.6.

Figure 2.6 Sample function for discretized control.

With this for a forcing function we can write out the state at time $(k+1)T$ as

$$\mathbf{x}((k+1)T) = \boldsymbol{\Phi}((k+1)T, kT)\,\mathbf{x}(kT) + \int_{kT}^{(k+1)T} \boldsymbol{\Phi}(t, \xi)\,d\xi\ \mathbf{u}(kT) \tag{2.38}$$

Define the function $\Gamma(kT)$ as

$$\Gamma(kT) = \int_{kT}^{(k+1)T} \boldsymbol{\Phi}(t, \xi)\,d\xi \tag{2.39}$$

Now it is convenient to suppress the dependence on T to denote the state equation as

$$\mathbf{x}(k+1) = \boldsymbol{\Phi}(k+1, k)\,\mathbf{x}(k) + \Gamma(k)\mathbf{u}(k) \tag{2.40}$$

It should be pointed out that although $\boldsymbol{\Phi}(k+1, k)$ has an inverse if it comes from a continuous-time system, for an arbitrary discrete-time system this need not be necessary. Thus, not all discrete-time systems have continuous-time realizations.

In similar fashion, the measurement equation can be written as

stability can be ascertained. It will be found to be a quadratic form involving the transition matrix and the state variables. The theorem will involve the proof of equivalency of the five statements made in it. It should be realized at the start that the generation of Lypanov functions may not be obvious. For here one requires a very structured function, which is not frequently obtained by mere observation.

THEOREM 4.2.

Consider the discrete-time linear dynamic system

$$\mathbf{x}(k+1) = \Phi(k+1, k)\mathbf{x}(k) + \mathbf{B}(k)\mathbf{u}(k) \tag{4.63}$$

and assume that

$$\frac{\|\Phi(k+1, k) - \mathbf{I}\|}{T} \leqq C_1 < \infty \tag{4.64}$$

for all k and

$$0 < C_2 \leqq \frac{\|\mathbf{B}(k)\mathbf{x}\|}{T} \leqq C_3 < \infty \tag{4.65}$$

for all $\|\mathbf{x}\| = 1$ and all k.

Then the following propositions concerning the system are equivalent:

(a) Assuming $\mathbf{x}(0) = \mathbf{0}$ any *uniformly bounded excitation*

$$\|\mathbf{u}(k)\| \leqq C_4 < \infty \qquad ; k \geqq 0 \tag{4.66}$$

gives rise to a *uniformly bounded response* for all $k > 0$; that is,

$$\|\mathbf{x}(k)\| = \left\|\sum_{i=0}^{k-1} \Phi(k, i+1)\,\mathbf{B}(i)\mathbf{u}(i)\right\| \leqq C_5(C_4) < \infty \tag{4.67}$$

(b) for all $k > 0$.

$$\sum_{i=0}^{k} \|\Phi(k, i)\| \leqq C_6 < \infty \tag{4.68}$$

(c) The *equilibrium state* $\mathbf{x}_e = \mathbf{0}$ of the free system is *uniformly asymptotically stable*.

(d) There exist positive constants C_7, C_8 such that whenever $k > 0$,

$$\|\Phi(k, 0)\| \leqq C_7 e^{-C_8 kT} \tag{4.69}$$

(e) Given any positive definite matrix $\mathbf{Q}(k)$ satisfying for all $k > 0$.

$$0 < C_9\mathbf{I} \leqq \mathbf{Q}(k) \leqq C_{10}\mathbf{I} < \infty \tag{4.70}$$

the scalar function defined by

$$V(\mathbf{x}(k), k) = \sum_{i=k}^{\infty} \mathbf{x}^T(k)\Phi^T(i, k)\mathbf{Q}(i)\Phi(i, k)\mathbf{x}(k) = \mathbf{x}^T(k)\mathbf{P}(k)\mathbf{x}(k) \tag{4.71}$$

exists and is a *Lyapanov function* of the free system satisfying the requirements of the previous theorem with

$$\Delta V(\mathbf{x}, k) = -\mathbf{x}^T(k)\mathbf{Q}(k)\mathbf{x}(k) \tag{4.72}$$

The proof of this theorem is quite long and will be performed in the following fashion. First we shall show that (a) implies (b), then (b) implies (c), which implies (d). Then we will show (d) implies (a) and (e). Finally, we will show (e) implies (c). The proof is for the discrete version of the theorem originally found in Kalman and Bertram [2] and follows the proof of the continuous version of Kalman and Bertram [1].

Proof. We shall prove this by contradiction, by showing that unless (a) implies (b) we will always be capable of finding a control which will lead to a contradiction. Assume that for some j, l pair

$$\sum_{i=0}^{k} \Phi_{jl}(k, i) \to 0 \tag{4.73}$$

where $\Phi_{jl}(k, i)$ is the jl th component of $\Phi(k, i)$. Now from (a) we know that $\|\mathbf{x}(k)\|$ is bounded for all bounded $\mathbf{u}(k)$. This implies that

$$\frac{1}{n^{1/2}} \sum_{i=1}^{n} |x_i(k)| \leqq \|\mathbf{x}(k)\| < \infty \tag{4.74}$$

or

$$\|\mathbf{x}(k)\| \geqq \frac{1}{n^{1/2}} |x_j(k)|; \qquad \text{for any } j \tag{4.75}$$

But $|x_j(k)|$ can be lower bounded by

$$\begin{aligned} |x_j(k)| &= \Big|\sum_{r=1}^{n} \sum_{s=1}^{m} \Big(\sum_{i=0}^{k} \Phi_{jr}(k, i)\, B_{r,s}(i) u_s(i)\Big)\Big| \\ &\geqq \frac{1}{n^{1/2}} \Big|\sum_{s=1}^{m} \Big(\sum_{i=0}^{k} \Phi_{jl}(k, i) B_{l,s}(i) u_s(i)\Big| \end{aligned} \tag{4.76}$$

which follows from Lemma 4.2. Now since (a) holds for all bounded $\mathbf{u}(k)$ choose the following $\mathbf{u}(k)$:

$$u_s(i) = |\mathbf{B}^{-1}(i)|_{ls}\ \operatorname{sgn}\Phi_{jl}(i) \tag{4.77}$$

where $|\mathbf{B}^{-1}(i)|_{ls}$ represents the lsth element of $\mathbf{B}^{-1}(i)$ and "sgn" is the sign function. Then we can interchange summations to show

$$\begin{aligned} &\Big|\sum_{s=1}^{m} \Big(\sum_{i=0}^{k} \Phi_{jl}(k, i) B_{ls}(i)\, u_s(i)\Big)\Big| \\ &= \Big|\sum_{i=0}^{k} \Phi_{jl}(k, i) \operatorname{sgn}\Phi_{jl}(k, i)\Big| \\ &= \Big|\sum_{i=0}^{k} \Phi_{jl}(k, i)\Big| \end{aligned} \tag{4.78}$$

But this implies that

$$\infty > \|\mathbf{x}(k)\| > \frac{1}{n} \sum_{i=0}^{k} |\Phi_{ji}(k, i)| \tag{4.79}$$

which by assumption is unbounded below, thus a contradiction, and so, (a) must imply (b).

(b) → (c): We now want to show u.a.s., which means that the system is uniformly stable and there will exist a μ such that for all $T(\mu)$, such that $t_k > t_0 + T(\mu)$.

$$\|\phi(t_k; \mathbf{x}_0, t_0)\| < \mu \tag{4.80}$$

whenever $\|\mathbf{x}_0\| < r$ for some r. Now, (4.64) of the theorem, we have

$$\|\Phi(k+1, k) - \mathbf{I}\| < C_1 < \infty \qquad ;\forall k \tag{4.81}$$

We now want to show that for some x(0)

$$\|\Phi(k, 0)\mathbf{x}(0)\| < \mu \tag{4.82}$$

Now(b) implies

$$\sum_{i=0}^{k} \|\Phi(k, i)\| < C_6 \tag{4.83}$$

Using (b) and (4. 64), we obtain

$$\begin{aligned} C_6 C_1 &> \sum_{i=0}^{k} \|\Phi(k, i)\| \ \|\Phi(i, i-1) - \mathbf{I}\| \\ &\geqq \sum_{i=0}^{k} \|\Phi(k, i)\Phi(i, i-1) - \Phi(k, i)\| \\ &\geqq \left\| \sum_{i=0}^{k} (\Phi(k, i-1) - \Phi(k, i)) \right\| \end{aligned} \tag{4.84}$$

But it is easily shown that

$$\sum_{i=0}^{k} (\Phi(k, i-1) - \Phi(k, i)) = \Phi(k, 0) - \mathbf{I} \tag{4.85}$$

Thus, this implies that

$$\|\Phi(k, 0) - \mathbf{I}\| < C_6 C_7 \tag{4.86}$$

Now we also have

$$\|\Phi(k, 0) - \mathbf{I}\| \geqq \|\Phi(k, 0)\| - \|\mathbf{I}\| \tag{4.87}$$

So that

$$\|\Phi(k, 0)\| \leqq \|\mathbf{I}\| + C_6 C_7 < C_{11} < \infty \tag{4.88}$$

Furthermore, using (b) and the above bound, we have

$$\begin{aligned} C_6 C_{11} &> \sum_{i=0}^{k} \|\Phi(k, i)\| \ \|\Phi(i, 0)\| \\ &> \sum_{i=0}^{k} \|\Phi(k, 0)\| = k \ \|\Phi(k, 0)\| \end{aligned} \tag{4.89}$$

Thus, we have for some k

$$\|\Phi(k, 0)\| < \frac{C_6 C_{11}}{k} \tag{4.90}$$

Now if we choose $T(\mu)$ as

$$T(\mu) = \frac{C_6 C_{11} r}{\mu} \tag{4.91}$$

where $\|\mathbf{x}_0\| < r$ then we have

$$\|\Phi(k, 0)\mathbf{x}(0)\| \leqq \|\Phi(k, 0)\| \ \|\mathbf{x}_0\| < \frac{C_6 C_{11} r}{k} \tag{4.92}$$

Thus,

$$\|\mathbf{x}(k)\| < \frac{C_6 C_{11} r}{k} < \mu \tag{4.93}$$

for all $T(\mu) < k$ which proves u.a.s.

(c) → (d) Now u.a.s. implies that

$$\|\phi(t_k; \mathbf{x}_0, t_0)\| < \mu ; \qquad \forall k > T(\mu) \tag{4.94}$$

Now choose $T(\mu)$ such that

$$\|\phi(l; \mathbf{x}_0, 0)\| < \tfrac{1}{2} \tag{4.95}$$

where time l is equal to $T(\mu)$ and such that $k = nl + m$ where l, n, and m are integers. Clearly n represents the number of l multiples in k and m is the remainder. Since the system is also uniformly stable, we know that

$$\|\Phi(q; \mathbf{x}_0, 0)\| < \frac{C_7}{2} \tag{4.96}$$

for all q and that the restriction on $t_0 = 0$ is arbitrary because of the uniform nature of the stability. Thus we have

$$\begin{aligned} \|\phi(k; \mathbf{x}_0, 0)\| &= \|\Phi(k, k - m)\Phi(k - m - l, k - m - 2l) \\ &\quad \cdots \Phi(k - m - (n - 1)l, 1)\Phi(1, 0)\mathbf{x}(0)\| \\ &\leqq \|\Phi(k, k - m)\| \ \|\Phi(k - m - l, k - m - 2l)\| \\ &\quad \cdots \|\mathbf{x}(0)\| \end{aligned} \tag{4.97}$$

But since the system is u.a.s.

$$\|\Phi(k - m - pl, k - m - (p - 1)l)\| < \tfrac{1}{2} \tag{4.98}$$

and

$$\|\Phi(k, k - m)\| < \frac{C_7}{r^2} \tag{4.99}$$

Thus,

$$\|\phi(k; \mathbf{x}_0, 0)\| < \frac{C_7}{r}\left(\frac{1}{2}\right)^{n-1} \|\mathbf{x}_0\| \tag{4.100}$$

Let

$$C_8 = -\frac{\log 2}{k} n + 1 \tag{4.101}$$

Then

$$\|\phi(k; \mathbf{x}_0, 0)\| < C_7 \exp(-kC_8) \tag{4.102}$$

which proves the contention.

(d) → (a): Now from part (d) we have

$$\|\Phi(k, 0)\| < C_7 \exp(-C_8 k) \tag{4.103}$$

The output to a uniformly bounded excitation is

$$\sum_{i=0}^{k} \Phi(k, i)\mathbf{B}(i)\mathbf{u}(i) \tag{4.104}$$

But

$$\left\| \sum_{i=0}^{k} \Phi(k, i)\mathbf{B}(i)\mathbf{u}(i) \right\| \leqq \sum_{i=0}^{k} \|\Phi(k, i)\| \ \|\mathbf{B}(i)\| \ \|\mathbf{u}(i)\| \tag{4.105}$$

But by hypothesis

$$\|\mathbf{u}(i)\| < C_4 \tag{4.106}$$

$$\|\mathbf{B}(i)\| < C_3 \tag{4.107}$$

Therefore,

$$\begin{aligned} &\left\| \sum_{i=0}^{k} \Phi(k, i)\mathbf{B}(i)\mathbf{u}(i) \right\| \\ &\quad < \sum_{i=0}^{k} \|\Phi(k, i)\| C_4 C_3 \\ &\quad < C_4 C_3 C_7 \sum_{i=0}^{k} \exp(-(k - i)C_8) \\ &\quad \leqq C_4 C_3 C_7 \sum_{i=0}^{k} \exp(-iC_8) \\ &\quad \leqq C_4 C_3 C_7 \sum_{i=0}^{k} [\exp(-C_8)]^i \end{aligned} \tag{4.108}$$

But $\exp(-C_8) < 1$ since $C_8 > 0$, so that

$$\begin{aligned} &\left\| \sum_{i=0}^{k} \Phi(k, i)\mathbf{B}(i)\mathbf{u}(i) \right\| \\ &\qquad < C_3 C_4 C_7 \frac{1}{1 - e^{-C_8}} < C_{15} \end{aligned} \tag{4.109}$$

which proves the contention.

(e) → (c): If $V(\mathbf{x}(k), k)$ is a Lyapanov function, then from the previous theorem the system is u.a.s.i.l., which implies u.a.s.

(d) → (e): From part (d) we know that

$$\|\Phi(k, 0)\| < C_7 \exp(-C_8 k) \tag{4.110}$$

Using this we must now prove that

$$V(\mathbf{x}(k), k) = \sum_{i=k}^{\infty} \mathbf{x}^T(k)\, \Phi^T(i, k)\mathbf{Q}(i)\Phi(i, k)\mathbf{x}(k) \tag{4.111}$$

is a Lyapanov function. Clearly $V(\mathbf{x}(k), k)$ is positive definite since $\mathbf{Q}(k)$ is positive definite for all k. Now

$$\begin{aligned} V(\mathbf{x}(k), k) &\leqq \sum_{i=k}^{\infty} \|\mathbf{Q}(i)\| \; \|\Phi(i, k)\mathbf{x}(k)\|^2 \\ &< C_{10} \sum_{i=k}^{\infty} \|\Phi(i, k)\mathbf{x}(k)\|^2 \\ &< C_{10} \sum_{i=k}^{\infty} \|\Phi(i, k)\|^2 \, \|\mathbf{x}(k)\|^2 \\ &< C_{10} C_7^2 \, \|\mathbf{x}(k)\|^2 \sum_{j=0}^{\infty} [\exp(-2C_8)]^j \\ &< C_{15} \|\mathbf{x}\|^2 = \beta(\|\mathbf{x}\|) \end{aligned} \tag{4.112}$$

since C_8 is > 0 and e^{-2C_8} is less than one. Similarly, we can lower bound the function by

$$\begin{aligned} V(\mathbf{x}(k), k) &> \sum_{i=k}^{\infty} C_9 \|\Phi(i, k)\, \mathbf{x}(k)\|^2 \\ &> \frac{C_9}{C_1^2} \sum_{i=k}^{\infty} \|\Phi(i+1, i) - \mathbf{I}\|^2 \, \|\Phi(i, k)\mathbf{x}(k)\|^2 \end{aligned} \tag{4.113}$$

where we have used (4.64) in the bound. Now we can further lower bound it by

$$\mathbf{V}(\mathbf{x}(k), k) > \frac{C_9}{C_1^2} \Big\| \sum_{i=k}^{\infty} [\Phi(i+1, k) - \Phi(i, k)]\mathbf{x}(k) \Big\|^2 \tag{4.114}$$

But it is easily shown that

$$\sum_{i=k}^{\infty} [\Phi(i+1, k) - \Phi(i, k)] = -\Phi(k, k) = -\mathbf{I} \tag{4.115}$$

Thus,

$$V(\mathbf{x}(k), k) > \frac{C_9}{C_1^2} \|\mathbf{x}(k)\|^2 = \alpha(\|\mathbf{x}\|) \tag{4.116}$$

Thus, $V(\mathbf{x}(k), k)$ is properly upper and lower bounded. Finally,

$$\Delta V(\mathbf{x}(k), k) = -\mathbf{x}^T(k)\mathbf{Q}(k)\mathbf{x}(k) \leq -C_9\|\mathbf{x}(k)\|^2 = \gamma(\|\mathbf{x}\|) \tag{4.117}$$

which shows that $V(\mathbf{x}(k), k)$ is a Lyapanov function, which completes the proof of the theorem. ∎

The relationships that this theorem provides are indispensable with regards to the analysis of the stability of linear time-varying discrete-time systems. We shall use these results in Chapter 6 when we analyze the stability of the optimum estimate equation. More general discussions of Lyapanov stability theory are in LaSalle and Lefschetz and in Ogata. These techniques are used widely for the analysis of a variety of stability problems as discussed in Kalman and Bertram [1, 2].

We have completed all that is necessary for an understanding of the deterministic model. Yet there are many more theorems and techniques that are available and extend the basic concepts introduced here. Yet if one can grasp the general nature of what has been presented, then a reasonable understanding of the theory will be obtained. There will be questions posed later where we will have to relate to the concepts developed in this chapter. For example, the convergence of filters will become a stability problem whose answer is easily obtained by the Lyapanov method. The concepts of observability and controllability will be essential when we do modeling. If the model is not controllable, then some state is deterministic. Thus, this chapter has presented a great deal of information that will be expanded upon later.

2.5 CONCLUSIONS

The state-space approach to the analysis of dynamic systems provides a very useful medium for the discussion of stochastic systems. Thus, the results reviewed in the chapter should be considered a brief refresher of those concepts from deterministic systems analysis that will be useful in the development of a stochastic model. Many of the concepts contained in a deterministic analysis have analogues in the stochastic format; however, as we shall see, there are stark contrasts that exist also. In this section we shall review the results of this chapter and put them in a context which will be useful for the development of the stochastic model in the following chapters.

The first topic discussed in this chapter was the structure of the state-space description of nonlinear systems. These multidimensional systems were first presented in their most general form with the definition of the state of the system being presented. The assumption made throughout this text it that this state description of a dynamical system is known or can be obtained. Careful study leads one to believe that most physical systems have state descriptions that can be obtained from Newton's laws of motion or from Maxwell's equations or from some other set of well-defined physical laws. The

most important subclass of dynamical systems described in a state-space form are those whose dynamics are linear. Some examples such as simple harmonic motion, linear electrical circuits should indicate the wealth of useful systems that fall into this category. These systems are also useful because they lend themselves to careful analysis, particularly the linear time-invariant dynamic system. For this reason linearization techniques are most useful. The linearization techniques we discussed considered the system solution about an a priori trajectory: that is, we considered the perturbed solution. Using Taylor-series techniques these nonlinear systems can be reduced to linear ones. This technique will be used again when we obtain the filtering equations for nonlinear estimators.

The last topic discussed in the first section discussed the effect of driving or forcing functions on dynamical systems. In the next chapter we shall introduce random forcing functions to generate stochastic systems. These will rely heavily upon the deterministic structure.

The transition matrix is basically the Green's function for a linear time-varying (or -invariant) dynamic system. It projects the effect of the initial state into the present and that of the continuum of inputs into the current state of the system. The solution of the differential equation in terms of the transition matrix can be considered as an integral equation for the operator $\Phi(t, t_0)$. This interpretation has provided some insight into the solutions to cartain integral equations (see Baggeroer [1]). In Chapter 5 we will use the transition matrix extensively in the analysis of the structure of stochastic systems, particularly their second moment properties.

For the linear time-invariant case the transition matrix takes on a particularly simple form. The evaluation of this matrix is simple in many cases. and for more difficult ones there are abundant algorithms now available for this purpose. The adjoint system was also presented, and its use will be elaborated on in Chapter 6 when solutions to estimators are obtained.

In the second section we also introduced the concept of a discrete-time system. Since all computation for estimators is performed on a digital computer, the systems we are investigating should actually be phased in the discrete-time structure. However, for rapid enough sampling intervals the continuous-time system is necessary. We should also note that the discrete-time systems are analogues of the linear time-varient dynamic systems of the form

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t)$$

where the discrete analogue is

$$\mathbf{x}(k+1) = \Phi(k+1, k)\mathbf{x}(k) + \mathbf{B}(k)\mathbf{u}(k)$$

The structure of $\Phi(k+1, k)$ is determined directly from the transition matrix of the continuous-time system. The terminology for the forcing func-

tion part may vary depending on how $\mathbf{u}(t)$ is sampled. The reader is referred to Brockett or Ogata for a comparison.

The second important issue developed in this section was that of a measurement model. The system or state equation may be considered as the inner structure of a black-box system where the measurement equation tells us what we can observe about this system. For example, the system may be a highly complex electrical circuit and we measure only certain node voltages. The measurement concept is important because in almost every system we have access to only a subset of the states. Understanding the relationship of these measurements to the state is essential if we hope to obtain information concerning the states themselves.

As with the system equation, the measurement equation can be expressed in a highly complex nonlinear form. However, the simpler linear or linearized continuous-time version is most useful, namely,

$$\mathbf{z}(t) = \mathbf{C}(t)\mathbf{x}(t)$$

Likewise the discrete-time analogue is important and will be used for both deterministic analysis and stochastic analysis, namely,

$$\mathbf{z}(k+1) = \mathbf{C}(k+1)\mathbf{x}(k+1)$$

Note that as with the discrete-time state model, a sampling interval is assumed.

The second set of topics in this section refer to two important properties of linear time-varying dynamic systems: controllability and observability. Both of these topics, first introduced in the systems context by Kalman, have analogues in stochastic systems. For example, if a deterministic system is not observable, it means that by observing the output alone we cannot reconstruct the state of system. In contrast we shall see that a deterministically nonobservable system may be stochastically observable as a result of correlation properties.

Similar remarks concerning controllability can be made. In Chapter 7 we shall discuss these relationships between deterministic and stochastic controllability and observability. Also, these two concepts can be applied to nonlinear systems but require an analysis beyond the scope of the present discussion. The concept of observability and controllability for nonlinear systems is much more involved. The work by Kou, Elliot, and Tarn discusses some of these issues.

The final section introduced the concept of stability and the Lyapanov function. This topic completes the review of the deterministic theory that will be used in the analysis of stochastic systems. Viewed from the position of a deterministic model, the study of stability allows us to determine the conditions under which our system will be stable. It furthermore delineates the

various forms of stability. For stochastic systems the stability theory developed and reviewed in this chapter will be used in determining the stability and divergence of stochastic estimators in Appendix C (see Price: Deyst and Price; or Bucy [2]). Furthermore, studies of the stability of stochastic systems have been made by Kushner [5] and Wonham [5].

In Appendix A we develop one further point that is an immediate extension of this chapter, that of existence and uniqueness of equations of the form

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), t)$$

That is, we consider what properties of $\mathbf{f}(\ )$ are sufficient to guarantee that there exists at least one solution to this equation and, further, that this is the only solution. This issue arises again in the stochastic model of Chapter 3.

The structures developed in this chapter will be used extensively throughout the remainder of the book. This chapter thus represents a review, albeit a very succinct one, of the basics necessary to develop a theory of stochastic state estimation.

2.6 PROBLEMS

2.1. Write a state variable representation for the following equations:

(a) $\dddot{x} + a\ddot{x} + bx = u$

(b) $\dddot{x} + a\ddot{x} + bx = u + c\dot{u}$

(c) $\dddot{x} + a\ddot{x} + b\dot{x} + cx = u$

2.2. Consider a series $R - L - C$ circuit with a voltage source $u(t)$connected in series also.

(a) Write a differential equation for this system.

(b) Write a state-space representation for this system.

2.3.* Which of the following functions are Lipschitz?

(a) x^2

(b) $x^{1/2}$

(c) $\log x$

(d) $\sin x$

(e) $\sinh x$

2.4.* Prove Theorem A.2.

2.5. Find the transition matrix for the following systems:

(a) $\dddot{x} + 10\ddot{x} + 3x = u(t)$

(b) $\dddot{x} + 3\ddot{x} + 5\dot{x} + x = u(t)$

(c) $\dddot{x} + 3\ddot{x} + 2x = u(t)$

2.6. Let the $R - L - C$ circuit discussed in Problem 2.2 have values $R = 1$, $L = 2$, $C = 2$. Let $u(t)$ be given by

$$u(t) = \begin{cases} 1 & t \geqq 0 \\ 0 & t < 0 \end{cases}$$

(a) Find the transition matrix for this system.
(b) Find the loop current as a function of time.

2.7. Consider the percent distillate on the ith plate of a distillation column to be given by $x_i(t)$. The time rate of change of distillate on the ith plate is given by

$$\frac{dx_i(t)}{dt} = -\alpha_{ii}x_i(t) + \beta_{i,i-1}x_{i-1}(t) + \gamma_{i,i+1}x_{i+1}(t)$$

where α_{ii} is the boil-off rate from the ith plate, $\beta_{i,i-1}$ is the vaporization rate from the $i-1$ plate and $\gamma_{i,i+1}$ is the liquid drop rate from the $i+1$ plate. There are a total of N plates, and at the top(Nth) plate the input is $u_0(t)$, while at the bottom (first plate) the output is $u_1(t)$. Measurements are made by sampling the percent distillate on each plate and are given by

$$z_i(t) = \delta_i x_i(t)$$

Write a state variable model for this system.

2.8. Let a dynamical system be given by

*Problems marked by an asterisk depend on Appendix A, p. 683.

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t)$$

(a) Let $\mathbf{u}(t)$ be given by

$$\mathbf{u}(t) = \sum_{i=0}^{\infty} \mathbf{u}_i\,\delta(t - iT)$$

where δ is the delta function. Find a discrete version of $\mathbf{x}(t)$ in the form

$$\mathbf{x}(k+1) = \Phi(k+1, k)\mathbf{x}(k) + \mathbf{B}(k)\mathbf{u}(k)$$

(b) Let $\mathbf{u}(t)$ be given by

$$\mathbf{u}(t) = \sum_{i=0}^{\infty} \mathbf{w}_i u_{-1}(t - iT)$$

where $u_{-1}(t)$ is the unit step function. Obtain a discrete-time version of $\mathbf{x}(t)$ as in (a).

2.9. Consider the discrete-time system given by

$$\mathbf{x}(k+1) = \Phi(k+1, k)\,\mathbf{x}(k) + \mathbf{B}(k)\mathbf{u}(k)$$
$$\mathbf{z}(k+1) = \mathbf{C}(k+1)\mathbf{x}(k+1)$$

Let $\mathbf{x}(n)$ be the state at time n. Let there be N measurements made from time $n-N, n$.

(a) Let $\mathbf{M}_s(n, N)$ be the observability matrix of state $\mathbf{x}(n)$ given N measurements. Show that $\mathbf{M}_s(n, N)$ is given by

$$\mathbf{M}_s(n, N) = \sum_{i=n}^{n-N} \Phi(i, n)\mathbf{C}^T(i)\mathbf{C}(i)\Phi(i, n)$$

(b) Let $\mathbf{W}_s(n, N)$ be the controllability matrix of the above problem. Show that

$$\mathbf{W}_s(n, N) = \sum_{i=n}^{n-N} \Phi(n, i+1)\mathbf{B}(i)\mathbf{B}^T(i)\Phi^T(n, i+1)$$

2.10. Prove Corollary 3.1.

2.11. Prove Theorem 3.2.

2.12. Prove Corollary 3.2.

2.13. Prove Theorem 3.4.

2.14. Let $\mathbf{M}_c(t_0, t_1)$ be the continuous-time observability matrix.

(a) Show that

$$\frac{d}{dt}\mathbf{M}_c(t, t_1) = -\mathbf{A}^T(t)\mathbf{M}_c(t, t_1) - \mathbf{M}_c(t, t_1)\mathbf{A}(t) - \mathbf{C}^T(t)\mathbf{C}(t)$$

with $\mathbf{M}(t_1, t_1) = 0$.

(b) Find

$$\frac{d}{dt}\mathbf{M}_c^{-1}(t, t_1)$$

2.15. Let $\mathbf{W}_c(t_0, t_1)$ be the continuous-time observability matrix.

(a) Show that

$$\frac{d}{dt}\mathbf{W}(t, t_1) = \mathbf{A}(t)\mathbf{W}(t, t_1) + \mathbf{W}(t, t_1)\mathbf{A}^T(t) - \mathbf{B}(t)\mathbf{B}^T(t)$$

(b) Find

$$\frac{d}{dt}\mathbf{W}_c^{-1}(t, t_1)$$

2.16. Consider the system

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix} = \begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix} + \begin{bmatrix}a\\ b\end{bmatrix}u$$

For what values of a, b is the system controllable?

2.17. Consider the system:

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix} = \begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}$$

$$y = [a \quad b]\begin{bmatrix}x_1\\ x_2\end{bmatrix}$$

For what values of a,b is the system observable?

2.18. Let a continuous-time system be given by

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}(t), t)$$

where $\mathbf{f}(\mathbf{0}, t) = \mathbf{0}$. Develop a continuous-time analogue to Theorem 4.1.

2.19. Let the continuous-time linear system be given by

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t)$$

Develop the continuous-time analogue for Theorem 4.2.

2.20. Consider a series R - L - C circuit where the inductor is linear but where the voltage across the resistor is $f(i)$, where i is the loop current, and the voltage across the capacitor is $g(q)$, where q is the charge. Let

$$V(\mathbf{x}) = \tfrac{1}{2} Lx_2^2 + \int_0^{x_1} g(x')\, dx'$$

where $x_1 = q$ and $x_2 = i$.

(a) Write a state variable representation for this network.

(b) Show that $V(x)$ is a Lyapanov function if and only if $f(x) > 0$ for all $x \neq 0$.

(c) Comment on the physical meaning of $V(x)$ in terms of the stored energy in the circuit.

2.21. Let the linear continuous-time time-invariant system be given by

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x}$$

Show that a necessary and sufficient condition for $\mathbf{x} = \mathbf{0}$ to be a u.a.s.i.l. solution there must exist a positive define matrix $\mathbf{P}$ such that

$$\mathbf{A}^T\mathbf{P} + \mathbf{P}\mathbf{A} = -\mathbf{I}$$

2.22. Let $\mathbf{A}$ be given by

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix}$$

and

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x}$$

Find the Lyapanov function

$$V(x) = \mathbf{x}^T\mathbf{P}\mathbf{x}$$

2.23. Let a discrete-time linear time-invariant system be given by

$$\mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k)$$

Show that the equilibrium state is u.a.s.i.l. if and only if for any positive definite matrix $\mathbf{Q}$ there exists a positive definite matrix $\mathbf{P}$ such that

$$\mathbf{G}^T\mathbf{P}\mathbf{G} - \mathbf{P} = -\mathbf{Q}$$

Then $V(x) \equiv \mathbf{x}^T\mathbf{P}\mathbf{x}$ and $\dot{V}(x) = -\mathbf{x}\mathbf{Q}\mathbf{x}$

2.24. A function $\mathbf{f}(\mathbf{x})$ is called a contraction mapping if

$$\|\mathbf{f}(\mathbf{x}_1) - \mathbf{f}(\mathbf{x}_2)\| < \|\mathbf{x}_1 - \mathbf{x}_2\|$$

For the system

$$\mathbf{x}(k+1) = \mathbf{f}(\mathbf{x}(k)),\ \mathbf{f}(\mathbf{0}) = \mathbf{0}$$

where $\mathbf{f}(\ )$ is a contraction mapping, show that the system is u.a.s.i.l and one of its Lyapanov functions is

$$V(\mathbf{x}) = \|\mathbf{x}\|$$

2.25. Let $C_1 \geqq 0$ and $u(t)$ and $K(t)$ be continuous functions. Let $K(t) \geqq 0$. Show that if

$$u(t) \leqq C_1 + \int_{t_0}^{t} \mathbf{K}(\xi)u(\xi)\,d\xi$$

then

$$u(t) \leqq C_1 \exp\left(\int_{t_0}^{t} \mathbf{K}(\xi)\,d\xi\right)$$

This is called the Bellman-Gromwall lemma.

2.26. Consider the time-varying linear system given by

$$\dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t)$$
$$\mathbf{z}(t) = \mathbf{C}(t)\ \mathbf{x}(t)$$

Show that it satisfies all the conditions to be a dynamic system.

2.27. A dynamical system is described by the scalar equation

$$\dddot{y}(t) + 6\ddot{y}(t) + 11\dot{y}(t) + 6y(t) = u(t)$$
$$z(t) = 5y(t) + 8\ddot{y}(t)$$

(a) Draw a block diagram of this system.
(b) Write the system in state form.
(c) Find the transition matrix and write the corresponding discrete-time system with a sample time of T sec.
(d) Let $\mathbf{x}(0)$ equal 0 and $u(t) = t$. Find $\mathbf{x}(t)$.

CHAPTER 3

THE STOCHASTIC MODEL

Chapter 2 presented the deterministic portion of the model we wish to employ in the ensuing chapters. In that chapter we formulated the deterministic model and discussed several salient issues that will be needed later. If, indeed, our model were completely deterministic and observable, then based upon a set of measurements we could without error give the state of the system for *any* time. Unfortunately, both system and measurements are disturbed by stochastic processes. It will be the purpose of this chapter to discuss the theoretical properties of the processes that are used in the analysis.

The models we develop will depend strongly on the deterministic structure and the processes to be defined in this chapter. For example, we may be told that a system is driven by white noise, that is, noise whose power-density spectrum is constant for all frequencies. Yet, as we shall see, mathematically no such noise process exists and that system whose definition depends on such behavior must be carefully constructed. There is another facet of the problem we must deal with, that of obtaining results. In Chapter 5, by choosing the noise to be an independent increment process, we can obtain solutions to the filtering problem. For this reason we shall then concentrate on independent increment processes and investigate their effects on dynamical systems. Another simplification that is reasonable and almost necessary is that the processes be Markov processes. Such processes fall easily into the framework of dynamical systems because they allow descriptions of systems behavior to be determined from initial conditions.

This chapter will be highly theoretical by necessity. The reader may find if this is his first acquaintance with these concepts that a reading of the theorems and examples should suffice in order to understand the following chapters. Yet he should be aware of the model that is used so that he does not try to extend the results of the latter chapters to cases for which they are not valid.

We first discuss the structure of the probability space and introduce the concept of a Markov process. In Section 3.2 independent increment processes are defined and discussed. The Wiener process and the generalized Poisson

process are introduced. General properties of independent increment processes are discussed. Specific properties of the Wiener process are developed in Section 3.3 with the introduction of the concept of a martingale. In this section we show that the Wiener process is continuous but not of bounded variation.

We discuss in depth the nature and structure of stochastic integrals in Section 3.4 and stochastic differential equations. In some of the most recent work in the area of solving nonlinear filtering problems these results are found to be invaluable. In this discussion we present the Ito integral and the Strantonovich integral. The reason for this presentation is to let the reader know that there is not necessarily a unique representation of integrals when the measure is an independent increment process. This is not an unrelated problem of only academic interest, as the reader will notice, for these two interpretations led Kushner and Strantonovich to obtain two different equations for the optimum filter.

The final step in the development of the stochastic model will be to apply these integral representations to the solution of stochastic differential equations. It will be interesting to see that their solutions do not follow the usual rules applied to normal deterministic differential equations. We also present Ito's differentiation rule in the last section as an extension to stochastic differential equations.

The theory contained in this chapter can be found in varying degrees in Doob; Ito; Ito and McKean; McKean; and several current papers referenced in the body of the chapter. The greatest contributors to this area of investigation have been Wiener and Ito, to whom the entire theory of filtering owes a great debt.

3.1 STOCHASTIC PROCESSES

The accurate presentation of the theory of estimating stochastic processes initially requires an introduction to the fundamental structure of probability spaces. With this structure defined, a stochastic process can easily be defined thereupon, and its properties become closely aligned with that of the underlying probability space. A probability space is built up from a certain or sure event Ω that occurs with probability 1 and other less certain events A_i. On this space we define a probability measure P that gives us a quantitative description of the chance of any of the A_i events occurring.

The first restriction that must be applied to this concept of a probability space is that the sets or events, A_i must have certain intuitive properties. The first of these is that if A_i is an event then the complement of A_i must also be an event. Second, if A_i and A_j are events, then their union must also be an event; similarly their intersection must also be an event. However, there is

one further property that this class of events must satisfy: they must be such that if each A_i for a countable number of i is an event, then their union is an event also. This restriction is fundamental to probability, wherein we often ask questions concerning convergence, and such questions must be properly defined in terms of events (see Breiman, p. 402). The requirements that these events must possess leads us to define a structure for it.

DEFINITION 1.1. A class of sets $\{A_i\}$ is called a *σ-field*, $\mathcal{A}$ if

1. $\Omega \in \mathcal{A}, \phi \in \mathcal{A}$ where ϕ is the null set
2. $A_i \cup A_j \in \mathcal{A}$ if $A_i, A_j \in \mathcal{A}$
3. $A_i^c \in \mathcal{A}$ if $A_i \in \mathcal{A}$
4. $A_i \cap A_j \in \mathcal{A}$ if $A_i, A_j \in \mathcal{A}$
5. $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$ $\forall A_i \in \mathcal{A}$.

The concept of a σ-field (or σ-algebra) follows from measure theory (see Halmos [2]) and is used to construct a consistent theory of measure and integration on abstract spaces. The sets A_i are composed of points $\omega \in \Omega$ that have certain properties. The assignment of quantitative values to each of these events or ω-sets is called a probability measure P. This measure has the following properties.

DEFINITION 1.2. A *probability measure* P is a function defined on the sets $A \in \mathcal{A}$ such that

$$0 \leqq P[A] \leqq 1, P[\phi] = 0, P[\Omega] = 1$$

and

$$P[A] = \sum_{i=1}^{\infty} P[A_i]$$

if

$$A = \bigcup_{i=1}^{\infty} A_i \quad \text{and} \quad A_i \cap A_j = \phi \qquad (\forall i \neq j)$$

This probability measure thus defines intuitively a quantity that will yield the relative occurrence of any event. With the set Ω, the field $\mathcal{A}$, and the measure P we can define a probability space.

DEFINITION 1.3. A space Ω with points ω, together with a σ-field $\mathcal{A}$ of sets in Ω and a probability measure $P[\ ]$ that is defined on all sets in $\mathcal{A}$, is called a probability space and is denoted by the triple $(\Omega, \mathcal{A}, P)$.

The events $A \in \mathcal{A}$ are called measurable events, and we call $P[\ ]$ a probability distribution on the sets in $\mathcal{A}$ (see Cramer and Leadbetter, Chapter 2, for extensions). Now the abstract nature of $(\Omega, \mathcal{A}, P)$ can be reflected into quantities that are seen in reality—for example, voltages, currents, pressures.

All of these quantities are random and may be defined relative to the underlying abstract probability space. We shall let $x(\omega)$ represent any one of these quantities. The quantity x (pressure, voltage) is random because it depends upon some point $\omega \in \Omega$. This quantity is called a random variable if it satisfies certain conditions.

DEFINITION 1.4. Let $(\Omega,\mathscr{A},P)$ be a probability space. Let x be a function that maps Ω into $\mathbf{R}$ where $\mathbf{R}$ is the real line. Then, if $x_0 \in \mathbf{R}$ and if the set,

$$\{\omega: x(\omega) < x_0\} \in \mathscr{A}$$

for all $x_0 \in \mathbf{R}$, where $\mathscr{A}$ is the σ-field defined on Ω, then $x(\omega)$ is called a *random variable*.

Random variables are therefore special classes of mappings from Ω into $\mathbf{R}$ such that the inverse images of the open intervals in $\mathbf{R}$ are events. This requirement is extremely useful, since now, if we ask what is the probability that the pressure $x(\omega)$ is between 50 psi and 90 psi, such a statement defines an event, $A_i \in \mathscr{A}$, to which we have ascribed a probability. The reason for choosing the open intervals in $\mathbf{R}$ is that the class of open intervals in $\mathbf{R}$ generate a σ-field called the Borel field, $\mathscr{B}$. The sets belonging to the Borel field are called Borel sets. Thus the set

$$\{x: a \leqq x < b\}$$

is a Borel set. A random variable then is a mapping whereby all inverse images of $\mathscr{B}$ are events and belong to the σ-field $\mathscr{A}$. That is,

$$\{\omega: x(\omega) \in B, B \in \mathscr{B}\} = A \qquad (A \in \mathscr{A})$$

Functions that are random variables are also called *measurable functions* or *transformations*.

We can generalize this to the case where x maps Ω into $\mathbf{R}^n$, the n dimensional euclidean space. Then x is an $n \times 1$ vector, and we define probabilities as

$$P[\{\omega: x_1(\omega) < \xi_1 \cdots x_n(\omega) < \xi_n\}].$$

This quantity appears quite frequently, so we shall call it

$$P[x_1(\omega) < \xi_1 \cdots x_n(\omega) < \xi_n]$$

so that the correct notation is understood.

The next obvious extension is to let x be a function of time, thus generating a stochastic process. Stochastic processes play a central role in our discussions. These processes are obtained by considering the underlying probability space Ω and an interval of time T and letting x map the product space $\Omega \times T$ into $\mathbf{R}^n$. The mapping x from $\Omega \times T$ into $\mathbf{R}^n$ is a measurable function of the σ-field $\mathscr{A}$ of Ω.

This leads us to defining a stochastic process as follows.

DEFINITION 1.5. A *stochastic process* is a finite real valued function $x(t, \omega)$ that is a mapping from $\Omega \times T$, for some interval T, into $\mathbf{R}^n$ such that for each fixed $t \in T$, $x(t, \omega)$ is a measurable function of $\mathscr{A}$. That is, for each $t \in T$ the event

$$\{\omega: a_1 < x_1(t) < b_1 \cdots a_n < x_n(t) < b_n\} \in \mathscr{A}$$

where $\mathscr{A}$ is the σ-field of events on Ω.

A great deal more can be said about stochastic processes in an abstract sense, but for our purposes such a digression whould be of little value. Thus, we shall assume that a stochastic process $\{x(t, \omega), \omega \in \Omega\}$ represents an ensemble of possible wave forms that one could measure. We shall then let $x(t)$ be the representation of the process with the understanding that $x(t,\omega)$.

Human speech is a stochastic process, as is the noise observed on radar sets. In general, the stochastic process is a function of time. This is not a necessary restriction, for such processes may also be considered to be functions of space. For example, the pressure on the surface of the earth varies from point to point. Here, then, the random or stochastic process is a function of the spatial coordinates. In such a case, the definition of a stochastic process would have to be amended to let $T \subset \mathbf{R}^m$, where m is the dimension of the parameter space of the process. Such processes are also called *random fields*.

We would like to be able to describe stochastic processes in a consistent fashion. It is obvious that listing $x(t,\omega)$ for all $\omega \in \Omega$ would be prohibitive. A useful extension, though, would be to investigate and structure the process based upon its probabilistic nature. For example, we could consider the process evaluated at some point $t = t_i$. At that point $x(t_i)$ is now a random variable and, as such enables us to, define a probability distribution function. That is, we could let $P[x(t_i) \geqq \lambda]$ be the probability that the process at time t_i is greater than some constant λ. Such a description is quite useful. But it is also limited because it says nothing about the process except at a single point. Therefore, in general, we would like to know the statistical nature of the process at many points $t = t_1, \cdots, t_n$. This is given by

$$P[x(t_1) \geqq \lambda_1; x(t_2) \geqq \lambda_2; \cdots; x(t_n) \geqq \lambda_n]$$

This is still incomplete since it is for finite n and discrete t. Ultimately, we would like to know this for all n and a set of $\{t_n\}$ dense on some interval T. Such a joint distribution function is almost impossible to obtain. The exception is for a class of processes called Markov processes. A Markov process is one in which knowledge of the present given knowledge of the past depends solely upon the most recent past knowledge. In this case, we shall find the

conditional distribution of $x(t_n)$ given $x(t_{n-1})$ and other $x(t_j)$ for all j less than $n-1$.

Thus, we shall first discuss the Markov process and its implications. The class of Markov processes that we wish to study consists of continuous-time and continuous-state processes. This means that T is some interval of time and the process $x(t)$, called the state, can take on any value in $\mathbf{R}^n$. There are other classes of Markov processes that are useful. One that we shall encounter later will be the continuous-state discrete-time process. In that case, $x \in \mathbf{R}^n$ but T is countable and possibly even finite. What we shall now do is to first develop the concept of conditional expectation and then define what we mean by a Markov process, provide an example of how it relates to a dynamical system of the continuous-state discrete-time form and then prove the Chapmann-Kolmogorov theorem, which will allow one to evaluate an arbitrary distribution function. We shall also assume where necessary that the density function, the derivative of the distribution function, exists.

The ability to describe a stochastic process defined for a countable number of times $\{t_i\}$ given the probability distribution for a finite number of times is given in the classical Kolmogorov extension theorem (see Billingsley, Appendix II). The theorem says that there exists a well-defined distribution for $x(t_i)$, for all $\{t_i\}$, given the distribution for some finite set $\{t_i\}$. The underlying assumption is that the process can be adequately defined by a countable set $\{t_i\}$. We shall assume that all our processes can adequately be defined by some countably dense set $\{t_i\}$, $t_i \in \mathbf{R}$. This property of processes is called *separability*. Thus, we shall assume that all processes are separable. The concepts of separable processes are discussed in Doob and in Loeve but involve ideas beyond the scope of the present analysis. The concept of separability will be looked at in Chapter 5 when we discuss conditional probabilities.

One of the more important concepts of probability theory and one that we will rely on heavily is that of conditional expectation. The standard definition used in elementary treatments of probability quickly becomes of little use in the area of continuous estimation theory and a more rigorous and complete definition is required. In order to develop this understanding, we shall first present a constructive definition of the conditional expectation and from it obtain a descriptive definition. This approach follows Loeve (pp. 337–349), and other approaches are found in Doob (pp. 15–34), Gikhman and Skorokhod (pp. 134–143), and Breiman (pp. 73–80).

From the elementary point of view, the conditional probability of some event A, given an event B, is defined as the ratio of the joint probability of both events to that of event B alone. Likewise, if $x(\omega)$ is a random variable on $(\Omega, \mathscr{A}, P)$, then the expected value of $x(\omega)$ given B, $E[x|B]$ is defined in a similar manner. Now B can be any set belonging to $\mathscr{A}$, so that the conditional

expectation can be interpreted not just as some constant but as an ω-function also. That is, $E[x|B]$ is a function of ω, whose value is constant, $E[x|B]$ on the set B. This can be generalized if we consider a class of sets $\{B_i\}$ and define the indicator function I_{B_i} as an ω function such that

$$I_{B_i} = \begin{cases} 1 & \omega \in B_i \\ 0 & \omega \notin B_i \end{cases}$$

Let $\mathscr{B}$ be the σ-field generated by the set $\{B_i\}$. Then we can think of the conditional expectation as an ω-function

$$E[x|\mathscr{B}] = \sum_i E[x|B_i]\, I_{B_i}$$

That is, $E[x|\mathscr{B}]$ is an ω-function that takes on the value $E[x|B_i]$ whenever $\omega \in B_i$. Thus, $E[x|\mathscr{B}]$ is an ω-function that is constant on the sets B_i. Therefore, rather than defining the conditional expectations of x with respect to a given set, we can define it with respect to the σ-field generated by those sets. For example, if $x(\omega)$ and $y(\omega)$ are random variables measurable with respect to $(\Omega,\mathscr{A},P)$ and if B is a Borel set, then $E[x|y \in B]$ can be defined as $E[x|\mathscr{B}_y]$, where $\mathscr{B}_y$ is the σ-field generated by y. That is, since y is a random variable from Ω into say $\mathbf{R}$, then if $\{B_i\}$ are the Borel sets in $\mathbf{R}$, the inverse sets $y^{-1}(B_i) = \{\omega: y(\omega) \in B_i\}$ generate a σ-field $\mathscr{B}_y$ called the σ-field generated by y. The σ-field $\mathscr{B}_y$, so defined, is also called a sub σ-field of $\mathscr{A}$ ($\mathscr{B}_y \subset \mathscr{A}$), since by definition $\mathscr{B}_y$ is a sub σ-field of $\mathscr{A}$ if all $B_i \in \mathscr{B}_y$ are such that $B_i \in \mathscr{A}$.

Thus, we can consider the following constructive definition of conditional expectation.

DEFINITION 1.6. Let x be a random variable on $(\Omega,\mathscr{A},P)$ and let $\mathscr{B}$ be a sub σ-field of $\mathscr{A}$ with $B_i \in \mathscr{B}$. The conditional expectation of x is defiend as

$$\begin{aligned} E[x|\mathscr{B}] &= \sum_i E[x|B_i]\, I_{B_i} \\ &\triangleq \sum_i \left(\frac{1}{P[B_i]} \int_{B_i} x(\omega) dP(\omega) \right) I_{B_i} \end{aligned}$$

and is an ω-function measurable with respect to $\mathscr{B}$.

Thus, from the elementary theory, such as expression as $E[x|y \leqq 5]$ may be thought of as the value of $E[x|\mathscr{B}]$ at a specific point in Ω. Up to this point. such a definition is more cumbersome than the elementary approach, but the advantage of such an approach becomes clear when we use the conditioning on a segment of a random process. We shall consider this in great detail in Chapter 4 but shall briefly outline its structure in the descriptive definition of the conditional expectation.

Before proceeding, it should be clear that the conditional probability is merely a special case of conditional expectation. That is, if we let x be I_A for some $A \in \mathscr{A}$, then $E[x|\mathscr{B}]$ is equal to $P[A|\mathscr{B}]$.

We can now present a second and more general definition of conditional expectation. What motivates it is that in the constructive approach the conditional expectation is an ω-function measurable with respect to a given sub σ-field $\mathscr{B}$. Thus:

DEFINITION 1.7. Let x be a random variable on $(\Omega,\mathscr{A},P)$ and let $\mathscr{B}$ be a sub σ-field of $\mathscr{A}$. The condition expectation of x, $E[x|\mathscr{B}]$ is any $\mathscr{B}$-measurable function such that for all $B \in \mathscr{B} \subset \mathscr{A}$

$$\int_B E[x|\mathscr{B}]\, dP = \int_B x\, dP$$

As an example of how such a definition can be used, let $x(t)$ be a random process. Then $E[x(t)|x(s), s \leqq u]$ is given as $E[x(t)|\mathscr{F}_u]$ where $\mathscr{F}_u$ is the sub σ-field generated by $\{x(s),\ s \leqq u\}$. Thus, $E[x(t)|\mathscr{F}_u]$ is an $\mathscr{F}_u$-measurable function, an ω-function on $\mathscr{F}_u$.

In the preceding we said $E[x|\mathscr{B}]$ was an ω-function. We can extend this if we let y be a random variable and let $\mathscr{B}_y$ be the sub σ-field generated by y.

THEOREM 1.1.

Let x be a random variable on $(\Omega,\mathscr{A},P)$ into $(\Omega',\mathscr{B}_x)$. Let y be a function from $(\Omega',\mathscr{B}_x)$ into $(\Omega',\mathscr{B}_y)$. Then

$$E[x|y] = f(y)$$

where $f(\ \ )$ is a $\mathscr{B}_x$-measurable function.

Proof. The proof of the theorem is in Gikhman and Skorokhod (p. 136) or Loeve (pp. 342–343).

This theorem says that the conditional expectation $E[x|\mathscr{B}_y]$ can also be written as a function of y and not just as a function of ω. This therefore relates the ω-function definition to the elementary definition of a conditional expectation.

The class of all stochastic processes is too general for many purposes, but by considering subclasses with specified properties, a great deal can be said. A specific class of processes that lends itself to being quite useful is that of Markov processes. This class possesses the property that the statistics of the process at the present time, conditioned upon knowledge of the past, depend only upon the most recent past. This property will allow us in Chapter 5 to write probability densities in a simple form for many processes.

DEFINITION 1.8. A *Makrov process* is a stochastic process $x(t)$, $t \in T$, that satisfies the following conditions: For any integer $n \geqq 1$, if $t_1 < t_2 < \cdots < t_n$ are parameter values, the conditional probabilities of $x(t_n)$, given $x(t_1), \cdots, x(t_{n-1})$, are the same as those just given $x(t_{n-1})$. That is,

$$P[x(t_n) \geqq \lambda \mid x(t_1) \cdots x(t_{n-1})] = P[x(t_n) \geqq \lambda \mid x(t_{n-1})]. \tag{1.1}$$

Now if the times are continuous intervals, we infer that if $s < t$, then if $\mathscr{F}_s$ is the sub σ-field generated by $\{x(u);\ u \leqq s\}$, then

$$P[x(t) \geqq \lambda \mid \mathscr{F}_s] = P[x(t) \geqq \lambda \mid x(s)] \tag{1.2}$$

Thus, Markov processes are independent of the past. We can now consider a Markov process generated by a discrete-time system.

Example. Given the process

$$\mathbf{x}(k+1) = \boldsymbol{\Phi}(k+1, k)\mathbf{x}(k) + \mathbf{u}(k) \tag{1.3}$$

Let the $\mathbf{u}(k)$ be independent Gaussian random variables such that

$$E[\mathbf{u}(k)\mathbf{u}^T(j)] = \mathbf{Q}(k)\delta_{jk} \tag{1.4}$$

Then also assume that $\mathbf{x}(0)$ is zero mean a Gaussian random variable with

$$E[\mathbf{x}(0)\mathbf{x}^T(0)] = \mathbf{P}(0) \tag{1.5}$$

and assume that $\mathbf{x}(0)$ and $\mathbf{u}(k)$ are independent random variables. Now we want to find

$$P[\mathbf{x}(k+1) \geqq \lambda \mid \mathbf{x}(k) \cdots \mathbf{x}(0)] \tag{1.6}$$

It should be obvious that $\mathbf{x}(k)$ is Gaussian since it is the sum of Gaussian random variables. Furthermore, it should also be obvious that $\mathbf{x}(k+1)$, given $\mathbf{x}(k) \cdots \mathbf{x}(0)$, depends only on $\mathbf{x}(k)$. Then,

$$\begin{aligned} &P[\mathbf{x}(k+1) \geqq \lambda \mid \mathbf{x}(k) \cdots \mathbf{x}(0)] \\ &= P[\mathbf{x}(k+1) \geqq \lambda \mid \mathbf{x}(k)] \qquad (1.7) \\ &= \int_{\lambda}^{\infty} \frac{1}{(2\pi)^{n/2}|\mathbf{Q}(k)|^{1/2}} \exp\left[-\tfrac{1}{2}\|\mathbf{x}(k+1) - \boldsymbol{\Phi}(k+1, k)\mathbf{x}(k)\|_{\mathbf{Q}^{-1}(k)}\right] d\mathbf{x}(k+1) \end{aligned}$$

Let us now consider that we only know $\mathbf{x}(k-2)$ and all the past beyond it. Namely, what is

$$P[\mathbf{x}(k+1) \geqq \lambda \mid \mathbf{x}(k-2) \cdots \mathbf{x}(0)] \tag{1.8}$$

Now

$$\mathbf{x}(k-1) = \boldsymbol{\Phi}(k-1, k-2)\mathbf{x}(k-2) + \mathbf{u}(k-2) \tag{1.9}$$

Further,

$$\begin{aligned} \mathbf{x}(k) &= \boldsymbol{\Phi}(k, k-1)\mathbf{x}(k-1) + \mathbf{u}(k-1) \\ &= \boldsymbol{\Phi}(k, k-1)[\boldsymbol{\Phi}(k-1, k-2)\mathbf{x}(k-2) + \mathbf{u}(k-2)] + \mathbf{u}(k-1) \\ &= \boldsymbol{\Phi}(k, k-2)\mathbf{x}(k-2) + \boldsymbol{\Phi}(k, k-1)\mathbf{u}(k-2) + \mathbf{u}(k-1) \end{aligned} \tag{1.10}$$

where we have used the property of the transition matrix. Also,

$$\begin{aligned} \mathbf{x}(k+1) &= \boldsymbol{\Phi}(k+1, k)[\mathbf{x}(k)] + \mathbf{u}(k) \\ &= \boldsymbol{\Phi}(k+1, k-2)\mathbf{x}(k-2) + \boldsymbol{\Phi}(k+1, k-1)\mathbf{u}(k-2) \\ &\quad + \boldsymbol{\Phi}(k+1, k)\mathbf{u}(k-1) + \mathbf{u}(k). \end{aligned} \tag{1.11}$$

It is obvious that the process is Markov, since

$$P[\mathbf{x}(k+1) \geqq \lambda \mid \mathbf{x}(k-2) \cdots \mathbf{x}(0)] = P[\mathbf{x}(k+1) \geqq \lambda \mid \mathbf{x}(k-2)] \quad (1.12)$$

We can define a new noise as

$$\mathbf{u}^*(k) = \Phi(k+1, k-1)\mathbf{u}(k-2) + \Phi(k+1, k)\mathbf{u}(k-1) + \mathbf{u}(k) \quad (1.13)$$

which has a convariance function

$$\begin{aligned} E[\mathbf{u}^*(k)\mathbf{u}^{*T}(k)] &= \Phi(k+1, k-1)\mathbf{Q}(k-2)\Phi^T(k+1, k-1) \\ &\quad + \Phi(k+1, k)\mathbf{Q}(k-1)\Phi^T(k+1, k) + \mathbf{Q}(k) \\ &\equiv \mathbf{Q}^*(k) \end{aligned} \quad (1.14)$$

Therefore,

$$P[\mathbf{x}(k+1) \geqq \lambda | \mathbf{x}(k-2)] = \int_{\lambda}^{\infty} \frac{1}{(2\pi)^{n/2}|\mathbf{Q}^*(k)|^{1/2}}$$
$$\cdot \exp\left[-\tfrac{1}{2}\|\mathbf{x}(k+1) - \Phi(k+1, k-2)\mathbf{x}(k-2)\|_{\mathbf{Q}^{*-1}(k)}\right] d\mathbf{x}(k+1) \quad (1.15)$$

Such examples will be common in the development of the filtering equation. We will also be interested in the continuous-time version of the above that will be done later.

Such processes when exciting continuous-time dynamic systems in the proper fashion will produce Markov processes.

In most cases of interest the probability distribution function $P[\ ]$ is differentiable and its derivative is easier to manipulate. The derivative of the distribution function is called the *probability density function* and is denoted in the following definition.

DEFINITION 1.9. Let $P[\ ]$ be a probability distribution function. Let the probability distribution function of the random vector $\mathbf{x} \in \mathbf{R}^n$ be given by

$$P[x_1(\omega) < u_1; \cdots; x_n(\omega) < u_n]$$

Then the probability density of the random vector $\mathbf{x}$ is given by

$$p_{\mathbf{x}}(\mathbf{u}) \cdots = \frac{\partial^n}{\partial u_1 \cdots \partial u_n} P[x_1(\omega) < u_1; \cdots; x_n(\omega) < u_n] \quad (1.16)$$

where $\mathbf{u} \in \mathbf{R}^n$ and u_i are the components of the $n \times 1$ vector $\mathbf{u}$.

Similarly, for conditional density functions we shall write $p_{\mathbf{x}}(\mathbf{u}|\mathbf{y})$ for the conditional density of $\mathbf{x}$ at $\mathbf{u}$, given random variable $\mathbf{y}$. In general, we shall assume that these derivatives exist and are bounded. If not, then a more general analysis can be made, but use of $P[\ ]$ must be made (see Loeve or Doob).

We will now want to prove a simple theorem called the Chapmann-Kolmogorov theorem, which will be important in the study and analysis of more complicated processes. It provides the tool necessary to give a complete statistical description to a Markov process in terms of a density function.

It will be used extensively later in the development of the estimation equations.

THEOREM 1.2

Let $p_{\mathbf{x}}(\mathbf{u}, t\,|\mathbf{x}_0, t_0)$ be a probability density function on the Markov process $\mathbf{x}(t)$; given that $\mathbf{x}(t_0) = \mathbf{x}_0$, and $t_0 < t$, then.

$$p_{\mathbf{x}}(\mathbf{u}, t\,|\mathbf{x}_0, t_0) = \int p_{\mathbf{x}}(\mathbf{u}, t\,|\mathbf{v}, s) p_{\mathbf{x}}(\mathbf{v}, s|\mathbf{x}_0, t_0)\, d\mathbf{v} \tag{1.17}$$

where $t_0 < s < t$.

Note that the conditional probability density $p_{\mathbf{x}}(\mathbf{u}, t\,|\mathbf{x}_0, t_0)$ is an evaluation of the ω-function used to define the corresponding conditional expectation. We shall use this notation wherever convenient to define such an evaluation.

Proof. What we wish to show is that by transferring a process from some initial state $\mathbf{x}(t_0)$ to a final state $\mathbf{x}(t)$, we can do it in two steps. First, we can go from $\mathbf{x}(t_0)$ to a state $\mathbf{x}(s)$ that lies between $\mathbf{x}(t_0)$ and $\mathbf{x}(t)$, as shown in Figure 3.1. Then we transfer from the state $\mathbf{x}(s)$ to the state $\mathbf{x}(t)$.

In general, we know that

$$p_{\mathbf{x}}(\mathbf{u}, t\,|\mathbf{x}_0, t_0) = \int p_{\mathbf{x}}(\mathbf{u}, t\,|\mathbf{v}, s; \mathbf{x}_0, t_0) p_{\mathbf{x}}(\mathbf{v}, s|\mathbf{x}_0, t_0)\, d\mathbf{v} \tag{1.18}$$

Figure 3.1 Geometric interpretation of Champmann-Komlogorov theorem.

But, since the process is Markov, we know that

$$p_{\mathbf{x}}(\mathbf{u}, t \,|\mathbf{v}, s; \mathbf{x}_0, t_0) = p_{\mathbf{x}}(\mathbf{u}, t \,|\mathbf{v}, s) \tag{1.19}$$

Then

$$p_{\mathbf{x}}(\mathbf{u}, t \,|\mathbf{x}_0, t_0) = \int p_{\mathbf{x}}(\mathbf{u}, t \,|\mathbf{v}, s) p_{\mathbf{x}}(\mathbf{v}, s|\mathbf{x}_0, t_0)\, d\mathbf{v} \blacksquare \tag{1.20}$$

The density $p_{\mathbf{x}}(\mathbf{u}, t \,|\mathbf{v}, s)$ is called the *transition probability density,* and it shows how the Markov process progresses in time. Note that given $p_{\mathbf{x}}(\mathbf{u}, t \,|\mathbf{v}, s)$ for any $\mathbf{u}$, t, $\mathbf{v}$, s we can obtain it for any other set of states and times. Thus, the transition density is very important in the study of Markov processes, and it is this density to which we will direct our studies in Chapter 5. The transition density acts for stochastic systems as the transition matrix acts for deterministic systems. It projects the state of the system from one instant to the state at some other instant of time. Furthermore, a complete statistical description is possible with only the transition function and some initial density. Namely, to obtain $p_{\mathbf{x}}(\mathbf{u}_1, t_1; \cdots; \mathbf{u}_n, t_n)$, we note that it is equal to $p_{\mathbf{x}}(\mathbf{u}_n, t_n|\mathbf{u}_{n-1}, t_{n-1}) \cdots p_{\mathbf{x}}(\mathbf{u}_2, t_2|\mathbf{u}_1, t_1) p_{\mathbf{x}}(\mathbf{u}_1, t_1)$ as a result of the Markov property.

This establishes the material we need concerning Markov processes. A more structured definition is contained in Ito [2] [p. 20].

3.2 PROCESSES WITH INDEPENDENT INCREMENTS

In this section we shall discuss processes that are called *independent increment processes.* The study of these will take up the greatest portion of our interest in later sections. The first important independent increment process will be the Wiener process, which is the basis for the study of most processes that are continuous in a state space. By this, we mean that the state variables as discussed in the preceding chapter are allowed to take on a continuum of values. Contrasted to this would be the discrete state process typified by the simple Poisson process. This type of process takes on only discrete values in the state space. We are in general interested in the estimation of the state of processes that are corrupted by noises of the two preceding forms.

The Wiener process is an independent increment process that, when formally differentiated, yields white noise. The white noise process is an ideal process for use in modeling such effects as measurement noise and system forcing functions. It is only one of the class of independent increment processes.

A second member of the class of independent increment processes that we concentrate upon is the Poisson process, because it too is important in many areas of estimation. Many physical problems fall into the category of discrete levels. For example, if we are trying to estimate the trajectory of a molecule

or some virus by the use of a scanning electron microscope, we receive electrons for signal measurements and these received signals are usually modeled by Poisson statistics. We may desire the state, given by position and momentum, of this particle and also wish to estimate its mass and viscous damping coefficient.

Another topic will be a discussion of processes with orthogonal increments. It will be easy to see that if the process has independent increments and if the second moments are bounded, it will also have orthogonal increments. Thus, the class of processes with orthogonal increments will contain the class of processes with independent increments.

We end this section with a discussion of the properties of other classes of independent increment processes and show how they relate to the Wiener and Poisson process defined. For extensions of the following, the reader is referred to Doob [2] (pp. 46–101), Ito [2] (pp. 1–17), and McKean [2] (pp. 1–19).

DEFINITION 2.1. Let $x(t,\omega)$ be a random process defined on $(\Omega,\mathcal{A},P)$ and let T be a finite set of times $\{t_i\}$. The process $x(t,\omega)$ is said to be a *Gaussian process* if for all sets T and all constants c_i the random variable $z(\omega)$

$$z(\omega) = \sum_{t_i \in T} c_i x(t_i, \omega) \tag{2.1}$$

is a Gaussian random variable.

Recall that if $\mathbf{x}(\omega)$ is an $n \times 1$ random vector and if it is Gaussian, the probability density of $\mathbf{x}(\omega)$ is

$$p_{\mathbf{x}}(\mathbf{u}) = \frac{1}{(2\pi)^{n/2}|\Lambda|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{m})^T \Lambda^{-1}(\mathbf{x} - \mathbf{m})\right) \tag{2.2}$$

where Λ is the $n \times n$ covariance matrix defined by

$$\Lambda = E[(\mathbf{x} - \mathbf{m})(\mathbf{x} - \mathbf{m})^T] \tag{2.3}$$

and $\mathbf{m}$ is the mean $E[\mathbf{x}]$. Thus, it should be clear from the definition of a Gaussian process that the set of variables $\{\mathbf{x}(t_i)\}$ are jointly Gaussian random variables. Thus, for a Gaussian random process, as shown in Figure 3.2, the variables $\mathbf{x}(t_i,\omega)$ are each separately Gaussian and also jointly so, having the probability density given above.

DEFINITION 2.2. The random process $x(t)$, $t \in T$ is an *independent increment process* if the increments

$$x(t_i) - x(t_j);\ x(t_n) - x(t_m) \tag{2.4}$$

where

$$t_j < t_i \leqq t_m < t_n \tag{2.5}$$

have conditional probabilities of the form

Figure 3.2 Example of a random process.

Figure 3.3 An example of an independent increment process.

$$P[x(t_i) - x(t_j) \in \lambda | x(t_n) - x(t_m) = \xi] = P[x(t_i) - x(t_j) \in \lambda] \qquad (2.6)$$

for any λ in $\mathbf{R}^1$ and ξ a point in $\mathbf{R}^1$.

This implies that the increments of the process over two nonoverlapping

times are independent (this is shown in Figure 3.3). That is, the random variables $\theta = x(t_2) - x(t_1)$ and $\psi = x(t_4) - x(t_3)$ are independent. Note that this is *not* true of the variables $x(t_4) - x(t_3)$ and $x(t_6) - x(t_5)$, because they have a common overlap area. Yet, if we create three random variables $x(t_3) - x(t_5)$, $x(t_6) - x(t_3)$, and $x(t_4) - x(t_6)$, then for any λ_1, λ_2, and λ_3

$$\begin{aligned} &P[x(t_3) - x(t_5) \in \lambda_1;\, x(t_6) - x(t_3) \in \lambda_2;\, x(t_4) - x(t_6) \in \lambda_3] \\ &= P[x(t_3) - x(t_5) \in \lambda_1]P[x(t_6) - x(t_3) \in \lambda_2]P[x(t_4) - x(t_6) \in \lambda_3]. \end{aligned} \tag{2.7}$$

That is, the independence implies a factoring of the probabilities. The following two definitions will give the structure of two important independent increment processes, the Wiener process and the Poisson process.

DEFINITION 2.3. Let $x(t, \omega)$; $t \in [0, \infty]$ be a scalar stochastic process such that

1. $x(0, \omega) = 0$ for almost all ω;
2. the process $x(t)$ is an independent increment process;
3. for $t \geqq s$,

$$P[x(t) - x(s) < \lambda] = \frac{1}{[2\pi(t - s)]^{1/2}} \int_{-\infty}^{\lambda} \exp\left(\frac{-\xi^2}{2(t - s)}\right) d\xi \tag{2.8}$$

That is, the increments have a Gaussian distribution. Then this is called a normalized *Wiener process*. When the variance is $\sigma^2|t - s|$, we have a Wiener process that is not normalized.

One should note that the variance of the Wiener process has values that are proportional to the time difference between the samples. We may generalize this to several important cases, as shown in the following example.

Example. Let $x(t)$ be a normalized Wiener process as given in the previous definition. Define

$$u(t) = \sigma x(t) \tag{2.9}$$

Then

$$E[u(t)] = E[\sigma x(t)] = \sigma E[x(t)] \tag{2.10}$$

But from (3) of the definition, we see that the process has zero mean. Thus,

$$E[u(t)] = 0 \tag{2.11}$$

Furthermore, if we consider the increments, they also have zero means

$$E[u(t) - u(s)] = 0 \tag{2.12}$$

Therefore, the variance of the increment process is

$$E[(u(t) - u(s))^2] = \sigma^2 E[(x(t) - x(s))^2] \tag{2.13}$$

and using (3) of Definition 2.3, we obtain

$$E[(u(t) - u(s))^2] = \sigma^2|t - s| \tag{2.14}$$

for any t and s. Thus, any Wiener process can be formed from a normalized Wiener process.

Let us now consider the process $x(t)$ itself and not the increments. Now, by definition,

$$x(t) = x(t) - x(0) \tag{2.15}$$

Also

$$E[x(t)] = 0 \tag{2.16}$$

The variance of the process $x(t)$ is given by

$$E[x^2(t)] = t \tag{2.17}$$

The correlation function is defined by

$$E[x(t)x(s)] \tag{2.18}$$

Assume that $t > s$. Let

$$x(t) = x(s) + x(t) - x(s) \tag{2.19}$$

Clearly, the increment $x(t) - x(s)$ is independent of $x(s)$—that is, $x(s) - x(0)$—so that

$$E[x(t)x(s)] = s \tag{2.20}$$

if $s < t$. Likewise, when $t < s$, we find the correlation is t. Thus,

$$E[x(t)x(s)] = \min(t, s) \tag{2.21}$$

where min (t,s) is a function equal to the minimum of the time t or s.

The covariance of a normalized Wiener process $E[x(t)x(s)]$ was shown to be min (t,s). Let us go one step further with this process $u(t)$ defined in the example. If we formally take $du(t)/dt$, what is the covariance function of the process? It can be formally written as

$$E\left[\frac{du(t)}{dt}\,\frac{du(s)}{ds}\right] = \frac{\partial^2}{\partial t\,\partial s}\,E[u(t)u(s)] \tag{2.22}$$

where we have interchanged expectation and differentiation; that is, we can obtain it by taking the partial derivatives of the covariance function.

In Figure 3.4(a) we have plotted the covariance function $E[u(t)u(s)]$ for the process. In Figure 3.4(b) we plot

$$\frac{\partial}{\partial t}\,E[u(t)u(s)] \tag{2.23}$$

We see that the value is 1 on the t side of the line $t = s$ and 0 along the s side. There is a discontinuity at $t = s$. Now if we take the partial derivative with respect to s we get 0 everywhere, except for an impulse along the line $t = s$. This would imply an infinite amount of energy, obviously not a realistic

Figure 3.4 White noise process.

assumption. This is figuratively shown in Figure 3.4(*c*). Such a process is called a *white noise process*. Its correlation function is given by

$$E\left[\frac{du(t)}{dt}\,\frac{du(s)}{ds}\right] = \sigma^2\,\delta(t-s) \tag{2.24}$$

Such a process, although theoretically distasteful, allows one to obtain useful engineering results (see Kalman [1]). One method around the dilemma of using the impuse is to introduce the concept of generalized random processes (Yaglom, Appendix II; Gelfand and Vilenkin, pp. 237–302). This approach rests on the theory of distributions of Schwartz and legitimizes the impulse. Some readable introductions to the theory of distributions are those given by Lighthill and Zemanian.

We can now introduce the second independent increment process of interest, the Poisson process. The Poisson process is a step process where the points of discontinuity or steps are at most countable. We shall find that this process parallels the results for the Wiener process quite closely.

DEFINITION 2.4. Let $x(t)$, $t \in [0, \infty]$, be a stochastic process such that

1. for almost all ω sample points, the sample function $x(t, \omega)$ is a step function increasing with jump one and vanishes at $t = 0$, that is, $x(0) = 0$;

2. the probability that the number of jumps is k between time t and s is given by

$$P[x(t) - x(s) = k] = \exp\left[-\lambda|t-s|\right]\frac{(|t-s|)^k\lambda^k}{k!} \tag{2.25}$$

where $\lambda > 0$.

3. the process is an independent increment process.

Such a process is called a *Poisson process*.

4. The discontinuities are of the first kind: that is $x(t-0) \neq x(t) = x(t+0)$ as shown in Figure 3.5.

A typical sample function is shown in Figure 3.6. More general comments and structures for Poisson processes are given in Parzen [1] (Chapter 4). In this definition we defined λ as independent of time. λ may also be a function of time, $\lambda(t)$, and such a process is called a *nonhomogeneous Poisson process*.

A natural extension of the Wiener process for the case of the Poisson process is the generalized Poisson process. As we have seen, the Poisson process $x(t)$ of Definition 2.4 was a simple unit-step process with the steps occurring at Poisson intervals. An immediate generalization is to allow the steps themselves to be random variables. Thus:

DEFINITION 2.5. Let $x(t)$ be a simple Poisson counting process with

Figure 3.5 Discontinuous process

Figure 3.6 Sample function of a Poisson process.

$$P[x(t) - x(s) = k] = \frac{(\lambda|t - s|)^k}{k!} \exp(-\lambda|t - s|) \qquad (2.26)$$

Then

$$y(t) = \sum_{i=1}^{x(t)} a_i u_{-1}(t - \tau_i) \qquad (2.27)$$

Figure 3.7 Sample function of a generalized Poisson process.

is a *generalized Poisson process* where the random variables a_i are independent of $x(t)$ and distributed with density function $p_a(\alpha)$, $u_{-1}(t)$ is the unit-step function and $\{\tau_i\}$ are the arrival times of the Poisson counting process $x(t)$.

Clearly, $y(t)$ is also an independent increment process. Thus, it is also a Markov process. A sample path of $y(t)$ is shown in Figure 3.7 where the process $y(0)$ is zero. A useful generalized Poisson process is the one where the $\{a_i\}$ are independent identically distributed zero mean Gaussian random variables with variance σ_a^2. In that case,

$$E[y(t)] = E\left[\sum_{i=1}^{x(t)} a_i u_{-1}(t-\tau_i)\right] = 0 \tag{2.28}$$

which follows directly from the zero mean a_i. An important property of random variables such as $y(t)$ is its characteristic function, for from it we can obtain the probability density function. In the following example we evaluate the characteristic function for this process.

Example. The characteristic function is defined as

$$M_x(u, t) = E[\exp [jux(t)]] \tag{2.29}$$

Let $y(t)$ be a generalized Poisson process with amplitudes a_i being independent identically distributed zero mean Gaussian random variables with variance σ_a^2. From the definition

$$M_y(u, t) = \sum_{N=0}^{\infty} E\left[\exp\left(j \sum_{i=1}^{N} a_i u\right)\right] P[x(t) = N] \tag{2.30}$$

But $\sum_{i=1}^{N} a_i$ is a zero mean Gaussian random variable with variance $N\sigma_a^2$. Thus,

$$M_y(u, t) = \sum_{N=0} \exp(-\lambda t) \frac{[\exp(-\frac{1}{2}\sigma_a^2 u^2)\lambda t]^N}{N!} \tag{2.31}$$

Factoring out the $\exp(-\lambda t)$ term and factoring within the sum, we obtain

$$M_y(u, t) = \exp(-\lambda t) \sum_{N=0}^{\infty} \frac{[\exp(-\frac{1}{2}\sigma_a^2 u^2)\lambda t]^N}{N!} \tag{2.32}$$

which becomes

$$M_y(u, t) = \exp(\lambda t[\exp(-\tfrac{1}{2}\sigma_a^2 u^2) - 1]) \tag{2.33}$$

The inverse Fourier transform of this gives the probability density function of the random variable $y(t)$. Now, in a similar fashion, we can obtain the characteristic function of the Gaussian random process $x(t)$. This is

$$\begin{aligned} M_x(u, t) &= \int_{-\infty}^{\infty} \exp(juv) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{v^2}{\sigma^2}\right) dv \\ &= \exp[-\tfrac{1}{2}\sigma^2 u^2] \end{aligned} \tag{2.34}$$

The joint characteristic functions for $x(t_1)\cdots x(t_n)$, where the $x(t_i)$ are jointly zero mean Gaussian random variables, follows directly. That is, if $\mathbf{u}$ is an $n \times 1$ vector, then

$$M_{\mathbf{x}}(\mathbf{u}; t_1, \cdots, t_n) = E[\exp(j\mathbf{u}^T\mathbf{x})] \tag{2.35}$$

where

$$\mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} \tag{2.36}$$

and

$$\mathbf{x} = \begin{bmatrix} x(t_1) \\ \vdots \\ x(t_n) \end{bmatrix} \tag{2.37}$$

Let Λ be the covariance matrix, and if the process is zero mean, then

$$M_{\mathbf{x}}(\mathbf{u}; t_1, \cdots, t_n) = \exp(-\tfrac{1}{2}\mathbf{u}^T\Lambda\mathbf{u}) \tag{2.38}$$

This is easily extended to vector processes also. The covariance of the generalized Poisson process is evaluated in the following example.

Example. Let $y(t)$ be a generalized Poisson process given by

$$y(t) = \sum_{i=1}^{x(t)} a_i u_{-1}(t - \tau_i) \tag{2.39}$$

and let the a_i be independent identically distributed Gaussian random variables with zero mean and variance σ_a^2. $x(t)$ is a Poisson step process with

rate λ. We want to evaluate the covariance, which is $E[y(t_1)y(t_2)]$. To evaluate this, we find it convenient to use the characteristic function. Now the joint characteristic function is also easily obtained. That is, we want to evaluate

$$M_y(u_1, t_1; u_2, t_2) = E[\exp\{ju_1 y(t_1) + ju_2 y(t_2)\}] \tag{2.40}$$

Let us assume $t_2 > t_1$. Thus.

$$\begin{aligned} M_y(u_1, t_1; u_2, t_2) = E[&\exp\{j(u_1 + u_2)y(t_1)\} \\ &\cdot \exp\{ju_2[y(t_2) - y(t_1)]\}] \end{aligned} \tag{2.41}$$

But since $y(t)$ is an independent increment process we have

$$M_y(u_1, t_1; u_2, t_2) = E[e^{j(u_1+u_2)y(t_1)}]\, E[e^{ju_2[y(t_2) - y(t_1)]}] \tag{2.42}$$

which from the previous example becomes

$$\begin{aligned} M_y(u_1, t_1; u_2, t_2) = \exp\{\lambda t_1\{\exp[-\tfrac{1}{2}\sigma_a^2(u_1 + u_2)^2] - 1\}\} \\ \cdot \exp\{\lambda(t_2 - t_1)[\exp(-\tfrac{1}{2}\sigma_a^2 u_2^2) - 1]\} \end{aligned} \tag{2.43}$$

The correlation between $y(t_1)$ and $y(t_2)$ is given by

$$\begin{aligned} E[y(t_1)y(t_2)] &= -\frac{\partial^2}{\partial u_1\, \partial u_2} M_y(u_1, t_1; u_2, t_2)\Big|_{u_1 = u_2 = 0} \\ &= \lambda t_1 \sigma_a^2 \end{aligned} \tag{2.44}$$

or, in general.

$$E[y(t)y(s)] = \lambda\sigma_a^2 \min(t, s) \tag{2.45}$$

which is the same correlation one would obtain if $y(t)$ were a Wiener process. Also the process

$$z(t) = \frac{dy(t)}{dt} \tag{2.46}$$

has a correlation function

$$E[z(t)z(s)] = \lambda\sigma_a^2\delta(t - s) \tag{2.47}$$

or it is a white noise process although non-Gaussian. The usefulness of this process is that $z(t)$ is stationary zero mean, has Gaussian amplitudes. and has a power spectrum that is flat. Thus, to second moment properties it is indistinguishable from the Wiener process.

Extensions to the case of nonstationary processes, where λ is not constant but a function of time, are carried out in Problem 3.3. The vector case where $z(t)$ is an $n \times 1$ generalized Poisson process is treated in Problem 3.7. Both the Wiener process and the Poisson process will be used extensively in Chapter 5 in the analysis of the nonlinear estimation problem.

Having defined the independent increment process we will find it worthwhile now to show some of the consequences of its structure. The following

two theorems consider two different types of independent increment processes. The first is the type whose path in time $x(t)$ is continuous. It is then possible for us to show that the increments of those processes are Gaussian: namely, if we let I be the interval $[t_i, t_i + \Delta]$, the random variable

$$x(I) \triangleq x(t_i + \Delta) - x(t_i) \tag{2.48}$$

is a Gaussian random variable. The second type of process is that which undergoes only unit jumps; that is, the process $x(t)$ takes on only a countable number of integer values. It is then possible to show that the $x(I)$ for this independent increment process is a Poisson random variable; that is, the random variable $x(I)$ assumes only integer values, say n. with probability

$$P[x(I) = n] = \frac{(\lambda\Delta)^n}{n!} e^{-\lambda\Delta} \tag{2.49}$$

These two theorems provide us with some important information. In Chapter 5 we will drive a system with an independent increment process. These theorems tell us then that if the process is continuous. then it is Gaussian; if it has jumps, it is Poisson; or if it is composed of both a continuous and discontinuous part, then a Poisson and Gaussian decomposition may be possible. This decomposition theorem is discussed in detail by Ito [2] and by Hida.

Before proving the theorem, let us recall the definiton of the "almost everywhere" concept in probability. We want to show that an event occurs for almost all $\omega \in \Omega$. This implies that if Ω_0 is the set of Ω for which the event does not occur and P is the probability measure on Ω and $\mathcal{B}$ is its σ-field, then $\Omega_0 \in \mathcal{B}$ and $P(\Omega_0) = 0$. That is the sets on which the exception occurs have probability 0. Thus. the event occurs for almost all ω.

THEOREM 2.1

(Gaussian) Let $x(t, \omega)$ be an independent increment process. If $x(t, \omega)$ is continuous in t for almost all ω, then $x(I)$ is a Gaussian variable.

Proof. Let $I = (t_0, t_1)$. Since almost all sample functions are continuous for any $\varepsilon > 0$, there exists a $\delta(\varepsilon) > 0$ such that

$$P[\forall s, t \in I; |t - s| < \delta \text{ implies } |x(t) - x(s)| < \varepsilon] > 1 - \varepsilon$$

This follows directly from the preceding discussion of probability 1 or almost everywhere. Now, for each n, let $t_0 = t_{n_0} < t_{n_1} < \cdots < t_{n_{p_n}} = t_1$ be a subdivision of the interval $[t_0, t_1]$ with

$$0 < t_{n_i} - t_{n_{i-1}} < \delta(\varepsilon_n) \tag{2.50}$$

where ε_n approaches 0 from above. Let us define the increment variable

$$x_{n_i} = x(t_{n_i}) - x(t_{n_{i-1}}) \tag{2.51}$$

Define the sum of such variables as

$$x = x(I) = \sum_{k=1}^{P_n} x_{n_k} \tag{2.52}$$

where P_n denotes the total partitioning. Then define $x'_{n_k} = x_{n_k}$ if $|x_{n_k}| < \varepsilon_n$ and 0 otherwise. That is, let

$$x'_{n_k} = \begin{cases} x_{n_k}; & |x_{n_k}| < \varepsilon_n \\ 0; & |x_{n_k}| \geqq \varepsilon_n \end{cases} \tag{2.53}$$

Put

$$x_n = \sum_{k=1}^{P_n} x'_{n_k} \tag{2.54}$$

Then, from the assumption of continuity,

$$P(x = x_n) > 1 - \varepsilon_n \tag{2.55}$$

That is, x_n approaches x in probability. Since the x_{n_k} are independent, so are the x'_{n_k}. Therefore,

$$E[e^{jux}] = \lim_{n\to\infty} E[e^{jux_n}] = \lim_{n\to\infty} \prod_{k=1}^{P_n} E[e^{jux'_{n_k}}] \tag{2.56}$$

Let

$$M_{n_k} = E[x'_{n_k}] \tag{2.57}$$

and

$$V_{n_k} = E[(x'_{n_k} - E[x_{n_k}])^2] \tag{2.58}$$

Then, by the independence hypothesis, we know that the sum of the means is

$$M_n = \sum_{k=1}^{P_n} M_{n_k} \tag{2.59}$$

and the sum of the variances is

$$V_n = \sum_{k=1}^{P_n} V_{n_k} \tag{2.60}$$

the variance of the sum. Then, by definition of x'_{n_k} as given in (2.53), we have

$$|E[x'_{n_k}]| = |M_n| \leqq \varepsilon_n \tag{2.61}$$

and

$$E[(x'_{n_k} - E[x'_{n_k}])^2] \leqq E[(2\varepsilon_n)^2] \leqq 4\varepsilon_n^2 \tag{2.62}$$

which also follows directly from (2.53) by the way x'_{n_k} is defined. Therefore,

$$\begin{aligned} E[e^{jux}] &= \lim_{n\to\infty} e^{juM_n} \prod_{k=1}^{P_n} E[e^{ju(x'_{n_k} - Mn_k)}] \\ &= \lim_{n\to\infty} e^{juM_n} \prod_{k=1}^{P_n} \left[1 - \frac{u^2}{2} V_{n_k} + o(\varepsilon_n)\right] \end{aligned} \tag{2.63}$$

This should follow simply enough if we realize that all higher-order central moments are of the order

$$E[(x_{n_i} - M_{n_i})^q] \leqq 2^q \varepsilon_n^q \tag{2.64}$$

and if ε_n is small enough, then indeed these higher moments vanish quite quickly. We can easily prove that if we let

$$z_n = \sum_{i=1}^{P_n} z_{n_i} \tag{2.65}$$

where the sum is absolutely bounded, then as z_n approaches z as n approaches ∞ we have

$$\lim_{n\to\infty} \prod_{i=1}^{P_n} [1 - z_{n_i}] = e^{-z} \tag{2.66}$$

For our problem

$$\max_k |V_{n_k}| \leqq 4\varepsilon_n^2 \to 0$$

and

$$\sum_{k=1}^{P_n} V_{n_k}[1 + o(\varepsilon_n)] \to V \tag{2.67}$$

and each $V_{n_k} \geqq 0$, so that using the previous limiting result

$$\lim_{n\to\infty} \prod_{k=1}^{P_n} \left[1 - \frac{u^2}{2} V_{n_k}(1 + o(t_n))\right] = \exp\left(-\frac{u^2}{2} V\right) \tag{2.68}$$

Therefore, the characteristic function is given by

$$M_x(ju) = \exp\left(jum - \frac{u^2 V}{2}\right) \tag{2.69}$$

since

$$\lim_{n\to\infty} e^{jum_n} = e^{jum} \tag{2.70}$$

The existence of this limit can be shown to be true. If it were not, the expected value would be 0, which would yield a contradiction. Thus, (2.69) shows that the process has a characteristic function that converges to a Gaussian form, so that the processes converge to a Gaussian in distribution. This means that the *distribution* of the *interval* variable $x(I)$ is Gaussian. It does not mean that x evaluated at some point t in the interval is Gaussian. This should be intuitively obvious, since $x(I)$ is becoming a sum of an infinite number of random variables. This is shown in Figure 3.8. ▮

We should also note that with this theorem we could have relaxed the assumption of Gaussianess in defining the Wiener process and merely required zero mean and variance of the increments as $|t - s|$. But we would then have

Figure 3.8 Gaussian approximation.

had to add the assumption that the process be almost everywhere continuous. By Theorem 2.1 we have shown that such an almost everywhere continuous process $x(t)$ is Gaussian. In the next section we shall show that the Wiener process is almost everywhere continuous, thus complementing this proof. We can now prove a similar theorem for the Poisson process. These two theorems provide the basis of a decomposition of an arbitrary independent increment process.

To prove a theorem similar to the previous for the Poisson process it is first necessary to intorduce the definition of a Levy process.

DEFINITION 2.5.: A stochastic process $x(t)$ is a *Levy process* if:

i. $x(0)=0$

ii. $x(t)$ is an independent increment process

iii. $x(t)$ has no fixed discontinuities, that is

$$\lim_{t \to s} P\left[\left| x(t)-x(s)\right| > 0\right]=0 \tag{2.71}$$

iv. The sample paths of $x(t)$ have discontinuities of the first kind.

The following theorem can then be stated.

THEOREM 2.2

(Poisson) Let $x(t)$ be a Levy process. If almost all sample functions are step functions with jump 1, then $x(t)$ is a Poisson variable.

Proof. From the continuity in probability of $x(t)$,

$$\sup P[|x(t) - x(s)| \geqq 1] \to 0 \qquad \text{as} \qquad n \to \infty, \ |t - s| < n^{-1} \tag{2.72}$$

For each n let

$$t_0 = t_{n_0} < t_{n_1} < \cdots < t_{n_{p_n}} = t_1 \tag{2.73}$$

and let

$$t_{n_i} - t_{n_{i-1}} \leqq \frac{1}{n} \tag{2.74}$$

be a subdivision of $[t_0, t_1]$ and let

$$x_{n_i} = x(t_{n_i}) - x(t_{n_{i-1}}) \tag{2.75}$$

and

$$x'_{n_i} = \begin{cases} 1 & \text{if} \quad x_{n_i} \geqq 2 \\ x_{n_i} & \text{if} \quad x_{n_i} = 0, 1 \end{cases} \tag{2.76}$$

Put

$$x_n = \sum x'_{n_i} \tag{2.77}$$

Then since $P[x_n \to x] = 1$ where $x = x(t_1) - x(t_0) = x(I)$ as given in (2.52) we have for the characteristic function of the process

$$\begin{aligned} E[e^{jux}] &= \lim_{n\to\infty} E[e^{jux_n}] = \lim_{n\to\infty} \prod_{k=1}^{p_n} E[e^{jux'_{n_i}}] \\ &= \lim_{n\to\infty} \prod_{k=1}^{p_n} [(1 - P_{n_i}) + P_{n_i} e^{ju}] \end{aligned} \tag{2.78}$$

where

$$P_{n_i} = P[x_{n_i} \geqq 1] = P[x'_{n_i} = 1] \tag{2.79}$$

Now, continuing, we obtain

$$\begin{aligned} E[e^{jux}] &= \lim_{n\to\infty} \prod_{k=1}^{p_n} [1 - P_{n_i}(1 - e^{ju})] \\ &= \lim_{n\to\infty} \left[\prod_{k=1}^{p_n} [e^{-P_{n_i}(1 - e^{ju})}] + o\left(\frac{1}{n}\right) \right] \\ &= \lim_{n\to\infty} e^{-P_n(1 - e^{ju})} = e^{-(1 - e^{ju}) \lim P_n} \end{aligned} \tag{2.80}$$

where

$$P_n = \sum_{k=1}^{p_n} P_{n_i} \tag{2.81}$$

and $P_n \to \lambda T$ as $n \to \infty$, where $T = |t - s|$ and λ is the arrival rate of the process. Therefore,

$$E[e^{jux}] = e^{-\lambda T(1 - e^{ju})} \tag{2.82}$$

which is the characteristic function for a homogeneous Poisson process. ∎

As we stated, these two preceding theorems provide the basis for the decomposition of independent increment processes. We saw that if an indepen-

dent increment process were continuous for almost all $\omega \in \Omega$, it was a Gaussian process. If we further required that it have a covariance of the form $|t - s|$ for the increments and have $x(0) = 0$, then it would be a Wiener process. We will in the next section reverse the argument and show that a Wiener process is continuous almost everywhere. Thus we could say that an independent increment process with $x(0) = 0$ and possessing a covariance $|t - s|$ is a Wiener process if and only if it is continuous for almost all $\omega \in \Omega$. This continuity will be of prime importance in characterizing the Wiener process. We should also carefully note the structure of the Poisson process. Hida contains a great deal on this process. For example, the derivative of the Wiener process and that of the zero mean generalized Poisson process both have the same correlation function, that is an impulse. Thus, there exists an isometry between systems driven by these two processes.

3.3 PROPERTIES OF THE WIENER PROCESS

The Wiener process—or, as it is sometimes called, Brownian motion—is of vital importance in the modeling of stochastic systems. Recall that the process was such that it had independent increments and that the increments were normally distributed. The name "Brownian motion" is derived from the basic observations the botanist Robert Brown made on the motions of particles suspended in a liquid. A historical account of the developments that led to the recognition of this process is given by Nelson. The abstract process of particles in Brownian motion is called a Wiener process because it was Norbert Wiener who formulated much of the present theory. Masani relates the story of how Wiener, while looking out of his M.I.T. office at the Charles River, first thought of describing the random motion of the waters by means of a stochastic process with independent increments. In our model we will find it useful to use a Wiener process as a stochastic driving function on a dynamic system.

This section will set out to show three facts concerning the Wiener process. First, we shall show that such a process is uniformly continuous. By this, we will mean that if $x(t)$ is a Wiener process and $x(t, \omega)$ is a sample path, then for almost all $\omega \in \Omega$, $x(t, \omega)$ will be uniformly continuous. This is a valuable property since it will mean that when driving a system with such a process, the output may also be continuous.

The second main fact is that the Wiener process is not of bounded variation. This will be directly related to the fact that the derivative of such a process will not exist. It will force us in the following section to carefully define integrals of such processes.

The third fact requires the introduction of process one more concept. It is the idea of a *martingale*. A martingale is a stochastic whose expectation at

some time t conditioned on a past history of the process is equal to a point in the past history. For example, if we are given that at time t we are at $x(t)$ and that the process $x(s)$ is a martingale, then the expected value of our position at time t_1, $x(t_1)$, given $x(t)$, is $x(t)$. Thus, with this concept we show that every random process that is a continuous martingale and has a suitable variance is a Wiener process. It is also trivial to show the converse, that a Wiener process is a martingale. This is a most powerful result. For example, in detection theory one has a likelihood ratio. Such a ratio can be shown to be a continuous martingale. Thus, if the likelihood ratio has the proper variance, it is a Wiener process.

We will now need the following three lemmas in order to prove continuity.

LEMMA 3.1. The following inequality holds for all $x > 0$:

$$\int_x^\infty e^{-\xi^2/2}\,d\xi \leqq \frac{1}{x}\int_x^\infty \xi e^{-\xi^2/2}\,d\xi = \frac{e^{-x^2/2}}{x} \tag{3.1}$$

The proof of the above lemma is trivial and will be omitted.

LEMMA 3.2. Let $y_1, \cdots, y_n$ be a set of mutually independent random variables and let

$$x_k = y_1 + \cdots + y_k \tag{3.2}$$

If the differences $x_n - x_1$, $x_n - x_2$, $x_n - x_3, \cdots$ have symmetric distributions, then

$$2P[x_n \geqq \lambda + 2\varepsilon] - 2\sum_{j=1}^{n} P[y_j \geqq \varepsilon] \leqq P[\max_{j<n} x_j \geqq \lambda] \leqq 2P[x_n \geqq \lambda] \tag{3.3}$$

for every λ, $\varepsilon > 0$. The right half of the above is valid if each $x_n - x_k$ is zero mean but not symmetrically distributed.

Proof. The proof of this lemma is in Doob (p. 107). ∧

LEMMA 3.3. Let $x(t)$ be a Wiener process on $[0, T]$, then

$$\begin{aligned} &P[\sup_{0\leq t\leq T}[x(t) - x(0)] \geqq \lambda] \\ &= 2P[x(T) - x(0) \geqq \lambda] < \frac{\sigma}{\lambda}\sqrt{\frac{2T}{\pi}}\; e^{-\lambda^2/2\sigma^2 T} \end{aligned} \tag{3.4}$$

Proof. See Doob (p. 392). ∧

With these three lemmas we can now show that the Wiener process is continuous. To do so we must assume that the process has a separable version (see Billingsley). This theorem is closely related to the result in the preceding section for continuous independent increment processes. We show in tihs theorem that the Wiener process is uniformly continuous. By uniform continuity we imply that the probability of any discontinuity is vanishingly small. We now proceed with the theorem.

THEOREM 3.1.

Almost *all sample functions* of a (*separable*) Wiener process are uniformly *continuous* in $t \in [0, \infty]$.

Proof. Choose intervals

$$t \in I_{j,N}$$

where

$$I_{j,N} = \left\{ t: \frac{j-1}{N} \leqq t \leqq \frac{j+1}{N} \right\} \qquad \text{and} \qquad j = 1, \cdots, N^2 \tag{3.5}$$

Such intervals are shown in Figure 3.9. We should note that the maximum length of the interval is

$$\frac{N^2+1}{N} \tag{3.6}$$

so that the total length is increasing but the width of the intervals is decreasing. Now we want to show that for any $t \in I_{j,N}$.

Figure 3.9 Examples of $I_{j,N}$

$$\left| x(t) - x\left(\frac{j}{N}\right) \right| \leqq N^{-1/4} \tag{3.7}$$

except for a set of probability 0. Note that $I_{j,N}$ is centered about the point j/N and goes $\pm(1/N)$ on the sides. We now want to show that

$$P\left[\sup_{\substack{|t-j/N| \leqq 1/N \\ 1 \leqq j \leqq N^2}} \left| x(t) - x\left(\frac{j}{N}\right) \right| \geqq N^{-1/4} \right] \to 0 \tag{3.8}$$

with increasing N. This above expression is

$$\begin{aligned} & P\left[\sup_{\substack{|t-j/N| \leqq 1/N \\ 1 \leqq j \leqq N^2}} \left| x(t) - x\left(\frac{j}{N}\right) \right| \geqq N^{-1/4} \right] \\ &= P\Big[\sup\Big\{ \sup_{|t-1/N| \leqq 1/N} \left| x(t) - x\left(\frac{1}{N}\right) \right| ; \\ &\qquad \sup_{|t-2/N| \leqq 1/N} \left| x(t) - x\left(\frac{2}{N}\right) \right| ; \cdots ; \\ &\qquad \sup_{|t-N| \leqq 1/N} |x(t) - x(N)| \Big\} \geqq N^{-1/4} \Big] \end{aligned} \tag{3.9}$$

But

$$P[\sup[y_1 \cdots y_{N^2}] \leqq N^{-1/4}] \leqq \sum_{i=1}^{N^2} P[y_i \leqq N^{-1/4}] \tag{3.10}$$

so that if we define the random variables y_i as

$$y_i = \sup_{|t-i/N| \leqq 1/N} \left| x(t) - x\left(\frac{i}{N}\right) \right| \tag{3.11}$$

we have a bound for (3.10) of the form

$$\begin{aligned} & P\left[\sup_{\substack{|t-j/N| \leqq 1/N \\ 1 \leqq j \leqq N^2}} \left| x(t) - x\left(\frac{j}{N}\right) \right| > N^{-1/4} \right] \\ & \leqq \sum_{i=1}^{N^2} P\left[\sup_{|t-i/N| \leqq 1/N} \left| x(t) - x\left(\frac{i}{N}\right) \right| \leqq N^{-1/4} \right] \end{aligned} \tag{3.12}$$

But since the process has increments whose distributions are stationary in time we have

$$\begin{aligned} & P\left[\sup_{|t-i/N| \leqq 1/N} \left| x(t) - x\left(\frac{i}{N}\right) \right| \leqq N^{-1/4} \right] \\ &= P\left[\sup_{|t-j/N| \leqq 1/N} \left| x(t) - x\left(\frac{j}{N}\right) \right| \leqq N^{-1/4} \right] \end{aligned} \tag{3.13}$$

which is true for all i,j in the interval. Choose $i,j = 1$, and then (3.12) becomes

$$P\left[\sup_{\substack{|t-j/N|\le 1/N \\ 1\le j\le N^2}} \left|x(t) - x\left(\frac{j}{N}\right)\right| \ge N^{-1/4}\right]$$
$$< N^2 P\left[\sup_{|t-1/N|<1/N} \left|x(t) - x\left(\frac{1}{N}\right)\right| \le N^{-1/4}\right] \tag{3.14}$$

But from the previous lemma we know that

$$P\left[\sup_{|t-1/N|\le 1/N} \left|x(t) - x\left(\frac{1}{N}\right)\right| \le N^{-1/4}\right]$$
$$\le \frac{\sigma}{(N^{-1/4})}\sqrt{\frac{2(2/N)}{\pi}}\exp\left[-\frac{(N^{-1/4})^2}{2\sigma^2(2/N)}\right] = 2\sigma\sqrt{\frac{1}{\pi N^{1/2}}}\exp\left(-\frac{N^{1/2}}{4\sigma^2}\right) \tag{3.15}$$

Now, substituting 3.3 into (3.14), we obtain

$$P\left[\sup_{\substack{|t-j/N|\le 1/N \\ 1\le j\le N^2}} \left|x(t) - x\left(\frac{j}{N}\right)\right| \ge N^{-1/4}\right] \le \frac{2\sigma N^{7/4}}{\sqrt{\pi}}\exp\left(-\frac{N^{1/2}}{4\sigma^2}\right) \tag{3.16}$$

And, indeed, as $N \to \infty$, the probability that the difference between points that are getting closer and closer exceeding 0 is going to 0. Thus, except for a set of probability 0 the process is continuous on $[0, \infty)$ everywhere. ∎

We now want to show that although the process is continuous uniformly, it may traverse an infinitely long path. This concept is best presented by showing that the process is not of bounded variation. This concept is defined as follows.

DEFINITION 3.1. Let f be a function on $\mathbf{R}^1$. The *total variation* T_f of f at point x is defined by

$$T_f(x) = \sup \sum_{j=0}^{N} |f(x_{j+1}) - f(x_j)| \tag{3.17}$$

for all values of $x \in (-\infty, \infty)$ where the supremum is taken over all N and over all choices of $\{x_j\}$ such that

$$-\infty < x_0 < x_1 < \cdots < x_n = x$$

If T_f is a function of bounded variation, then if $x < y$,

$$0 \le T_f(x) \le T_f(y) \le \infty \tag{3.18}$$

and

$$V(f) = \lim_{x\to\infty} T_f(x) \tag{3.19}$$

exists and is finite. If this is so, then we say that f is of *bounded variation* (Rudin [1], pp. 160–161; Kestleman, pp. 185–187).

A function is also called of bounded variation on an interval $[a, b] \in \mathbf{R}^1$ if

$$T_f(a, b) = \sup \sum_{j=0}^{N} |f(x_{j+1}) - f(x_j)| \tag{3.20}$$

is finite (see Gikhman and Skorokhod, p. 279). It is a simple matter to show that if $b > a$.

$$T_f(a, b) = T_f(b) - T_f(a) \tag{3.21}$$

so that if f is of bounded variation on $(-\infty, b]$, it is of bounded variation on any subinterval. To be complete a function is of bounded variation if and only if it is of bounded variation for any subinterval [a,b]$\subset \mathbf{R}_1$. This implies that if the interval $[a, b]$ is replaced by $[t, t+\Delta]$, then $T_f(t, t+\Delta)$ is finite. Thus, if f is of bounded variation, f can be differentiated (e.g., the derivative exists: see Rudin [2], Chapter 8). As we shall see in the next section, the common Riemann-Steiltjes integral is defined only for functions of bounded variation. Thus, integrals with respect to Wiener processes (or *measures*, as they are also called) do not exist in the ordinary sense.

We can now consider the Wiener process and show that it is *not* of bounded variation and thus is not differentiable.

THEOREM 3.2

Let Δ be the interval $[t, u]$ and let the partition be given by $t = s_0 < s_1 < \cdots < s_n = u$, and let $\delta(\Delta) = \max_i (s_{i+1} - s_i)$. Then the following hold for the Wiener process, $x(t)$:

First,

$$I_2(\Delta) = \sum_i (x(s_{i+1}) - x(s_i))^2 \to (u - t) \tag{3.22}$$

in quadratic mean as $\delta(\Delta) \to 0$.

Second,

$$I(x, t, u) = \sup \sum_i |x(s_{i+1}) - x(s_i)| = \infty \tag{3.23}$$

in quadratic mean; that is, the Wiener process is not of bounded variation.

Proof. We shall first prove (3.22):

$$\begin{aligned} E[I_2(\Delta)] &= \sum_i E[(x(s_{i+1}) - x(s_i)^2] \\ &= \sum_i (s_{i+1} - s_i) = (u - t) \end{aligned} \tag{3.24}$$

To show that we have convergence in quadratic mean, we must now show that

$$\lim_{\Delta \to 0} E[(I_2(\Delta) - (u - t))^2] \to 0 \tag{3.25}$$

Substituting for $I_2(\Delta)$, we obtain

$$E[[I_2(\Delta) - (u - t)]^2] = E[I_2^2(\Delta)] - 2E[I_2(\Delta)](u - t) + (u - t)^2 \quad (3.26)$$

But we already calculated $E[I_2(\Delta)]$, so that the above becomes

$$E[[I_2(\Delta) - (u - t)]^2] = E[I_2^2(\Delta)] - (u - t)^2 \quad (3.27)$$

This now requires the second moment of $I_2^2(\Delta)$

$$\begin{aligned} E[I_2^2(\Delta)] &= E\{\sum_i [x(s_{i+1}) - x(s_i)]^2 \sum_j [x(s_{j+1}) - x(s_j)]^2\} \\ &= \sum_i E[(x(s_{i+1}) - x(s_i))^4] \\ &+ \sum_i \sum_{i \neq j} \sum_j E[(x(s_{i+1}) - x(s_i)^2] \\ &\cdot E[(x(s_{j+1}) - x(s_j))^2] \end{aligned} \quad (3.28)$$

Now by the Gaussian nature of $x(t)$ process the fourth moment is equal to three times the variance squared. Thus,

$$\begin{aligned} E[I_2^2(\Delta)] &= \sum_i 3(s_{i+1} - s_i)^2 + \sum_i \sum_{i \neq j} \sum_j (s_{i+1} - s_i)(s_{j+1} - s_j) \\ &= 2\sum_i (s_{i+1} - s_i)^2 + \sum_i \sum_j (s_{i+1} - s_i)(s_{j+1} - s_j) \\ &= 2\sum_i (s_{i+1} - s_i)^2 + (u - t)^2 \end{aligned} \quad (3.29)$$

Thus,

$$E\{[I_2(\Delta) - (u - t)]^2\} = 2\sum_i (s_{i+1} - s_i)^2 \quad (3.30)$$

But recall that

$$\delta(\Delta) = \max(s_{i+1} - s_i) \quad (3.31)$$

Therefore,

$$E\{[I_2(\Delta) - (u - t)]^2\} \leqq 2\delta(\Delta) \sum_i (s_{i+1} - s_i) \leqq 2\delta(\Delta)(u - t) \quad (3.32)$$

Now, as $\delta(\Delta) \to 0$, the factor $I_2(\Delta)$ then approaches $(u - t)$ in quadratic mean which proves the first part of the theorem. We now want to show that the process is *not* of bounded variation. Now, from (3.22), we know that we can always find a sequence $I_2(\Delta)$ that will approach $(u - t)$ as we decrease $\delta(\Delta)$ with probability 1. (3.23) implies that the sup of the sum or its least upper bound goes to infinity. Now,

$$\begin{aligned} I(x, t, u) &= \sup \sum_i |x(s_{i+1}) - x(s_i)| \geqq \sum_i |x(s_{i+1}) - x(s_i)| \\ &\geqq \frac{\sum |x(s_{i+1}) - x(s_i)|^2}{\max_i |x(s_{i+1}) - x(s_i)|} \end{aligned} \quad (3.33)$$

Now, with probability 1 the numerator of the above function goes to $(u - t)$ (see Doob, p. 395), but the denominator goes to 0 because of the continuity

of the Wiener process as shown in the previous theorem. Thus, $I(x,t,u) \to \infty$ as $\delta(\Delta) \to 0$. ∎

This, therefore, implies that integrals of the form

$$\int_t^u dx(t)$$

where $x(t)$ is a Wiener process, are ill defined even though $x(t)$ is a continuous function. There are ways to circumvent this conclusion with either the use of distribution theory or generalized functions (Gelfand and Vilenkin, pp. 237–302) or the Ito integral (Doob [2], pp. 425–451; McKean [2], p. 21). We will in the next section pursue the latter course.

An immediate extension of the previous theorem applies to the increment of the Wiener process. For this extension we want to consider a vector Wiener process $\mathbf{u}(t)$ where the components of $\mathbf{u}(t)$ are correlated. That is,

$$E[d\mathbf{u}(t)\ d\mathbf{u}^T(t)] = \mathbf{Q}\ dt \tag{3.34}$$

where $\mathbf{Q}$ is an $n \times n$ positive definite matrix. What now results is that from the previous theorem not only is the expected value of this product equal to $\mathbf{Q}\ dt$ ~~(l.i.m. and wpl)~~.

COROLLARY 3.1. Let $\mathbf{u}(t)$ be an $n \times 1$ Wiener process with covariance $\mathbf{Q}$. Then,

$$d\mathbf{u}(t)d\mathbf{u}^T(t) = \mathbf{Q}\ dt \tag{3.35}$$

in quadratic mean (also wpl).

Proof. The proof is an immediate consequence of the preceding theorem since the theorem held for any arbitrary interval Δ. Thus, let Δ be the interval dt. Furthermore, the generalization to the vector process is easily done and is left as an exercise. ∎

This result is extremely important and will be used in Chapter 5 for the development of the filtering equations.

We have previously mentioned the term "martingale," and we will have need for its use greatly in the sections to come. We will find that the Wiener process is a martingale, that certain integrals necessary for the solution of stochastic differential equations are martingales, and that the likelihood ratios used in statistical detection theory are martingales. Doob [2] (Chapter 7) spends a great deal of time discussing them and elucidating many of their extremely useful properties, and we shall devote a minimal amount of space to indicate their worth.

DEFINITION 3.2. A *martingale* is a stochastic process $\{x(t), t \in T\}$ for which

$$E[|x(t)|] < \infty$$

for all admissible t and

$$x(t_n) = E\left[x(t_{n+1})\middle|x(t_n)\cdots x(t_0)\right] \tag{3.36}$$

where $t_{n+1} > t_n > \cdots > t_1 > t_0$.

The term "martingale" is derived from the French game in which a player doubles his bet everytime he loses. Feller [2, pp. 210–215] gives some examples of martingales other than those to which we shall turn our interest. If $x(t)$ is a random process over the interval T, then (3.36) becomes

$$x(t) = E[x(s)|\mathscr{F}_t]; \qquad s, t \in T \tag{3.37}$$

where $\mathscr{F}_t$ is the minimum sub σ-field generated by $\{x_u : u \leqq t\}$.

The martingale property is very important and we shall exploit it in Chapters 4 and 5 to further develop the ideas of conditional expectation. The first thing to note is that the Wiener process is a martingale. This follows directly from the fact that if $x(t)$ is a Wiener process, then since $x(s)$ can be written as $x(t) + x(s) - x(t)$, we have

$$E[x(s)|\mathscr{F}_t] = E[x(s) - x(t)|\mathscr{F}_t] + E[x(t)|\mathscr{F}_t] \tag{3.38}$$

But since $x(t)$ is an independent increment process of zero mean, the first expectation on the right vanishes and the second equals $x(t)$, which yields the desired result. We now want to show the converse statement—that is, if $x(t)$ is a martingale and if it is continuous, then it is a Wiener process if it has the proper covariance.

THEOREM 3.3

Let the process $x(t)$ be a martingale and suppose that almost all sample functions are continuous. Assume that

$$E[x^2(t)] < \infty; \qquad a \leqq t \leqq b \tag{3.39}$$

and that for each t, s such that $t > s$

$$E\left[[x(t) - x(s)]^2 \middle| \mathscr{F}_s\right] \tag{3.40}$$

with probability 1. Then $x(t)$ is a Wiener process.

Proof. Let $x(t)$ be defined on $[0, T)$ and let $x(0) = 0$. This can be done arbitrarily without changing the process. The method of the proof is to obtain the characteristic function for an arbitrary set of increments of the process and from this to show that the increments are independent and Gaussian. To this end, let $[0, T)$ be divided into m arbitrary nonoverlapping subintervals $\Delta_k = [t_{k-1}, t_k)$ where $t_k > t_{k-1}$. Let $\delta_k = t_k - t_{k-1}$ be the length of these intervals and let the characteristic function be given as

$$E\left[\exp\left\{j\sum_{k=1}^{m} u_k[x(t_k) - x(t_{k-1})]\right\}\right] \tag{3.41}$$

Now divide Δ_m into n equal subintervals and let $\tau(\omega)$ be the first value of $t \in \Delta_m$ such that

$$\max_{\substack{t_{m-1} \leq s_1,\, s_2 \leq t_m \\ |s_1 - s_2| < \delta_m/n}} |x(s_2) - x(s_1)| = \varepsilon \tag{3.42}$$

or t_m if there is no such t (e.g., $t_m = T$). Define the process $\bar{x}(t)$ as follows:

$$\bar{x}(t) = \begin{cases} x(t) & t \leq \tau(\omega) \quad t \in \Delta_m \\ x(\tau(\omega)) & t > \tau(\omega) \quad t \in \Delta_m \end{cases} \tag{3.43}$$

This process is defined so that all its increments are less than ε. Note that as n increases because of the continuity of the process, $\bar{x}(t)$ should converge to $x(t)$. It can also be shown that if $x(t)$ is a martingale, so too is $\bar{x}(t)$. Now define the random variables y_l as

$$y_l = \bar{x}\left(t_{m-1} + \frac{l}{n}\,\delta_m\right) - \bar{x}\left(t_{m-1} + \frac{l-1}{n}\,\delta_m\right) \tag{3.44}$$

and note that $\sum_{l=1}^{n} y_l$ equals $\bar{x}(t_m) - \bar{x}(t_{m-1})$. Also note that $|y_l| \leq \varepsilon$ for all l. Now, since $\bar{x}(t)$ is a martingale, we also have

$$\begin{aligned} E[y_1] &= E\left[\bar{x}\left(t_{m-1} + \frac{\delta_m}{n}\right) - \bar{x}(t_{m-1})\right] \\ &= E\left\{E\left[\bar{x}\left(t_{m-1} + \frac{\delta_m}{n}\right) - \bar{x}(t_{m-1}) \,\middle|\, \bar{x}(t_{m-1})\right]\right\} = 0 \end{aligned} \tag{3.45}$$

Similarly, we can show that

$$E[y_l | y_1, \cdots, y_{l-1}] = 0 \tag{3.46}$$

Also, we have—letting $\bar{x}_2 = \bar{x}(t_{m-1} + \delta_m/n)$, $\bar{x}_1 = \bar{x}(t_{m-1})$—

$$\begin{aligned} E[y_1^2] &= E[(\bar{x}_2 - \bar{x}_1)^2] \\ &= E[\bar{x}_2^2] - 2E[\bar{x}_2 x_1] + E[\bar{x}_1^2] \\ &= E[\bar{x}_2^2] - 2E[E[\bar{x}_2 \bar{x}_1 | \bar{x}_1]] + E[\bar{x}_1^2] \\ &= E[\bar{x}_2^2] - E[\bar{x}_1^2] \end{aligned} \tag{3.47}$$

But we can easily show that $E[\bar{x}_2^2]$ equals min (t_2, τ) where t_2 equals $t_{m-1} + \delta_m/n$ and $E[\bar{x}_1^2]$ equals min (t_1, τ) where t_1 equals t_{m-1}. Thus,

$$E[y_1^2] = \min(t_2, \tau(\omega)) - \min(t_1, \tau(\omega)) \tag{3.48}$$

It can also easily be shown that

$$E[y_1^2] < t_2 - t_1 = \frac{t_m - t_{m-1}}{n} \tag{3.49}$$

Thus, we have

$$E[y_1^2] \triangleq \sigma_1^2 < \frac{t_m - t_{m-1}}{n} \tag{3.50}$$

and

$$E[y_i^2 | y_1 \cdots y_{i-1}] \triangleq \sigma_i^2 < \frac{t_m - t_{m-1}}{n} \tag{3.51}$$

Now we want to show that the increment $\tilde{x}(t_m) - \tilde{x}(t_{m-1})$ is independent of all the preceding increments and, furthermore, that it is Gaussian. To this end, we write the characteristic function as follows:

$$\begin{aligned} &E\left[\exp\left\{j\sum_{k=1}^{m-1} u_k[\tilde{x}(t_k) - \tilde{x}(t_{k-1})] + ju_m[\tilde{x}(t_m) - \tilde{x}(t_{m-1})]\right\}\right] \\ &= E[\alpha(m-1)\exp\{ju_m[\tilde{x}(t_m) - \tilde{x}(t_{m-1})]\}] \end{aligned} \tag{3.52}$$

where $\alpha(m-1)$ is appropriately defined. Define

$$\tilde{x}_{i/n} = \sum_{k=1}^{i} y_k \tag{3.53}$$

with $\tilde{x}_{0/n}$ equal to zero and $\tilde{x}_{n/n}$ equal to $\tilde{x}(t_m) - \tilde{x}(t_{m-1})$. Also, $\tilde{x}_{i/n}$ equals $\tilde{x}_{i-1/n} + y_i$. Then, for any i, we have

$$\begin{aligned} &E[\alpha(m-1)\exp(ju_m\tilde{x}_{i/n})] \\ &= E[\alpha(m-1)\exp(ju_m\tilde{x}_{(i-1)/n})E[\exp(ju_m y_i)|y_1 \cdots y_{i-1}]] \\ &= E\left[\alpha(m-1)\exp(ju_m\tilde{x}_{(i-1)/n})\exp\left\{-\sigma_i^2\frac{u_m^2}{2}[1 + o(\Delta_m)]\right\}\right] \end{aligned} \tag{3.54}$$

The last equality follows by expanding $\exp(ju_m y_i)$ in its series form and recognizing that because of its martingale property it is zero mean and because of its continuity property the higher-order terms in the expansion are $o(\Delta_m)$. We now bound the characteristic function in the following manner:

$$\begin{aligned} &\left|E[\alpha(m-1)\exp(ju_m\tilde{x}_{i-1/n})] - E\left[\alpha(m-1)\exp(ju_m\tilde{x}_{(i-1)/n})\exp\left(-\frac{u_m^2}{2n}\tilde{\delta}_m\right)\right]\right| \\ &= \left|E\left[\alpha(m-1)\exp\left(ju_m\tilde{x}_{(i-1)/n} - \frac{u_m^2}{2n}\tilde{\delta}_m\right)\right.\right. \\ &\qquad \left.\left.\left[\exp\left(-\frac{u_m^2}{2}\{\sigma_i^2[1 + o(\Delta_m)] - \tilde{\delta}_m/n\}\right) - 1\right]\right]\right| \\ &\leqq E\left[\left|\exp\left[\frac{u_m^2}{2}\left(\frac{\tilde{\delta}_m}{n} - \sigma_i^2\right) + \sigma_i^2 o(\Delta)\right] - 1\right|\right] \\ &\leqq O(\Delta_m)[\delta_m/n - E[y_i^2]] + \frac{o(\Delta_m)}{n} \end{aligned} \tag{3.55}$$

where $O(\Delta_m)$ is any expression remaining bounded as $o(\Delta_m)$ vanishes. Now, by multiplying the above inequality by $\exp(\tilde{\delta}_m/2n)$, we can obtain a similar inequality:

$$\left| E[\alpha(m-1)\exp(ju_m\bar{x}_{i/n})]\exp\left(\frac{u_m^2}{2n}\bar{\delta}_m\right) \right.$$
$$\left. - E[\alpha(m-1)\exp(ju_m\bar{x}_{(i-1)/n})]\exp\frac{(i-1)u_m^2\bar{\delta}_m}{2n} \right|$$
$$\leqq O(\Delta_m)\left[\delta_m/n - E[y_i^2]\right] + \frac{o(\Delta_m)}{n} \tag{3.56}$$

Now, using the above inequality, we can add them up to show that

$$\left| E[\alpha(m-1)\exp\{ju_m[\bar{x}(t_m) - \bar{x}(t_{m-1})]\}]\exp\left(\frac{u_m^2\delta_m}{2}\right) - E[\alpha(m-1)]\right|$$
$$< O(\Delta_m)\left(\bar{\delta}_m - \sum_{i=1}^{n} E[y_i^2]\right) + o(\Delta_m) \tag{3.57}$$

But since $\bar{x}$ is a martingale, we have

$$\sum_{i=1}^{n} E[y_i^2] = E[[\bar{x}(t_m) - \bar{x}(t_{m-1})]^2] \tag{3.58}$$

Now for any ε we can make $P[\bar{x} = x]$ as close to 1 as possible. Thus, the characteristic function can be made arbitrarily close by using a sufficiently large n (see Doob [2] p. 387). Therefore, we write

$$\left| E(\alpha(m-1)\exp\{ju_m[x(t_m) - x(t_{m-1})]\}) - \exp\left(-\frac{u_m^2\delta_m}{2}\right)E[\alpha(m-1)]\right|$$
$$< O(\Delta_m)(\delta_m - E[[\bar{x}(t_m) - \bar{x}(t_{m-1})]^2]) + o(\Delta_m) \tag{3.59}$$

But since $\sigma_i^2 \leqq \delta_m/n$, we know that

$$E[[\bar{x}(t_m) - \bar{x}(t_{m-1})]^2] < \bar{\delta}_m \tag{3.60}$$

so that we have

$$0 \leqq \bar{\delta}_m - E[[\bar{x}(t_m) - \bar{x}(t_{m-1})]^2]$$
$$= E[[x(t_m) - x(t_{m-1})]^2 - [\bar{x}(t_m) - \bar{x}(t_{m-1})]^2]$$
$$= \int_{\tau(\omega)<t_m} [[x(t_m) - x(t_{m-1})]^2 - [\bar{x}(t_m) - \bar{x}(t_{m-1})]^2]\, dP(\omega) \tag{3.61}$$

But by choosing n large enough, $P[\tau(\omega) = T]$ can be made as near to 1 as one desires, so that the above expression can be bounded by $o(\Delta_m)$. This therefore implies that

$$\left| E[\alpha(m-1)\exp\{ju_m[x(t_m) - x(t_{m-1})]\}] - E[\alpha(m-1)]\exp\left(-\frac{u_m^2\bar{\delta}_m}{2}\right)\right|$$
$$< o(\Delta_m) \tag{3.62}$$

or that

$$E\left[\exp\left(j\sum_{k=1}^{m} u_k\{x[t_k - x(t_{k-1})]\}\right)\right]$$
$$= E\left[\exp\left\{j\sum_{k=1}^{m-1} u_k[x(t_k) - x(t_{k-1})]\right\}\right]\exp\left(-\frac{u_m^2}{2}\bar{\delta}_m\right) \tag{3.63}$$

Then, by induction, we can show that for all δ_k, Δ_k, and m we have for a characteristic function

$$\exp\left(-\sum_{k=1}^{m}\frac{u_k^2\delta_k}{2}\right) \tag{3.64}$$

or that the process $x(t)$ had independent Gaussian increments—in effect a Wiener process. ∎

The three theorems in this section provide the fundamental properties of the Wiener process that will be of use in further developments. That the process is continuous is important in understanding the nature of the sample paths. Yet despite the continuity the process was found not to be of bounded variation. This fact will be important in the next section when we introduce stochastic differential equations and integrals. The introduction of the martingale concept is also important because we shall return to it again in Chapter 4 when discussing conditional expectation. The relationship between the Wiener process and a martingale leads to different interpretations of stochastic integrals (see Meyer [1, 2] or Kunita and Watanabe). These interpretations, although important, are beyond the scope of this book. For a more complete discussion of continuous martingales, see Doob [2](Chapter 7).

The properties of the Poisson process have been sufficiently developed for our needs. Further material on this process is contained in Gikhman and Skorokhod (Chapter 7). It is simple to show that the simple Poisson process is also a martingale as well as a Markov process.

3.4 STOCHASTIC DIFFERENTIAL EQUATIONS

In Chapter 2 we discussed the state variable realization of dynamical systems and of measurements made on those systems. The disturbances to such systems that acted as driving forces were always deterministic. Thus, the state and the measurement were deterministic, and if the system were observable, then we could say what the initial state was. These types of deterministic systems describe a wide area of physical problems that can then subsequently be analyzed using various techniques. Yet for an even wider class of problems the disturbances are random in nature, and thus, deterministic methods fail to provide suitable models for the problem. It is these stochastic systems that we shall model and discuss in this section.

To motivate the development of stochastic differential equations, we can consider the motion of a particle in a viscous one-dimensional medium that is subjected to random forces. If we let $x(t)$ be the velocity of the particle, then a force balance is

$$m[x(t+\Delta t)-x(t)] = -\beta x(t)\Delta t + \Delta w(t)$$

where m is the mass of the particle, β is the damping factor, and $\Delta w(t)$ re-

presents an incremental force on the particle. In most physical problems the force is random, and further, the forces applied at different times are independent. Also $\Delta w(t)$ may be a Gaussian random variable. This model is a suitable description of Brownian motion in a viscous medium. Furthermore, with the assumptions on $\Delta w(t)$, it also is a Wiener process or Brownian motion. Now, as $\Delta t \to 0$, we want to describe the behavior of $x(t)$, but we know that if $w(t)$ is a Wiener process, then $dw(t)/dt$ does not exist. Thus, we must find an interpretation for the limiting behavior of $x(t)$. Formally, we write

$$m\ dx(t) = -\beta x(t)\ dt + dw(t)$$

In general, if $\mathbf{x}(t)$ is an $(n \times 1)$-vector random process we shall represent it in the form

$$d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}, t)\ dt + \boldsymbol{\sigma}(\mathbf{x}, t)\ d\mathbf{w}(t)$$

where $\mathbf{f}(x, t)$ is an $n \times 1$ vector, $\sigma(x, t)$ an $n \times m$ matrix, and $\mathbf{w}(t)$ an $(m \times 1)$-vector Wiener process. Equations of this form will be used to represent a variety of stochastic processes. For example, they may represent actual physical systems with random disturbances or they can be used to generate signals with given statistical properties. We shall discuss these points at length in Chapter 5, but our main object in this section is to give meaning to these expressions.

In a similar fashion, a measurement process $d\mathbf{y}(t)$ is also defined by

$$d\mathbf{y}(t) = \mathbf{h}(\mathbf{x}, t)\ dt + \boldsymbol{\beta}(\mathbf{x}, t)\ d\mathbf{w}(t)$$

where $\mathbf{y}(t)$ is an $(m \times 1)$-vector process and $\mathbf{h}(\mathbf{x}, t)$ an $m \times 1$ vector. $\boldsymbol{\beta}(\mathbf{x},t)$ is an $m \times k$ matrix, and $\mathbf{w}(t)$, a $k \times 1$ Wiener process.

In this section we shall concentrate on only scalar processes, although the vector processes have similar properties. Furthermore, we shall assume that the disturbances are Wiener processes, although generalized Poisson processes will do as well (see Skorokhod, Chapter 3). Extensions to these are presented in the problems.

Now if we assume that $x(t)$ is a scalar process on $[a, b]$, then a natural definition or interpretation of the formal notation is

$$x(t) = x(a) + \int_a^t f(x, t)\ dt + \int_a^t \sigma(x, t)\ dw(t)$$

for $t \leqq b$. This interpretation has still not solved the problem, for although the first integral on the right is well defined, the second is not. This is an immediate result of the fact that $w(t)$ is not of bounded variation. Thus, in order to interpret the differential equation, we must establish an interpretation for the integral.

We now want to interpret the integral

$$z(t) = \int_0^t \phi(x(s), s)\ dw(s) \tag{4.1}$$

in some appropriate form where $w(t)$ is a Wiener process. It will be the purpose of this section to discuss several meanings of this integral. The function $z(t)$ is called a *stochastic integral*. There are two formulations of the integral that we shall study in this section, the first was postulated by Ito in 1951 and the second by Stratonovich in 1963. The Ito integral, as it is called, is important in many areas of random processes. The Ito integral will also be used in the definition of stochastic differential equations.

We find that the Ito integral has some rather strange properties compared to ordinary integrals, because second-order terms are retained in expansions, contrary to what occurs in the case of the Riemann integral. This anomalous behavior will be attributed to the fact that $w(t)$ in (4.1) is a Wiener process and, as such, has second moments that are still of order dt, not dt^2 (see Corollary 3.1). If we recall from the past section, this factor led to other problems; namely, the process was not of bounded variation and thus did not have a derivative.

A second interpretation of the stochastic integral has been afforded us by Stratonovich, and it will then be called the *Stratonovich integral*. This was developed by Stratonovich [2] to show that with this integral formulation his development of the conditional probability density function was valid (see Stratonovich [1]). A similar approach appears in Fisk. Other approaches are discussed by McShane [1, 2] and by Millar. There exists a relationship between the two integrals, which we shall obtain. The existence of two interpretations of (4.1) thus leads one to see that uniqueness of approximation to (4.1) does not exist. The use of the Stratonovich interpretation is limited and is presented for completeness only. All interpretations will be in the Ito sense.

Before proceeding with the definition of the integral, we present several facts concerning independent increment processes. We also want to generalize the integral to one of the form

$$\psi = \int_a^b \phi(t, \mathscr{F}_t)\ dw(t)$$

where $\mathscr{F}_t$ is the σ-field generated by $\{w(s)\colon s \leqq t\}$ and the integral is over the interval $[a,b]$. Thus, $\phi(t,\mathscr{F}_t)$ is an $\mathscr{F}_t$ measurable function that clearly includes functions of the form $\sigma(t, x(t))$. Also note that ψ is a random variable, since the integral is over a fixed interval. Now assume $w(t)$ is any independent increment process such that

$$E\left[(w(t) - w(s))^2\right] = \begin{cases} R(t) - R(s); & t \geqq s \\ R(s) - R(t); & t \leqq s \end{cases} \tag{4.2}$$

where R is a monotone nondecreasing function. Also, we can formally write

$$E\,[[dw(t)]^2] = dR(t) \tag{4.3}$$

We recall that if $R(t)$ were given by $\sigma^2 t$, then we would have a Wiener process. Then it is obvious that since $w(t)$ is an independent increment process

$$E[\phi(t, \mathscr{F}_t)[w(t + \Delta) - w(t)]] = E[\phi(t, \mathscr{F}_t)]E[w(t + \Delta) - w(t)] = 0. \tag{4.4}$$

Furthermore,

$$E[\phi(t, \mathscr{F}_t)[w(t + \Delta) - w(t)]^2] = E[\phi(t, \mathscr{F}_t)]E[[w(t + \Delta) - w(t)]^2] \tag{4.5}$$

Then, using (4.2), we have

$$E[\phi(t, \mathscr{F}_t)[w(t + \Delta) - w(t)]^2] = E[\phi(t, \mathscr{F}_t)][R(t + \Delta) - R(t)] \tag{4.6}$$

We shall use these facts in the development of the stochastic integral.

In order to insure that we define the stochastic integral appropriately, we first outline our approach. We first define the integral for the case where $\phi(t, \mathscr{F}_t)$ is a series of step functions. In this case a particular definition is made so that the integral will eventually possess some special properties; namely, it will be zero mean and, under an extended definition, a martingale and thus a Markov process. Having defined it for a set of step functions $\phi_n(t, \mathscr{F}_t)$, we then consider a continuous $\phi(t, \mathscr{F}_t)$ such that $\phi_n \to \phi$ in some norm. Specifically, we require ϕ_n converge to ϕ in a limit in the mean (l.i.m.) that is L^2 convergence (see Chapter 4). Thus, having defined ϕ as the lim of ϕ_n, we have a ψ_n associated with ϕ_n, and we then prove that $\psi = \text{l.i.m.}\, \psi_n$. Therefore, based on a definition of ψ for a step ϕ, we can define ψ for any ϕ that is the limit of such a set of step functions (i.e., for all integrable ϕ). The limit depends on how we defined the integral for the step functions. The first definition we use will be the Ito definition. It is extremely powerful for under a more general integral it yields a martingale. We shall then extend our step definition to a wider class, called the *generalized Stratonovich equations*. It will be possible to relate the two.

DEFINITION 4.1. Let $\phi(t, \mathscr{F}_t)$ be a step function such that

$$\phi(t, \mathscr{F}_t) = \begin{cases} 0 & t < t_j \\ \phi_j(\mathscr{F}_{t_j}) & t_j \leqq t < t_j + 1 \\ 0 & t \geqq t_j + 1 \end{cases} \tag{4.7}$$

and let $w(t)$ be a Wiener process. Then the *Ito integral* over an interval T such that $t \in T$ is defined as

$$\psi = \int_{T,I} \phi(t, \mathscr{F}_t)\ dw(t) \triangleq \sum_{j=1}^{n} \phi_j(\mathscr{F}_{t_j})[w(t_{j+1}) - w(t_j)] \tag{4.8}$$

This is well defined since $\phi(t, \mathscr{F}_t)$ is constant over any t_i, t_{i+1} and it depends only on $\mathscr{F}_{t_i}$, which depends only on $w(t)$ for $t < t_i$. Thus, for any t_i, t_{i+1}

$$\int_{t_i}^{t_{i+1}} \phi(t, \mathscr{F}_t)\ dw(t) = \phi_i(\mathscr{F}_{t_i}) \int_{t_i}^{t_{i+1}} dw(t) \tag{4.9}$$

where we define the integral of the Wiener process as

$$\int_{t_i}^{t_{i+1}} dw(t) \triangleq w(t_{i+1}) - w(t_i) \tag{4.10}$$

One immediate result of this construction is that $E[\phi]$ is zero. This is obvious, since $w(t)$ is an independent increment process, so $\phi_i(\mathcal{F}_{t_i})$ and $w(t_{i+1}) - w(t_i)$ are independent and the increment is zero mean.

Let us now define the convergence norm on the functions $\phi(t, \mathcal{F}_t)$. That is, we want to define what we mean when we say that if $\phi_n(t, \mathcal{F}_t)$ is a set of step functions, it converges to $\phi(t, \mathcal{F}_t)$ relative to that norm.

DEFINITION 4.2. If ϕ is given by (4.8) and

$$\int_{-\infty}^{\infty} E[\phi[(t, \mathcal{F}_t)]^2] \; dR(t) < \infty \tag{4.11}$$

and

$$E[(\phi)^2] < \infty \tag{4.12}$$

then the distance between two ϕ functions and between pairs of ϕ are defined as

$$\|\phi_1 - \phi_2\| \triangleq \left[\int_{-\infty}^{\infty} E[[\phi_1(t, \mathcal{F}_t) - \phi_2(t, \mathcal{F}_t)]^2] \; dR(t)\right]^{1/2} \tag{4.13}$$

and

$$\|\phi_1 - \phi_2\| = [E[(\phi_1 - \phi_2)^2]]^{1/2} \tag{4.14}$$

We can now show that ϕ equals ~~lim~~ ϕ_n as $n \to \infty$, where

$$\phi_n = \sum_{i=1}^{n} \phi_i(t_i, \mathcal{F}_{t_i})[w(t_{i+1}) - w(t_i)] \tag{4.15}$$

To do so, we must show that

$$\lim_{n \to \infty} E[[\phi - \phi_n]^2] \to 0 \tag{4.16}$$

where from (4.8) we have

$$\phi = \int \phi(t, \mathcal{F}_t) \; dw(t) \tag{4.17}$$

Substituting (4.15) and (4.17) into (4.16) we obtain

$$\begin{aligned} & E[[\phi - \phi_n]^2] \\ &= E\left[\int \phi(t, \mathcal{F}_t) \; dw(t) \int \phi(u, \mathcal{F}_u) \; dw(u)\right] \\ &- 2\left[\int \phi(t, \mathcal{F}_t) \; dw(t)\right] \sum_{i=1}^{n} \phi_i(\mathcal{F}_{t_i})[w(t_{i+1}) - w(t_i)] \\ &+ \sum_{i=1}^{n} \sum_{j=1}^{n} \phi_i(\mathcal{F}_{t_i})\phi_j(\mathcal{F}_{t_j})[w(t_{i+1}) - w(t_i)][w(t_{j+1}) - w(t_j)]] \end{aligned} \tag{4.18}$$

Assuming the conditions of Fubini's theorem hold (Rudin [2], p. 140), we can bring the expectation operator inside the integrals and assume $t > u$; then,

$$\iint E[\phi(t, \mathscr{F}_t)\phi(u, \mathscr{F}_u)\ dw(t)\ dw(u)]$$
$$= \iint E_{w(u)}[\phi(u, \mathscr{F}_u)\ dw(u) E_{w(t)/w(u)}[\phi(t, \mathscr{F}_t)\ dw(t)]] \tag{4.19}$$

But recall that the Wiener process is an independent increment process and also martingale. Thus,

$$E_{w(t)/w(u)}[\phi(t, \mathscr{F}_t)\ dw(t)] = 0 \tag{4.20}$$

unless $t = u$. Thus, (4.19) becomes

$$\iint E[\phi(t, \mathscr{F}_t)\phi(u, \mathscr{F}_u)\ dw(t)\ dw(u)] = \int E[\phi(t, \mathscr{F}_t)\phi(t, \mathscr{F}_t)]\ dR(t) \tag{4.21}$$

where for the Wiener process $dR(t)$ equals $\sigma^2\ dt$. We proceed in like fashion with the remaining terms to obtain

$$E[(\phi - \phi_n)^2] = \int E[\phi(t, \mathscr{F}_t)\phi(t, \mathscr{F}_t)]\ dR(t) - \sum_{i=1}^{n} E[\phi_i^2(\mathscr{F}_{t_i})][R(t_{i+1}) - R(t_i)] \tag{4.22}$$

Now if the step functions are such that $\|\phi - \phi_n\|^2 \to 0$ as $n \to \infty$ where (4.13) is used and if $R(t)$ is continuous, then

$$\lim_{n\to\infty} E[(\phi - \phi_n)^2] = 0 \tag{4.23}$$

and our approximation is the limit in the mean of ϕ. The following theorem then summarizes this result.

THEOREM 4.1

If $w(t)$ is a Wiener process and if

$$\phi_n = \sum_{i=1}^{n} \phi(t_i, \mathscr{F}_{t_i})[x(t_{i+1}) - x(t_i)] \tag{4.24}$$

and if ϕ equals lim ϕ_n, where ϕ_n is $\phi(t_n, \mathscr{F}_{t_n})$ for any n; if $R(t)$ is continuous in t, then the integral ϕ is given by

$$\phi = \text{l.i.m.}\ \phi_n \tag{4.25}$$

where

$$\phi = \int_I \phi(t, \mathscr{F}_t)\ dw(t) \tag{4.26}$$

and the I stands for the definitions in the Ito sense, that is, (4.24). To the

reader familiar with measure theory the condition on ϕ is that it be integrable (see Halmos [2]).

Let us now consider a stochastic process $z(t)$ defined by

$$z(t) = \int_a^t \phi(s, \mathcal{F}_s)\ dw(s) \tag{4.27}$$

where as before $w(t)$ is a Wiener process. Note that this differs from ϕ in that the limits of integration vary with time. Yet we define the integral in the same way, that is, using the Ito interpretation. These processes are used to define the solution of differential equations driven by Wiener processes. We now show that these processes are martingales. It should be noted that if $\phi(s, \mathcal{F}_s)$ were $\phi(t, s, \mathcal{F}_s)$, then $z(t)$ would not be a martingale (see Frost).

THEOREM 4.2

Every process $z(t)$ defined by (4.27) is a martingale.

Proof. We know that

$$z_n(t) = \int_a^t \phi_n(s, \mathcal{F}_s)\ dw(s) \tag{4.28}$$

is the step approximation and that

$$z(t) = \text{l.i.m. } z_n(t) \tag{4.29}$$

Now, let

$$z_n(t_1) = \int_a^{t_1} \phi_n(s, \mathcal{F}_s)\ dw(s) \tag{4.30}$$

If $t < t_1$, then

$$z_n(t_1) = \int_t^{t_1} \phi_n(s, \mathcal{F}_s)\ dw(s) + \int_a^t \phi_n(s, \mathcal{F}_s)\ dw(s) \tag{4.31}$$

Now we know that

$$E\left[\int_t^{t_1} \phi_n(s, \mathcal{F}_s)\ dw(s) \middle| \mathcal{F}_t\right] = 0 \tag{4.32}$$

by (4.28) we have

$$z_n(t) = \int_a^t \phi_n(s, \mathcal{F}_s)\ dw(s) \tag{4.33}$$

so that

$$E\left[z_n(t) \mid \mathcal{F}_t\right] = z_n(t) \tag{4.34}$$

Therefore,

$$E\left[z_n(t_1) \mid \mathcal{F}_t\right] = z_n(t) \tag{4.35}$$

and since $z_n(t_1) \to z(t_1)$ as $n \to \infty$, we have

$$E\left[z(t_1) \mid \mathcal{F}_t\right] = z(t) \tag{4.36}$$

which shows the martingale property. ∎

There are several other properties of this integral that are worth mentioning without proof and the reader may find them explained in Doob [2] [9, Theorems 5.2 and 5.3]. The first of these theorems states that the integral defined in (4.1) yields a function $z(t)$ that is continuous with probability 1. In order to prove this fact, we need the Borel-Cantelli lemma, which is introduced in Appendix B. The proof of this fact is outlined in Problem 3.20. The second important property of representations of the form of (4.1) is that if the process $z(t)$ has bound second moment and is a.e. continuous on the desired interval and a suitable $\phi\ (t, x(t))$ exists, then a representation suitable for stochastic differential equations as we shall develop in (4.46) will exist. This fact is outlined in Problem 3.21.

We now want to consider the definition of the generalized Stratonovich integral (Stratonovich [2]). Let us assume that $\phi(t, \mathcal{F}_t)$ takes the form $\phi(t, w(t))$. Then the integral to be considered is

$$\psi = \int_a^b \phi(t, w(t))\ dw(t) \tag{4.37}$$

where $w(t)$ again is a Wiener process. The Ito interpretation of ψ was

$$\psi = \text{l.i.m.} \sum_{i=1}^{n} \phi(t_i; w(t_i))[w(t_{i+1}) - w(t_i)] \tag{4.38}$$

Here we really started with a step approximation of $\phi(t, w(t))$ and obtained it by expanding about the point $w(t) = w(t_i)$. We could now rightly ask what would happen if we were to expand $\phi(t, w(t))$ about some point between $w(t_i)$ and $w(t_{i+1})$. That is if we were to expand it out some arbitrary linear combination of these points.

DEFINITION 4.3. Let $\phi(t, w(t))$ be a set of step function on $t \in T$. Let $\phi(t, w(t))$ be given by

$$\phi(t, w(t)) = \begin{cases} 0 & (t < t_i) \\ \phi(t_i, \lambda w(t_i) + (1 - \lambda) w(t_{i+1})) & (t_i \leqq t < t_{i+1}) \\ 0 & (t_{i+1} \leqq t) \end{cases} \tag{4.39}$$

where $0 \leqq \lambda \leqq 1$. Then

$$\psi \equiv \sum_{i=1}^{n} \phi((t_i, \lambda w(t_i) + (1 - \lambda) w(t_{i+1}))[w(t_{i+1}) - w(t_i)] \tag{4.40}$$

is the *generalized Stratonovich integral* of (4.37).

Historically, Stratonovich used $\lambda = 1/2$ in his paper; the extension to an arbitrary λ was discussed by Frost. The first fact to note is that ϕ is no longer necessarily zero mean. We would now like to extend this to arbitrary $\phi(t, w(t))$ which are limits of step functions of the type $\phi_n\ (t, w(t))$. This is a simple

extension of Theorem 4.2. We shall therefore say that Theorem 4.2 also holds for the Stratonovich case.

We now will relate the Stratonovich and Ito integrals.

THEOREM 4.3

Let $\phi(w(t), t)$ have a continuous partial derivative

$$\frac{\partial \phi(w(t), t)}{\partial w(t)}$$

for $w \in (-\infty, \infty)$, $t \in [a, b]$. Let

$$\int_S \phi(t, w(t))\, dw(t)$$

be the generalized Stratonovich integral with $w(t)$ being a Wiener process. Let

$$\int_I \phi(t, w(t))\, dw(t)$$

be the Ito integral of that function. Then

$$\begin{aligned} &\int_S \phi(t, w(t))\, dw(t) \\ &= \int_I \phi(t, w(t))\, dw(t) + (1 - \lambda) \int \frac{\partial \phi(t, w(t))}{\partial w(t)}\, dt \end{aligned} \tag{4.41}$$

Note that the second integral on the right is well defined in the Riemann sense.

Proof. Recall that

$$\begin{aligned} &\int_S \phi(t, w(t))\, dw(t) \\ &= \text{l.i.m.} \sum_{i=1}^{n} \phi(t_i, \lambda w(t_i) + (1 - \lambda) w(t_{i+1}))[w(t_{i+1}) - w(t_i)] \\ &= \text{l.i.m.} \sum_{i=1}^{n} \phi(t_i, y_i)[w(t_{i+1}) - w(t_i)]. \end{aligned} \tag{4.42}$$

Now expand $\phi(t_i, y_i)$ in a Taylor series about $y_i = w(t_i)$. This yields

$$\begin{aligned} \phi(t_i, y_i) = \phi(t_i, w(t_i)) &+ \frac{\partial \phi(t_i, y_i)}{\partial y_i}\bigg|_{y_i = w(t_i)} \\ &\cdot [(\lambda - 1)w(t_i) + (1 - \lambda)w(t_{i+1})] + \cdots \end{aligned} \tag{4.43}$$

Substituting this into the Stratonovich limit yields

$$\begin{aligned} &\lim \sum_{i=1}^{n} \phi(t_i, y_i)[w(t_{i+1}) - w(t_i)] \\ &= \lim \sum_{i=1}^{n} \phi(t_i, w(t_i))[w(t_{i+1}) - w(t_i)] \\ &\quad + \lim (1 - \lambda) \sum_{i=1}^{n} \frac{\partial \phi(t_i, w(t_i))}{\partial w(t_i)} [w(t_{i+1}) - w(t_i)]^2 + o(t) \end{aligned} \tag{4.44}$$

But the first limit is the Ito definition, and the second is easily shown to be the Riemann limit. This follows directly from Theorem 4.1 and first part of Theorem 3.2. The extension to the case where $w(t)$ is a vector process is contained in Stratonovich [2] but will not concern us here. ∎

This previous theorem relates the Stratonovich integral to the Ito integral plus a Riemann integral. In order to evaluate the Stratonovich integral, we would still have to either evaluate its lim behavior or evaluate the Ito integral. Wong and Zakai have shown that it is possible to obtain the Stratonovich evaluation for $\lambda = 1/2$ from a Riemann rule applied to the stochastic integral. They also discuss the convergence of integrals that are defined on bounded processes to integrals on Wiener processes. It is beyond the scope of this book to discuss these extensions, since they really are in the realm of approximations. We may also extend the definition of these integrals to cases where $x(t)$ are independent increment processes but not necessarily Wiener processes. This is discussed in the problems.

We can now define a stochastic differential equation in terms of the integral interpretations just discussed (Wong [2], pp. 149–150).

DEFINITION 4.4. A *stochastic differential equation* is of the form

$$dx(t) = f(x, t)\ dt + \sigma(x, t)\, dw(t); \qquad t \in [a, b] \tag{4.45}$$

where $w(t)$ is an independent increment process; for each t the integral

$$\int_a^t \sigma(x, s)\ dw(s)$$

is interpreted in the Ito sense and for all $t \in [a, b]$, $x(t)$ is equal to the random variable defined by

$$x(t) = x(a) + \int_a^t f(x, s)\ ds + \int_a^t \sigma(x, s)\ dw(s) \tag{4.46}$$

An interesting extension of stochastic differential equations is *Ito's differentiation rule*. This rule allows one to obtain the filtering equations directly from suitable representation theorems. We shall discuss this in Chapter 5 when we develop Bucy's representation theorem. This theorem is presented for the vector case and for systems driven by Wiener processes.

THEOREM 4.4

(Ito's Lemma) Let $G(x_1(t), \cdots, x_n(t), t) = g(t)$ be a function of the random variables $x_i(t)$, where the $x_i(t)$ are solutions to the Ito equation.

$$dx_i(t) = f_i(x, t)\ dt + \sigma_i(x, t)\ d\mathbf{w}(t) \tag{4.47}$$

where $\mathbf{w}(t)$ is a $(k \times 1)$ - vector Wiener process with covariance matrix $\mathbf{I}$ and $\sigma_i(x, t)$ is a $1 \times k$ vector. Then,

$$dg(t) = \sum_{i=1}^{n} \frac{\partial G}{\partial x_i}\, dx_i(t) + \tfrac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 G}{\partial x_i\, \partial x_j}\, \sigma_i(x, t)\, \sigma_j^T(x, t)\, dt + \frac{\partial G}{\partial t}\, dt \tag{4.48}$$

Proof. This will only be a heuristic proof since a complete proof is quite detailed (see Ito [2], pp. 187–193) and beyond the scope of the book. Recall that

$$g(t) = G(x_1, \cdots, x_n, t) \tag{4.49}$$

Likewise,

$$g(t + dt) = G(x_1 + dx_1, \cdots, x_n + dx_n, t + dt) \tag{4.50}$$

Then, by definition,

$$dg(t) = g(t + dt) - g(t) \tag{4.51}$$

Now expand (4.50) about $x_1, \cdots, x_n, t$ and use this expansion in (4.51) to obtain

$$dg(t) = \sum_{i=1}^{n} \frac{\partial G}{\partial x_i}\, dx_i + \frac{\partial G}{\partial t}\, dt + \tfrac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 G}{\partial x_i\, \partial x_j}\, dx_i\, dx_j + \cdots \tag{4.52}$$

But recall that

$$\begin{aligned} dx_i\, dx_j &= (f_i\, dt + \sigma_i\, d\mathbf{w})(f_j\, dt + \sigma_j\, d\mathbf{w}) \\ &= f_i f_j (dt)^2 + f_i \sigma_j\, d\mathbf{w}\, dt + f_j \sigma_i\, d\mathbf{w}\, dt + \sigma_i\, d\mathbf{w}^T\, d\mathbf{w}^T \boldsymbol{\sigma}_j^T \end{aligned} \tag{4.53}$$

Now, retaining only terms of order dt and dropping those of order $(dt)^2$ or greater, (4.53) becomes

$$dx_i\, dx_j = \sigma_i \sigma_j^T dt \tag{4.54}$$

since we assumed $\mathbf{w}(t)$ is a reduced Wiener process. Thus, using (4.54) in (4.52) and dropping all higher-order terms, we have

$$dg(t) = \sum_{i=1}^{n} \frac{\partial G}{\partial x_i}\, dx_i + \frac{\partial G}{\partial t}\, dt + \tfrac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 G}{\partial x_i\, \partial x_j}\, \sigma_i \sigma_j^T dt \tag{4.55}$$

which "proves" the theorem. ∎

In Problems 3.18 and 3.19 we develop examples where Ito differentiation rule is used. Likewise, in Problem 3.16 we develop an Ito rule for Poisson forcing functions. Of greatest importance to our discussion is how the Ito rule is used to determine the propagation of the conditional densities in nonlinear estimation. It provides a rigorous tool for such analysis. We shall discuss this in Chapter 5.

What we have done in this section was to describe physical systems of interest wherein random forcing functions were present and to ascribe to

them interpretations suitable for our needs. The definition of the stochastic differential equation required the development of a stochastic integral, two of which we discussed in detail. The one which we find most suitable is the Ito interpretation because of its zero mean value and its martingale property.

3.5 CONCLUSIONS

The model developed in this chapter will be used in Chapter 5 as a basis for all the estimation schemes. This model will have a state equation given by

$$d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}, t)\, dt + \sigma(\mathbf{x}, t)\, d\mathbf{u}(t)$$

where $\mathbf{u}(t)$ will be an independent increment process. The interpretation of this stochastic differential equation will be in the Ito sense. In like manner, a measurement will be given by

$$d\mathbf{y}(t) = \mathbf{h}(\mathbf{x}, t)\, dt + \beta(\mathbf{x}, t)\, d\mathbf{w}(t)$$

Again the Ito interpretation. To get a thorough understanding of these equations and their interpretations, we reviewed the basics of stochastic processes, independent increment processes, and the properties of the Wiener process.

The review of stochastic processes was brief but covered the material of importance, that is, the structure of the underlying sample space and the probability measure induced on that space. The concept of conditional probability and expectation was introduced and elaborated upon. In the next chapter we shall further discuss conditional expectation with reference to the problem of estimation. This extension will lead us to consider ways to obtain conditional probability-density functions. As a final point in our brief discussion of stochastic processes was the introduction of the Markov process and the transition probability density. We saw that if a process was a Markov process, then knowing the density of any time and the transition density the joint density of any combinations was possible. We also saw that, after the fact, the processes that we are considering are all Markov.

In the second section we introduced the concept of the independent increment process. Two specific types were the Wiener process and the Poisson process. The formal derivative of both processes had impulsive correlation functions or flat power spectra. As such they are called white noise and are quite useful for exciting dynamical systems to obtain desired responses or for modeling measurement noise. In this section we also discussed some of the properties of independent increment processes as related to their sample paths. As the need demands we shall develop more properties of these independent increment processes.

The third section discussed properties of the Wiener process. The choice of the Wiener process as compared to the Poisson process was arbitrary, although historically it is more important. The fact that the Wiener process is

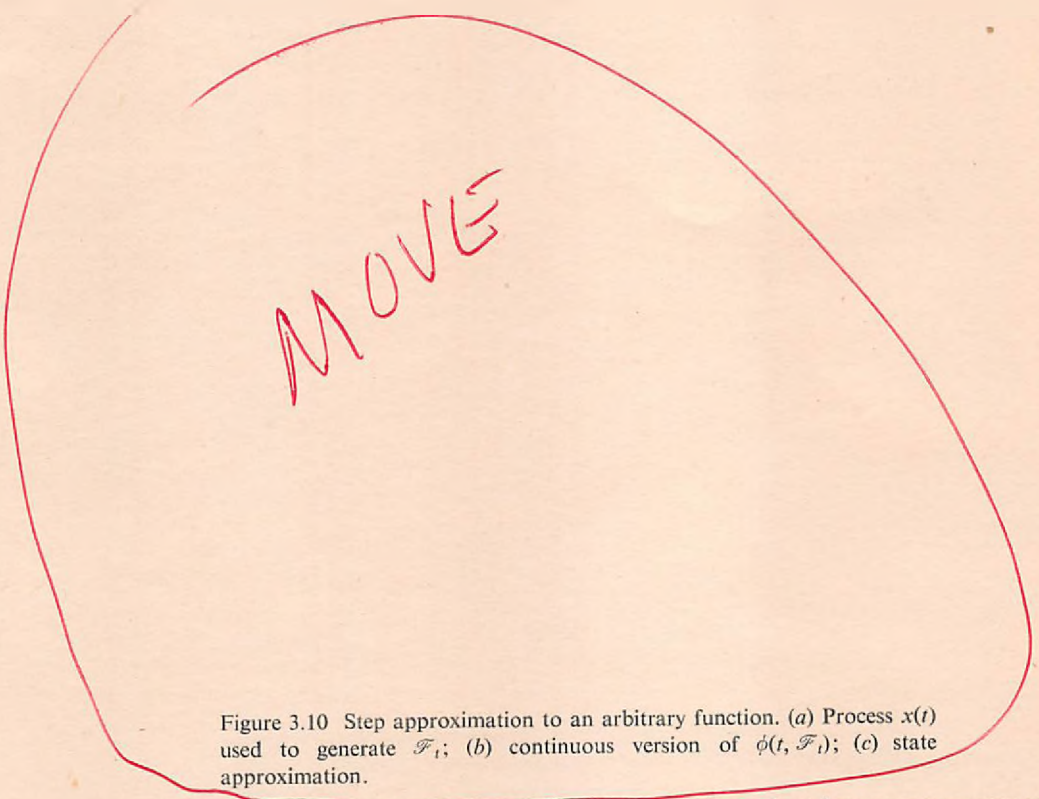

Figure 3.10 Step approximation to an arbitrary function. (*a*) Process $x(t)$ used to generate $\mathscr{F}_t$; (*b*) continuous version of $\phi(t, \mathscr{F}_t)$; (*c*) state approximation.

continuous leads us to conclude that any independent increment process with a suitable covariance is a Wiener process if and only if it is continuous. The two important facts concerning the integral nature of a Wiener process or Wiener measure is the fact that it is not of bounded variation and that $dw\,dw$ equals dt. The former fact influences our interpretation of stochastic differential equations, while the latter determines the nature of Taylor expansions of functionals of Wiener processes. This latter fact we shall use in Chapter 5. We finally introduced the martingale and showed that the Wiener process was a martingale. Conversely, we showed that a continuous martingale with a suitable covariance is a Wiener process. This thus states that any process with

a suitable covariance is a Wiener process if and only if it is a continuous martingale. Such a decomposition is important in other approaches to filtering where decompositions of this nature are used to determine optimal estimates. The decomposition is discussed in Meyer [1, 2], and it is used in J. R. Clark for the development of the estimation equations for doubly stochastic point processes.

In the fourth and last section of the chapter we developed the concept of stochastic differential equations. To do this, we saw that it was first necessary to define a stochastic integral. The interpretation we shall use is the Ito interpretation, because the resulting process is a zero mean martingale. The relationship of stochastic differential equations to limiting discrete-time versions is discussed in Wong and Zakai and in Wong [2, Chapter 4]. Such limit procedures are important when developing discrete-time simulations. The final issue of this section was the development of the Ito rule for differentiation. In a formal proof we showed how the fact that for a Wiener process $dw\,dw = dt$ played an important role in determining differentials.

The issues that we have discussed in this chapter reflect on only those that will be important for future use in the text. There are many other peripheral issues whose consideration would complement the material; these can be found in Doob [2]; Wong [2]; Dynkin [1,2]; Gikhman and Skorokhod; or Skorokhod. The problems at the end of this chapter lay the groundwork for this complimentation. There is one other issue that is important and is considered in Appendix A. This is the proof of the existence and uniqueness of solutions to stochastic differential equations. It parallels the analysis for the deterministic case of Chapter 2 but shows how probability 1 proofs are developed.

The extension of our results to vector processes and to Poisson forcing functions are developed in the problems and discussed in the references. In Chapter 5 we shall exploit both types of processes for modeling the estimation problem.

3.6 PROBLEMS

3.1. [Mehr and McFadden] A Gauss-Markov process can be characterized by a covariance matrix $R_x(t_1, t_2)$ given by

$$E[x(t_1)x(t_2)] = R_x(t_1, t_2)$$

A covariance matrix is called triangular if

$$R_x(t_1, t_2) = f(t_1)g(t_2) \qquad (t_1 \leqq t_2)$$

(a) Show that a Gaussian process $x(t)$ is Markov if and only if it has a triangular covariance.

(b) Let $T = [a,b]$ and let P be a continuous function defined on $T \times T$. Suppose

$$P(t_1, t_2) = m(t_1)n(t_2) \qquad (a \leqq t_1 \leqq t_2 \leqq b)$$

Show that P is the covariance of a zero mean Gauss-Markov process that is continuous in the mean and nondegenerate on (a,b) if and only if

(i) $m(t)/n(t)$ is positive and strictly increasing on (a,b);

(ii) $m(t)/n(t)$ can be defined to be continuous on (a,b), nonnegative and strictly increasing on (a,b).

3.2. Show that the class of all Gauss-Markov processes with stationary transition probabilities is characterized by covariance of one of two forms:

(a) $$R_1(t_1, t_2) = K_1 t_1 + K_2$$

where $0 \leqq t_1 \leqq t_2 < \infty$, $K_1 > 0$, $K_2 \geqq 0$;

(b) $$R_2(t_1, t_2) = K_3 e^{+K(t_1+t_2)} \mp K_4 e^{+K(t_2-t_1)}$$

where $0 \leqq t_1 \leqq t_2 < 0$, $k > 0$, $K_4 > 0$, $K_3 \mp K_3 \geqq 0$.

3.3. Let $N(t)$ be a Poisson jump process with rate parameter $\lambda(t)$, where $\lambda(t)\,dt$ is the number of arrivals in $t, t + dt$ seconds.

(a) Show that if $N(0) = 0$, then

$$P[N(t) = k] = \frac{\left[\int_0^t \lambda(\xi)\,d\xi\right]^k}{k!} \exp\left(-\int_0^t \lambda(\xi)\,d\xi\right)$$

(b) Obtain the characteristic function for this process.

(c) Find the correlation function $E[N(t)N(s)]$.

(d) Let

$$y(t) = N(t) - E[N(t)]$$

Find $E[N(t)]$ and obtain $E[y(t)y(s)]$.

(e) Let $z(t) = dy(t)/dt$; show that $z(t)$ is a time-varying white noise process.

3.4. Repeat Problem 3.3 for the case where $\mathbf{N}(t)$ is an $(n \times 1)$-vector process and where

$$\mathbf{N}(t) = \begin{bmatrix} N_1(t) \\ \vdots \\ N_n(t) \end{bmatrix}$$

and each $N_i(t)$ is independent of $N_j(t)$ $(i \neq j)$.

3.5. Prove Lemma 3.1.

3.6. Prove Lemma 3.2.

3.7. Let $N(t)$ be a Poisson process on $[0, T]$ with rate $\lambda(t)$. Let $\{t_i\}_{i=1,\,M}$ be the first M arrival times; where $t_i < t_{i+1} < t_{i+2}$;

(a) Let $p(t_1,\cdots,t_M)$ be the joint probability density of these arrival times. Show that

$$p(t_1,\cdots,t_M) = e^{-Q}\prod_{i=1}^{M}\lambda(t_i)$$

where $Q = \int_0^T \lambda(t)\ dt$.

(b) Show that the conditional probability of $p(t_1,\cdots,t_M \mid M = m)$ is

$$p(t_1,\cdots,t_M \mid M = m) = \left[\int_0^T \lambda(t)\ dt\right]^{-m} m! \prod_{i=1}^{m}\lambda(t_i)$$

and the joint probability is

$$p(t_1,\cdots,t_M;\ M) = \exp\left[-\int_0^T \lambda(t)\ dt\right]\prod_{i=1}^{M}\lambda(t_i)$$

(c) Show that for $\lambda(t) = \lambda$ the times $\{t_i\}$, given $M = m$, are uniformly distributed on $(0, T)$ (see Bar-David).

3.8. Show that the Wiener process is a Markov process. Show that it is a martingale.

3.9. Show that the Poisson process is a Markov process. Show that it is a martingale.

3.10. Let $x(t)$ be a Poisson process with rate parameter λ.

(a) Find the correlation function

$$E\left[\{x(t) - E[x(t)]\}\{x(s) - E[x(s)]\}\right] = K_x(t,s)$$

(b) Let $y(t) = dx(t)/dt$. Find $K_y(t,s)$. How does it relate to the white noise process?

3.11. Let $x(t)$ be a scalar Markov process given by

$$dx = -ax\ dt + dw \qquad (a > 0)$$

or

$$x = \int_0^t e^{-a(t-\tau)}\ dw$$

(a) Evaluate the mean of $x(t)$ when $w(t)$ is a normalized Wiener process.

(b) Evaluate the covariance of the process $x(t)$.

3.12. Let $N(t)$ be a simple Poisson process and let

$$I(t) = \int_0^t g(\sigma)\ dN(\sigma) = \text{l.i.m.}\sum_{i=0}^{n} g(\sigma_i)[N(\sigma_{i+1}) - N(\sigma_i)]$$

(a) Show that $I(t)$ is a martingale.

(b) Show that

$$\int_0^t N(\sigma)\ dN(\sigma) = \tfrac{1}{2}N^2(t) - \tfrac{1}{2}N(t)$$

3.13. Let $x(t)$ be a martingale with $\mathscr{B}_t$ a σ-field for which it is measurable. Let

$$E[x(t) - x(t_1)|\mathscr{B}_{t_1}] = 0$$
$$E[(x(t) - x(t_1))^2|\mathscr{B}_{t_1}] = \lambda(t) - \lambda(t_1)$$

Define ϕ

$$\phi = \int_0^T f(t)\ dx(t) = \sum_{k=0}^{n-1} f(t_k)[x(t_{k+1}) - x(t_k)]$$

for all step functions $f(t)$. Prove the following:

(a) If $f_n(t) \to f(t)$, where $f_n(t)$ are step functions, and the convergence is mean square, show that if

$$\phi_n = \int_0^T f_n(t)\ dx(t)$$
$$\phi = \int_0^T f(t)\ dx(t)$$

then $\phi_n \to \phi$ in mean square.

(b) If $\phi(t)$ is

$$\phi(t) = \int_0^t f(s)\ dx(s)$$

then $\phi(t)$ is a martingale.

3.14.* Prove Lemma B.4.

3.15.* Prove Theorem B.1 for the Markov process:

$$dx(t) = f(x(t), t)\ dt + \sigma(x(t))\ dN(t)$$

where $dN(t)$ is a scalar simple Poisson process and where $x(t)$ is a scalar process and both f and σ satisfy appropriate Lipschitz conditions. Combine this result with that of Theorem B. 1 to prove the existence and uniqueness of solutions to

$$dx(t) = f(x(t), t)\ dt + \sigma_1(x(t))\ dw(t) + \sigma_2(x(t))\ dN(t)$$

where $w(t)$ is a normalized Wiener process (see Skorokhod, Chapter 3).

3.16. Consider the $(n \times 1)$-vector Markov process given by

$$d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}, t)\ dt + \mathbf{B}(t)\ d\mathbf{N}(t)$$

where $\mathbf{N}(t)$ is a $(q \times 1)$-vector simple Poisson process and $\mathbf{B}(t)$ is an $n \times q$ matrix. Let $g(x, t)$ be a scalar function of the process $x(t)$. Show that the Ito differential of $g(x, t)$ is given by

$$dg(x, t) = \frac{\partial g(x, t)}{\partial t}\ dt + \sum_{i=1}^{n} \frac{\partial g(x, t)}{\partial x_i} f_i(x, t)\ dt + \sum_{i=1}^{\infty} [g(x(t) + \mathbf{B}(t)\gamma_i, t) - g(x(t), t)]\ d\mathbf{N}^T(t)\,\gamma_i$$

*Note. These problems depend on results developed in Appendix B.

where γ_i is a $q \times 1$ vector with all zeros except a 1 in the ith entry. (*Note.* Compare this to the Feller-Kolmogorov equation in Chapter 5.)

3.17. Let $N(t)$, $t \geq T$ be a scalar Poisson process with rate $\lambda(t)$ and arrival times t_i. Let $f(t)$ be arbitrary integrable function on T. Show that

$$E\left[\prod_{i=1}^{N(t)} f(t_i)\right] = \exp\left[\int_{t_0}^{t} \lambda(\tau)[f(\tau) - 1]\, d\tau\right]$$

(See J. R. Clark, p. 231.)

Hint. Let

$$\phi(t) = \prod_{i=1}^{N(t)} f(t_i) = \exp\left(\int_{t_0}^{t} \ln f(\tau)\, dN(\tau)\right)$$

and use Ito's lemma.

3.18. Kailath [3]: Let $g(L,t) = \ln L(t)$, where $L(t)$ is the stochastic process give by

$$dL(t) = a(t)\, dt + b(t)\, du(t)$$

(a) Using Ito's rule find $dL(t)$ and show that

$$d \ln L(t) = \frac{dL(t)}{L(t)} - \tfrac{1}{2} \frac{1}{L^2(t)} b^2(t)\, dt$$

(b) Show that the Ito integral equals

$$\int_{0,I}^{T} \frac{dL(t)}{L(t)} = \ln L(T) + \tfrac{1}{2} \int_0^T \frac{b^2(t)}{L^2(t)}\, dt - \ln L(0)$$

3.19. Kailath [3]: Let $g(x, t) = x^2$ and let $x(t)$ be given by

$$dx(t) = a(t)\, dt + b(t)\, dw(t)$$

where $E[dw(t)\, dw^T(t)] = 1$.

(a) Using Ito's rule find $dg(x,t)$.

(b) Let $a(\ ) = 0$, $b(\ ) = 1$, and show that

$$dw^2 = 2w\, dw + \tfrac{1}{2}\, dt$$

(c) Using the results of part (b) show that the Ito integral

$$\int_{0,I}^{T} w\, dw = \tfrac{1}{2} w^2(T) - \tfrac{1}{2} T$$

3.20. Let $w(t)$ be a Wiener process and let $z(t)$ be given by 4.1.

(a) Let $\phi_n(s,x(s))$ be a step function and define

$$z_n(t) = \int_a^t \phi_n(s, x(s))\, dw(s)$$

Show that

$$P\left[\sup|z(t) - z_n(t)| \leqq \frac{1}{n}\right] \leqq E[|z(b) - z_n(b)|^2]n^2 \leqq \frac{1}{n^2} \qquad (t \in [a, b])$$

(b) Use the Borel-Cantelli lemma to show that $z(t)$ is a.e. continuous.

3.21. Let $x(t)$ be a martingale with a bounded second moment for all $t \in [a,b]$ and assume that $x(t)$ is a.e. continuous on $[a,b]$. Show that there exists a Wiener process $w(t)$ such that for all $t \in [a,b]$

$$x(t) = x(a) + \int_a^t \phi(s, x(s))\ dw(s)$$

Be sure you explicitly state the conditions on ϕ $(s,x(s))$.

CHAPTER 4

OPTIMIZATION CRITERION

We have thus far established a model of a system and a measurement and have studied several of their properties. We now wish to combine our knowledge and obtain a suitable cost criterion defined on the set of measurements. A cost criterion is basically a measure of performance in a particular task. For our purpose the task will be the estimation of the state of the dynamical system. To this end, the only information we have will be noise corrupted measurements $\mathbf{z}(t)$. Thus, given $\mathbf{z}(t)$, we are asked to make the best possible estimate of $\mathbf{x}(t)$. In order to do so, we must define a cost criterion, which will be the objective of this chapter.

In the first section we introduce the concept of linear spaces and work toward the definition of a Hilbert space. The Hilbert space is a complete inner product space of functions from the underlying probability space Ω into $\mathbf{R}^n$. We show that the space of all random functions with finite second moments form a Hilbert space. Hilbert spaces have the property of orthogonal projections, which make them ideal for the use in estimation theory. That is, if we generate a subspace of a Hilbert space by some functionals of the measurements, then we can obtain an estimate of the random parameter $\mathbf{x}(\omega)$ such that the mean square error between $\mathbf{x}(\omega)$ and $\hat{\mathbf{x}}(\omega)$, the projection of $\mathbf{x}(\omega)$ onto the subspace, is minimum.

Having developed the properties of the Hilbert space in the first section, the second section considers the minimum mean square error problem. It shows that for a finite set of measurements the minimum mean square error (MMSE) estimate is the conditional mean. We then seek to obtain the MMSE estimate of a random variable given an observed process $\mathbf{y}(s,\omega)$ over an interval $s \in [t_0,t]$. To do this, we find our first definition of conditional expectation to be inadequate and are compelled to introduce a measure theoretic definition. Since our objective is not to study measure theory, we do not prove many of the theorems but reference them to the appropriate literature. Our objective is to provide a structure for continuous-time estimation. We conclude this section with a discussion of subspaces generated by the meas-

urements and relate the conditional mean to orthogonal projections in Hilbert spaces.

The third section is a presentation of discrete-time filtering. It is an example of how the concept of orthogonal projections can be used as a basis for optimum estimation. These results were first derived by Kalman and are called the discrete-time Kalman filter.

In the fourth and final section we discuss the extension that the Hilbert-space approach provides for both discrete- and continuous-time estimation. In this context we discuss reproducing kernel Hilbert spaces (RKHS), linear estimation using Wold decompositions (innovations), and other optimization criteria.

The purpose of this chapter is to show that the MMSE criterion is most suitable because of its ideal properties based upon the Hilbert-space nature of the L^2 space. Other criteria are not as general and such properties as existence and uniqueness do not follow as readily. The results of this chapter provide a basis for the optimal estimation construction to be developed in Chapter 5.

4.1. LINEAR SPACES

In the study of optimal estimation it becomes necessary to consider functions and functionals of the random variables to be estimated and of the measurement random variables used in the actual estimation. To perform this study in a consistent fashion, it is necessary to introduce the concept of linear space and certain structures on linear spaces that allow us to make precise statements concerning the optimal estimates. Linear spaces are also called *vector spaces*.

The linear space that we will be interested in is that space generated by functions from Ω, the underlying probability space, into $\mathbf{R}^n$.

DEFINITION 1.1. A nonempty set L of elements $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$, $\cdots$ is called a linear space (vector space) if it satisfies the following conditions:

1. For all elements $\mathbf{x},\mathbf{y} \in L$ there exists a unique element z called the sum *given by*

$$\mathbf{z} = \mathbf{x} + \mathbf{y} \qquad (\mathbf{z} \in L)$$

such that

a. $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ (commutative)

b. $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$ (associative)

c. There exists an element $\mathbf{0} \in L$ such that
$\mathbf{x} + \mathbf{0} = \mathbf{x}$ for all $\mathbf{x} \in L$

d. For all $\mathbf{x} \in L$ there exists an element
$-\mathbf{x} \in L$ such $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$

2. For any finite real number a and any $\mathbf{x} \in L$ there exists an element $a\mathbf{x} \in L$ such that

a. $a(b\mathbf{x}) = ab\mathbf{x}$; $\quad a,b \in \mathbf{R}$

b. $1\mathbf{x} = \mathbf{x}$; $\quad 1 \in \mathbf{R}$

3. Addition and multiplication are distributive;

a. $(a + b)\mathbf{x} = a\mathbf{x} + b\mathbf{x}$; $\quad a,b \in \mathbf{R}$; $\mathbf{x} \in L$

b. $a(\mathbf{x} + \mathbf{y}) = a\mathbf{x} + a\mathbf{y}$; $\quad a \in \mathbf{R}$; $\mathbf{x}.\mathbf{y} \in L$

The elements of L are called *vectors*. We shall also assume that the vectors belonging to L are measurable with respect to the σ-field $\mathcal{A}$ of Ω, so that they are also random variables. In this case the elements are called *random vectors*. Thus, we will consider L to be all measurable mappings from Ω into $\mathbf{R}^n$ that satisfy the conditions of a linear space.

We now want to consider the space L and induce a structure on it that will allow us to compare two elements within this space. A quantity called the *metric* plays an important role in proving convergence of sequences to given functions in L.

DEFINITION 1.2. A metric ρ is a mapping from L into the positive real line and has the following properties: Let $\mathbf{x}, \mathbf{y}, \mathbf{z} \in L$, then

1. $\rho(\mathbf{x},\mathbf{y}) = 0 \quad$ iff $\quad \mathbf{x} = \mathbf{y}$
2. $\rho(\mathbf{x},\mathbf{y}) = \rho(\mathbf{y},\mathbf{x})$
3. $\rho(\mathbf{x},\mathbf{z}) \leqq \rho(\mathbf{x},\mathbf{y}) + \rho(\mathbf{y},\mathbf{z})$

The double (L,ρ) is called a *metric space* M. Another measure is called the *norm*.

DEFINITION 1.3. A *norm* is a function defined over all elements of a linear space L such that it maps each $\mathbf{x} \in L$ into a real number $\|\mathbf{x}\|$ (the norm of x) with the following properties ($\mathbf{x}, \mathbf{y} \in L$, $a \in \mathbf{R}$):

1. $\|a\mathbf{x}\| = |a| \cdot \|\mathbf{x}\|$
2. $\|\mathbf{x}\| > 0; \mathbf{x} \neq 0$
3. $\|\mathbf{x} + \mathbf{y}\| \leqq \|\mathbf{x}\| + \|\mathbf{y}\|$

The space L with the norm is called a *normed linear space*. Every normed linear space is a metric space if we define the metric $\rho(\mathbf{x},\mathbf{y})$ in terms of the norm as follows:

$$\rho(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\| \tag{1.1}$$

We now want to consider the convergence properties of the spaces. Let the sequence $\{\mathbf{x}_n\}$ belong to L. The sequence is called Cauchy if, given any $\varepsilon > 0$, there exists an integer $N(\varepsilon)$ such that

$$\rho(\mathbf{x}_n,\mathbf{x}_m) < \varepsilon \tag{1.2}$$

for all $n,m > N(\varepsilon)$.

The metric may be thought of therefore as a distance. Thus, if as n approaches infinity the sequence converges to some $\mathbf{x}$, we would like to know if $\mathbf{x}$ is still in the linear space.

DEFINITION 1.4. A metric space M is complete if every Cauchy sequence in M converges to a point in M.

Not every metric space is complete. Consider the class of metrics obtained from norms. The following class of spaces play an important role in analysis:

DEFINITION 1.5. A complete normed linear space B is called a *Banach space*.

Thus, a Banach space is a space of functions on which we have a norm, thus a measure of closeness, and for which all Cauchy sequences converge to members of that space.

The structure of the Banach space was based upon the norm, which was a mapping from L into the set of positive integers. As we introduce more structure, the resulting spaces possess more properties. We now want to introduce one more function called the inner product.

DEFINITION 1.6. A complex vector space H is called an *inner product space* if for each pair $x, y \in H$ there exists a complex number $(\mathbf{x}, \mathbf{y})$ belonging to the field F (the complex numbers). The inner product has the following properties, for $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3 \in H$ and $a \in F$:

1. $(\mathbf{x}_1 + \mathbf{x}_2, \mathbf{x}_3) = (\mathbf{x}_1, \mathbf{x}_2) + (\mathbf{x}_2, \mathbf{x}_3)$
2. $(a\mathbf{x}_1, \mathbf{x}_3) = a(\mathbf{x}_1, \mathbf{x}_3)$
3. $(\mathbf{x}_1, \mathbf{x}_2) = (\mathbf{x}_2, \mathbf{x}_1)^*$

where the asterisk denotes complex conjugate

4. $(\mathbf{x}_1, \mathbf{x}_1) > 0$ iff $\mathbf{x}_1 \neq 0$ and $(\mathbf{x}_1, \mathbf{x}_1) = 0$ iff $\mathbf{x}_1 = 0$.

In general, we shall deal only with real vector spaces; thus, it is sufficient to define the inner product relative to the reals.

Now, given an inner product, we can define a norm from it. The norm is defined as follows:

$$\|\mathbf{x}\| = \sqrt{(\mathbf{x}, \mathbf{x})} \tag{1.3}$$

Thus, given an inner product, we can obtain a norm and from the norm we can obtain a metric $\rho(\mathbf{x},\mathbf{y})$. Specifically,

$$\rho(\mathbf{x}, \mathbf{y}) = \|\mathbf{x}-\mathbf{y}\| = \sqrt{(\mathbf{x}-\mathbf{y})(\mathbf{x}-\mathbf{y})} \tag{1.4}$$

We cannot however go in the reverse order. Thus, we should note that normed spaces or metric spaces where the norm or the metric are obtained from an inner product have special properties that warrant particular study. For this reason we introduce the concept of a Hilbert space.

DEFINITION 1.7. A Banach space whose norm is obtained from an inner product is called a *Hilbert space H*.

Thus, a Hilbert space is a complete inner product space. The inner product is also called the *dot*, or *scalar*, *product*. Physically, scalar products on vector spaces represent projections. It is this property of the inner product that we shall use to develop the special properties of the Hilbert space. The Hilbert

space H is called a *separable Hilbert space* if H contains a countably dense subset X. This implies that if H is separable, then it contains an enumerable number of elements $\mathbf{x}_1, \mathbf{x}_2, \cdots$ such that subspace spanned by $\{\mathbf{x}_i\}$ is identical to H. We shall restrict our attention to separable Hilbert spaces from now on.

We now want to prove the orthogonal projection theorem for Hilbert spaces. To do so, it is first necessary to present several more definitions.

DEFINITION 1.8. A subset M of a linear space L is called a *subspace of* L if M is itself a linear space, relative to addition and scalar multiplication defined on L. A necessary and sufficient condition for $M \subset L$ to be a subspace is that $\mathbf{x} + \mathbf{y} \in M$ and $a\mathbf{x} \in M$ whenever $\mathbf{x}$ and $\mathbf{y} \in M$ and $a \in \mathbf{R}$.

DEFINITION 1.9. A *closed subspace* M *of* H is a subsapce that is a closed set relative to the metric induced on H.

We will also need the definition of a convex set.

DEFINITION 1.10. A set E in a vector space L is said to be *convex* if it has the following property: If $\mathbf{x},\mathbf{y} \in E$ and for positive $p,q \in \mathbf{R}$ and

$$p + q = 1 \tag{1.5}$$

then the quantity

$$\mathbf{z} = p\mathbf{x} + q\mathbf{y} \tag{1.6}$$

belongs to the set E.

Thus, convexity requires that E contain the segments between any two of its points. This is shown in Figure 4.1.

The next property is that of orthogonality.

Figure 4.1 Convex vectors.

DEFINITION 1.11. If $(\mathbf{x},\mathbf{y}) = 0$ for some $\mathbf{x}, \mathbf{y} \in H$, we say that $\mathbf{x}$ is *orthogonal* to $\mathbf{y}$. This can be written $\mathbf{x} \perp \mathbf{y}$. It should be ovbious that

$$(\mathbf{x}, \mathbf{y}) = (\mathbf{y}, \mathbf{x}) = 0 \tag{1.7}$$

so that symmetry holds.

The concept of orthogonality can be visualized better if we consider a particular inner product. Let $\mathbf{x}$ and $\mathbf{y}$ be two vectors in the two-dimensional euclidean vector space. Let $|\mathbf{x}|$ and $|\mathbf{y}|$ be the standard euclidean lengths of these two vectors, and let θ be the angle between them. Then the inner product can be obtained from the vector dot product as follows:

$$(\mathbf{x}, \mathbf{y}) = \mathbf{x} \cdot \mathbf{y} = |\mathbf{x}|\ |\mathbf{y}| \cos \theta \tag{1.8}$$

In Figure 4.2 we show $\mathbf{x} \perp \mathbf{y}$ for all $\mathbf{y}$ belonging to a given plane. We have a set of $\mathbf{y}$ which are orthogonal to $\mathbf{x}$. For example, in Figure 4.2, $\mathbf{y}_1$, $\mathbf{y}_2$, $\mathbf{y}_3$, $\mathbf{y}_4$, $\mathbf{y}_5$ are all orthogonal to $\mathbf{x}$.

Now let us denote by $Y^{\perp}$ as the set:

$$Y^{\perp} = \{\mathbf{y} : (\mathbf{x}, \mathbf{y}) = 0\} \tag{1.9}$$

Figure 4.2 Orthogonality of vectors.

It is a simple matter to show that $Y^{\perp}$ ("y perp") is a subspace, since $\mathbf{x} \perp \mathbf{y}$ and $\mathbf{x} \perp \mathbf{y}'$ imply the closure $\mathbf{x} \perp (\mathbf{y} + \mathbf{y}')$. Furthermore, it can be shown that $Y^{\perp}$ is a closed subspace (Rudin [2], p. 78).

We are now prepared to prove a theorem useful in estimation. To motivate following theorem, we must return briefly to the problem of estimation. Recall that what we are seeking is an estimate of the state of a system $\mathbf{x}$. Call this estimate $\hat{\mathbf{x}}$. The error in our estimate is $\tilde{\mathbf{x}}$ and is merely $\mathbf{x} - \hat{\mathbf{x}}$. Now if we define a cost criterion as an inner product on Hilbert norm, say $\|\tilde{\mathbf{x}}\|$, then the theorem states that there exists a unique error vector $\tilde{\mathbf{x}}$ with a smallest Hilbert norm. We shall later show that the minimum mean square error is a suitable norm.

THEOREM 1.1

Every nonempty closed and convex set E in a Hilbert space H contains a unique element of a smallest norm. That is, there is only one $\mathbf{x}_0 \in E$ such that $\|\mathbf{x}_0\| \leqq \|\mathbf{x}\|$ for all $\mathbf{x} \in E$.

Proof. Now, since we have a norm, which is an inner product, we can express the following equality:

$$\|\mathbf{x} + \mathbf{y}\|^2 + \|\mathbf{x} - \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + 2\|\mathbf{x} \cdot \mathbf{y}\|^2 + \|\mathbf{y}\|^2 + \|\mathbf{x}\|^2 - 2\|\mathbf{x} \cdot \mathbf{y}\|^2 + \|\mathbf{y}\|^2 = 2\|\mathbf{x}\|^2 + 2\|\mathbf{y}\|^2 \tag{1.10}$$

We first want to show that if such an element exists, then it is unique. Let δ be defined as follows:

$$\delta = \inf\{\|\mathbf{x}\| : \mathbf{x} \in E\} \tag{1.11}$$

For any $\mathbf{x}$ and $\mathbf{y} \in E$ we apply the initial identity to $\frac{1}{2}\mathbf{x}$ and $\frac{1}{2}\mathbf{y}$.

$$\tfrac{1}{4}\|\mathbf{x} - \mathbf{y}\|^2 = \tfrac{1}{2}\|\mathbf{x}\|^2 + \tfrac{1}{2}\|\mathbf{y}\|^2 - \left\|\frac{\mathbf{x} + \mathbf{y}}{2}\right\|^2 \tag{1.12}$$

Since E is convex,

$$\frac{\mathbf{x} + \mathbf{y}}{2} \in E$$

Hence,

$$\|\mathbf{x} - \mathbf{y}\|^2 \leqq 2\|\mathbf{x}\|^2 + 2\|\mathbf{y}\|^2 - 4\delta^2 \tag{1.13}$$

since δ is the smallest possible norm. Now if both $\mathbf{x}$ and $\mathbf{y}$ are of the smallest norm, then we have

$$\|\mathbf{x} - \mathbf{y}\|^2 \leqq 0 \tag{1.14}$$

but this can only hold if

$$\mathbf{x} = \mathbf{y} \tag{1.15}$$

which shows that this element is unique.

Now we want to show that such a unique element of a smallest norm really exists. By the definition of δ as the infimum of the set there exists a sequence $\{\mathbf{y}_n\}$ in E such that

$$\|\mathbf{y}_n\| \to \delta \qquad \text{as} \qquad n \to \infty \tag{1.16}$$

But we want to show that this limit $\mathbf{x}_0$ belongs to E. Replace $\mathbf{y}_n$ and $\mathbf{y}_m$ in (1.13) to obtain

$$\|\mathbf{y}_n - \mathbf{y}_m\|^2 \leqq 2\|\mathbf{y}_n\|^2 + 2\|\mathbf{y}_m\|^2 - 4\delta^2 \tag{1.17}$$

Now, as $n, m \to \infty$, we have

$$\lim_{n,\, m \to \infty} \|\mathbf{y}_n - \mathbf{y}_m\|^2 < \varepsilon \tag{1.18}$$

which implies that the sequence is Cauchy. Now recall that H is a complete metric space and all Cauchy sequences converge in H. Therefore, there exists a limit of $\{\mathbf{y}_n\}$ called $\mathbf{x}_0$ that belongs to H. But since $\mathbf{y}_n \in E$ and E is closed—that is, E contains all its accumulation or limit points—then $\mathbf{x}_0$ also belongs to E. Therefore, it follows that

$$\|\mathbf{x}_0\| = \lim_{n \to \infty} \|\mathbf{y}_n\| = \delta \qquad \blacksquare \tag{1.19}$$

In order to develop the concept of orthogonal projection fully it is necessary to introduce the idea of a linear functional.

DEFINITION 1.12. Let L be a linear space. A functional f on the linear space L is a *linear functional* if

$$f(a\mathbf{x} + b\mathbf{y}) = af(\mathbf{x}) + bf(\mathbf{y}) \tag{1.20}$$

for all real numbers a, b and all vectors $\mathbf{x}, \mathbf{y} \in L$.

An example of a linear functional is the expectation operator $E[\ ]$.

We are now going to consider mappings of a function x onto different subspaces. In particular, we want to show that any vector can be uniquely decomposed into two components that are orthogonal. This is similar to the defining of a vector in terms of orthogonal unit vectors. Moreover, we shall prove that $\mathbf{x}$ can be decomposed into a vector belonging to a subspace M plus a vector that belongs to $M^\perp$, a subspace whose members are *all* orthogonal to all those in M. This concept can be shown geometrically in Figure 4.3. Here M is the $\mathbf{x}_1$, $\mathbf{x}_2$ plane and $\mathbf{x}^*$ is one vector in M. Now $\mathbf{x}$ is an arbitrary vector and $\mathbf{x}^\perp$ belongs to $M^\perp$, which is a plane perpendicular to M. The theorem we wish to prove is that such a representation is unique and exists—that is, that there is no other $\mathbf{x}^*$, $\mathbf{x}^\perp$ combination such that if $\mathbf{x}^* \in M$ and $\mathbf{x}^\perp \in M^\perp$ that

$$\mathbf{x} = \mathbf{x}^* + \mathbf{x}^\perp \tag{1.21}$$

Before proving this theorem we shall again motivate it in terms of the

Figure 4.3 Unique decomposition of x.

parameter estimation problem. Let us suppose that we want to estimate $\mathbf{x}$ and all we have is a set of measurements in the $\mathbf{x}_1$, $\mathbf{x}_2$ plane; the estimate will be called $\hat{\mathbf{x}}$ and it will belong to M. The error will be $\tilde{\mathbf{x}}$ and is the difference between $\mathbf{x}$ and $\hat{\mathbf{x}}$ Thus.

$$\tilde{\mathbf{x}} = \mathbf{x} - \hat{\mathbf{x}} \tag{1.22}$$

It should be obvious that $\tilde{\mathbf{x}}$ belongs to H. Furthermore, from the previous theorem there exists in H an $\tilde{\mathbf{x}}$ with the smallest norm, called $\tilde{\mathbf{x}}^*$. Thus, we would like to get this best $\tilde{\mathbf{x}}^*$ or, thus, the best $\hat{\mathbf{x}}^*$. It should be obvious to the reader that the distance $\mathbf{x}^\perp$ is the shortest distance from $\mathbf{x}$ to the plane of M, and $\mathbf{x}^*$ then is the best guess. This is what we now seek to prove rigorously.

THEOREM 1.2

Let M be a closed subspace of H. There exists a unique pair of mappings $P:H \rightarrow M$ and $Q:H \rightarrow M^\perp$ and

$$\mathbf{x} = \mathbf{x}^* + \mathbf{x}^\perp \tag{1.23}$$

where

$$\mathbf{x}^* = P(\mathbf{x}) \tag{1.24}$$

$$\mathbf{x}^\perp = Q(\mathbf{x}) \tag{1.25}$$

These mappings have the following further properties:

1. If $\mathbf{x} \notin M$, then

$$\mathbf{x} = \mathbf{x}^* \quad \text{and} \quad \mathbf{x}^\perp = 0 \tag{1.26}$$

If $\mathbf{x} \notin M^\perp$, then

$$\mathbf{x} = \mathbf{x}^\perp \quad \text{and} \quad \mathbf{x}^* = 0 \tag{1.27}$$

2. $\|\mathbf{x} - \mathbf{x}^*\| = \inf\{\|\mathbf{x} - \mathbf{y}\| : \mathbf{y} \in M\}$ (1.28)

This is the minimization property previously discussed. This says that $\mathbf{x}^\perp$ will then be the minimum "error."

3. $\|\mathbf{x}\|^2 = \|\mathbf{x}^*\|^2 + \|\mathbf{x}^\perp\|^2$ (1.29)

4. P and Q are linear functionals. (Indeed any inner product, as previously defined, is a linear functional on H.)

Now $\mathbf{x}^*$ and $\mathbf{x}^\perp$ are called the *orthogonal projections of H onto M and* $M^\perp$. We will formally write this as

$$\mathbf{x}^* = \text{o.p.}[\mathbf{x}; M] \tag{1.30}$$

and

$$\mathbf{x}^\perp = \text{o.p.}[\mathbf{x}; M^\perp] \tag{1.31}$$

Proof. Now, for any $\mathbf{x} \in H$ we will let

$$\mathbf{x} + M = \{\mathbf{x} + \mathbf{y}; \mathbf{y} \in M\} \tag{1.32}$$

This set is closed and convex. Let us define $\mathbf{x}^\perp$ to be the unique element of smallest norm in $\mathbf{x} + M$. We know this exists by the previous theorem. Now let

$$\mathbf{x}^* = \mathbf{x} - \mathbf{x}^\perp \tag{1.33}$$

Then indeed (1.23) holds. But $\mathbf{x}$ belongs to $\mathbf{x} + M$ and $\mathbf{x}^\perp$ belongs to $\mathbf{x} + M$. Then it should be obvious that

$$\mathbf{x}^* \in M \tag{1.34}$$

Therefore, $P(\mathbf{x})$ maps into M. We must now show that $\mathbf{x}^\perp (Q(\mathbf{x}))$ is orthogonal to all $\mathbf{y}$ that belong to M. That is,

$$(\mathbf{x}^\perp, \mathbf{y}) = 0; \qquad \forall \mathbf{y} \in M \tag{1.35}$$

We shall assume that $\|\mathbf{y}\| = 1$ for simplicity, but it should be obvious that there will be no loss in generality. Since $\mathbf{x}^\perp$ has a minimum inner product form

$$(\mathbf{x}^\perp, \mathbf{x}^\perp) = \|\mathbf{x}^\perp\|^2 \leqq \|\mathbf{x}^\perp - a\mathbf{y}\|^2 = (\mathbf{x}^\perp - a\mathbf{y}, \mathbf{x}^\perp - a\mathbf{y}) \tag{1.36}$$

This should be obvious since by hypothesis it is the smallest norm belonging to $\mathbf{x} + M$. Now, by subtracting any element $\mathbf{y} \in M$, we get the above inequality. Multiplying out the right hand side of (1.36), we obtain

$$(\mathbf{x}^\perp - a\mathbf{y}, \mathbf{x}^\perp - a\mathbf{y}) = (\mathbf{x}^\perp, \mathbf{x}^\perp) - \bar{a}(\mathbf{y}, \mathbf{x}^\perp) - a(\mathbf{x}^\perp, \mathbf{y}) + |a|^2(\mathbf{y}, \mathbf{y}) \tag{1.37}$$

But by assumption

$$(\mathbf{y}, \mathbf{y}) = 1 \tag{1.38}$$

We shall now choose a as

$$a = (\mathbf{x}^{\perp}, \mathbf{y}) \tag{1.39}$$

Now, using this in (1.37) yields

$$(\mathbf{x}^{\perp} - a\mathbf{y}, \mathbf{x}^{\perp} - a\mathbf{y}) = (\mathbf{x}^{\perp}, \mathbf{x}^{\perp}) - |(\mathbf{x}^{\perp}, \mathbf{y})|^2 \tag{1.40}$$

Now the choice of a is not arbitrary, but that of $\mathbf{y}$ is. That is any $\mathbf{y} \in M$ can be expressed as above so that we are assured of generality. Use (1.40) in (1.36) to obtain

$$0 \leqq -|(\mathbf{x}^{\perp}, \mathbf{y})|^2 \tag{1.41}$$

which implies that

$$(\mathbf{x}^{\perp}, \mathbf{y}) = 0 \tag{1.42}$$

for any $\mathbf{y} \in M$, so indeed $\mathbf{x}^{\perp} \perp M$. Therefore, $\mathbf{x}^{\perp} \in M^{\perp}$ by definition. This shows that there exists a decomposition of the form given in (1.23).

We now want to show uniqueness. We can do this by assuming a contradiction. Let us assume that

$$\mathbf{x} = \mathbf{x}_1^* + \mathbf{x}_1^{\perp} \tag{1.43}$$

and

$$\mathbf{x} = \mathbf{x}_2^* + \mathbf{x}_2^{\perp} \tag{1.44}$$

Subtract (1.44) from (1.43):

$$(\mathbf{x}_1^* - \mathbf{x}_2^*) = (\mathbf{x}_2^{\perp} - \mathbf{x}_1^{\perp}) \tag{1.45}$$

Now

$$\begin{aligned} \mathbf{x}_1^* &\in M \\ \mathbf{x}_2^* &\in M \\ \mathbf{x}_1^* - \mathbf{x}_2^* &\in M \end{aligned}$$

and

$$\begin{aligned} \mathbf{x}_1^{\perp} &\in M^{\perp} \\ \mathbf{x}_2^{\perp} &\in M^{\perp} \\ \mathbf{x}_1^{\perp} - \mathbf{x}_2^{\perp} &\in M^{\perp} \end{aligned}$$

Therefore, if we take the inner product of (1.45) with $\mathbf{x}_1^* - \mathbf{x}_2^*$, the right-hand side still is zero. That is.

$$(\mathbf{x}_1^* - \mathbf{x}_2^*, \mathbf{x}_1^* - \mathbf{x}_2^*) = (\mathbf{x}_2^{\perp} - \mathbf{x}_1^{\perp}, \mathbf{x}_1^* - \mathbf{x}_2^*) \tag{1.46}$$

But the right-hand side is zero, which implies that

$$\mathbf{x}_1^* = \mathbf{x}_2^* \tag{1.47}$$

and

$$\mathbf{x}_1^\perp = \mathbf{x}_2^\perp \tag{1.48}$$

which proves uniqueness. Properties (1), (2), and (3) follow directly from the basic properties. For example, property (2) follows simply from the fact that

$$\|\mathbf{x} - \mathbf{x}^*\| = \|\mathbf{x}^\perp\| \tag{1.49}$$

Now, for any other $\mathbf{y} \in M$ we have

$$\|\mathbf{x} - \mathbf{y}\| = \|\mathbf{x}^* + \mathbf{x}^\perp - \mathbf{y}\| \tag{1.50}$$

which can be written as

$$(\mathbf{x}^\perp, \mathbf{x}^\perp) - 2(\mathbf{x}^\perp, \mathbf{x}^* - \mathbf{y}) + (\mathbf{x}^* - \mathbf{y}, \mathbf{x}^* - \mathbf{y}) \tag{1.51}$$

But the center term is zero. Therefore,

$$\|\mathbf{x} - \mathbf{y}\|^2 = \|\mathbf{x}^\perp\|^2 + \|\mathbf{x}^* - \mathbf{y}\|^2 \tag{1.52}$$

Thus,

$$\|\mathbf{x} - \mathbf{x}^*\| \leqq \|\mathbf{x} - \mathbf{y}\|; \qquad \forall \mathbf{y} \in M \tag{1.53}$$

The final property is to show that indeed P and Q are linear. We shall prove this as we did uniqueness. Choose any two vectors $\mathbf{x}, \mathbf{y} \in H$. Then,

$$P(a\mathbf{x}) \in M$$
$$P(b\mathbf{y}) \in M$$
$$P(a\mathbf{x} + b\mathbf{y}) \in M$$

and

$$Q(a\mathbf{x}) \in M^\perp$$
$$Q(b\mathbf{y}) \in M^\perp$$
$$Q(a\mathbf{x} + b\mathbf{y}) \in M^\perp$$

Now recall that

$$a\mathbf{x} + b\mathbf{y} = P(a\mathbf{x} + b\mathbf{y}) + Q(a\mathbf{x} + b\mathbf{y}) \tag{1.54}$$

and

$$a\mathbf{x} = P(a\mathbf{x}) + Q(a\mathbf{x}) \tag{1.55}$$

and

$$b\mathbf{y} = P(b\mathbf{y}) + Q(b\mathbf{y}) \tag{1.56}$$

Now subtract (1.55) and (1.56) from (1.54)

$$P(a\mathbf{x} + b\mathbf{y}) - P(a\mathbf{x}) - P(b\mathbf{y}) = Q(a\mathbf{x} + b\mathbf{y}) - Q(a\mathbf{x}) - Q(b\mathbf{y}) \tag{1.57}$$

Taking the inner product of both sides, we find that by the orthogonality

$$(P(a\mathbf{x} + b\mathbf{y}) - P(a\mathbf{x}) - P(b\mathbf{y}), P(a\mathbf{x} + b\mathbf{y}) - P(a\mathbf{x}) - P(b\mathbf{y})) = 0 \tag{1.58}$$

which implies

$$P(a\mathbf{x} + b\mathbf{y}) = P(a\mathbf{x}) + P(b\mathbf{y}) \tag{1.59}$$

and the same holds true for Q. Thus, P and Q are linear. ∎

Thus, if we have any Hilbert space and a closed subspace of it, we can uniquely decompose any vector into a component in the subspace and a component orthogonal to the subspace, so that the orthogonal component has smallest length. Furthermore, all vectors in the subspace are orthogonal to the orthogonal vector.

We now want to apply this result to our probabilistic space. Recall that if x was a random variable, then x mapped Ω into R, the real line. The probability space was the triple $(\Omega, \mathcal{A}, \mathcal{P})$, where $\mathcal{P}$ was a probability measure and $\mathcal{A}$ a σ-field. An important class of random variables on Ω are those that have finite rth moments.

DEFINITION 1.13. Let m_r be the rth moment of x on $(\Omega, \mathcal{A}, \mathcal{P})$ given by

$$m_r = \int |x(\omega)|^r \, d\mathcal{P}(\omega) \tag{1.60}$$

The space of all random variables whose rth moments are finite are said to form the space L^r over the probability space $(\Omega, \mathcal{A}, \mathcal{P})$.

That is, $x \in L^r$ if $E[|x|^r] < \infty$. A special space of L^r spaces is the L^2 space, the space of functions with finite second moments. We now want to show that L^2 is a Hilbert space—that is, that the norm $E[|x|^2]$ is an inner product and the space L^2 is complete.

THEOREM 1.3

The L^2 space is an inner product space.

Proof. Let $x, y, z \in L^2$ and $a, b \in \mathbf{R}$. Let

$$(x, y) = \int xy^* d\mathcal{P}(\omega) \tag{1.61}$$

Then

$$((ax + by), z) = \int (ax + by)z^* d\mathcal{P}(\omega) = a(x, z) + b(y, z) \tag{1.62}$$

$$(x, y) = \int xy^* \, d\mathcal{P}(\omega) = \int \overline{x^* y} \, d\mathcal{P}(\omega) = \overline{(y, x)} \tag{1.63}$$

$$(x, x) = \int |x|^2 \, d\mathcal{P}(\omega) = 0 \tag{1.64}$$

implies that $x = 0$ (or the suitably generated equivalence class). Thus, the space is an inner product space. ∎

We can now define a norm on the space in the following manner:

$$\|x\| = (x, x)^{1/2} = \left[\int |x|^2 \, d\mathscr{P}(\omega)\right]^{1/2} \tag{1.65}$$

Thus, L^2 is a normed linear space or a normed linear function space since the x are really functions of Ω.

THEOREM 1.4

L^2 spaces are complete.

Proof. It is sufficient to show that if $x_n \in L^2$, then $x_n \to x$ if and only if $\{x_n\}$ is Cauchy. Let us first assume $x_n \to x$ for $x \in L^2$. Now

$$\begin{aligned} E[(x_n - x_m)^2] &= E[(x - x_n - x + x_m)^2] \\ &\leqq E[(x - x_m)^2] + E[(x - x_n)^2] \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \end{aligned} \tag{1.66}$$

since it is assumed that $x_n \to x$. Thus, it is Cauchy.

Now let $\{x_n\}$ be Cauchy and assume that

$$E[(x_n - x_m)^2] \to 0 \tag{1.67}$$

Also by definition

$$\liminf_n |x_n - x_m| = |x - x_m| \tag{1.68}$$

Thus we have

$$\int \liminf_n |x_n - x_m|^2 \, d\mathscr{P} = \int |x - x_m|^2 \, d\mathscr{P} \tag{1.69}$$

Now, using Fatou's lemma (see Halmos [2] or Loeve), we know that for any z_n

$$\int \liminf_n z_n \, d\mathscr{P}(\omega) \leqq \liminf_n \int z_n \, d\mathscr{P}(\omega) \tag{1.70}$$

Thus,

$$\int |x - x_m|^2 \, d\mathscr{P}(\omega) \leqq \liminf_n \int |x_n - x_m|^2 \, d\mathscr{P}(\omega) \tag{1.71}$$

and by assumption the right-hand side converges to zero as n, m approach infinity. Thus, x_m converges to x. To show that $x \in L^2$, we note

$$E[x^2] = E[((x - x_n) + x_n)^2] \leqq E[(x - x_n)^2] + E[x_n^2] \tag{1.72}$$

But, by assumption,

$$E[x_n^2] < \infty \tag{1.73}$$

and by the initial part of the proof

$$E[(x - x_n)^2] < \infty \tag{1.74}$$

Thus,

$$E[x^2] < \infty \tag{1.75}$$

and $x \in L^2$ and L^2 is complete. ∎

The results of the preceding two theorems are presented in the following corollary.

Corollary 1.1. The L^2 space is a Hilbert space.

Thus, the L^2 space is a natural space on which to structure an optimal estimate. If we are given a subspace of L^2, then we know that there exists a unique estimate of the vector in L^2, that projection of that vector upon the subspace.

The following theorem will prove useful in the next section when we desire estimates from different subspaces.

THEOREM 1.5

Let H be a Hilbert space and let M_1 and M_2 be subspaces of H. Let
$M_1 \subset M_2 \subset H$
Then there exist decomposition of vectors x in H such that

(i) $x = x_1^* + x_1^\perp$
$\quad = x_2^* + x_2^\perp$

where $x_1^* \in M_1$, $x_2^* \in M_2$ and

(ii) $\|x_1^\perp\| \geqq \|x_2^\perp\|$

Proof. Part (i) follows immediately from the fact that H is a Hilbert space. Now, since $M_1 \subset M_2$, we know that $x_1^* \in M_2$ as well as M_1. Thus, $x = x_1^* + x_1^\perp$ is a decomposition of x into a vector belonging to M_2 and another vector. But, by the fact that these are Hilbert spaces, there exists a unique minimal representation $x = x_2^* + x_2^\perp$. Thus,

$$\|x_1^\perp\| \geqq \|x_2^\perp\| \tag{1.76}$$

with equality if and only if $M_1 = M_2$, or $x_1^* = x_2^*$. ∎

This implies that as the subspace gets bigger the error gets smaller.

The estimation problem can be phrased as follows. Let $x(\omega)$ be a random variable from Ω into $\mathbf{R}$. Let $x(\omega)$ belong to L_2. M is a subspace of L^2 generated by a set of observations. We seek an estimate $\hat{x}(\hat{x}^*)$ from the subspace M such that

$$x - \hat{x} = \tilde{x} \tag{1.77}$$

is minimized. Thus, since L^2 is a Hilbert space, we know that such an estimate exists and furthermore is unique. We shall discuss this in greater detail in the next section.

4.2 CONDITIONAL EXPECTATION AND MMSE ESTIMATES

The estimation of a random variable from measurements of a continuous-time random process introduces many theoretical complexities that necessitate the introduction of several concepts from measure theory. This allows the results to be stated more precisely and with the rigor that is necessary for a thorough understanding of the analysis. Let us begin by considering a simple discrete-time problem and then from it proceed to the continuous-time version.

A random variable $x(\omega)$ that is a measurable function from Ω, the probability space into $\mathbf{R}$, the real line, is to be estimated based upon a set of n measurements $y(t_1)\cdots y(t_n)$. The measurements are given by

$$y(t_i, \omega) = x(\omega) + n(t_i, \omega) \tag{2.1}$$

where $n(t_i, \omega)$ is a random variable. For this example, we shall assume that $n(t_i, \omega)$ is a measurable function (an event) from $\Omega \times T$ into $\mathbf{R}$. The desired result is an estimate of $x(\omega)$, given the measurements $y(t_i, \omega)$. We let H be the Hilbert space of all measurable functions from Ω into $\mathbf{R}$ that have a finite second moment. The inner product on the space H is the L^2 norm given by

$$\|x - y\|^2 = \int (x(\omega) - y(\omega))^2 \, d\mathscr{P}(\omega) \tag{2.2}$$

where $\mathscr{P}$ is the probability measure on Ω. We define a subspace M as the set of all nonlinear functions of the data set $y(t_i, \omega)$, $G(y(t_1, \omega),\cdots, y(t_n, \omega))$ such that

$$\int \left|G(y(t_1, \omega),\cdots, y(t_n, \omega))\right|^2 d\mathscr{P}(\omega) < \infty \tag{2.3}$$

The MMSE estimation problem for this case can be phrased as follows: We seek the orthogonal projection onto M of the estimate using the given inner product; that is, we want to choose the $G(y(t_1, \omega),\cdots, y(t_n, \omega))$which minimizes the error ε, where

$$\varepsilon = \int [x(\omega) - G(y(t_1, \omega),\cdots, y(t_n, \omega))]^2 \, d\mathscr{P}(\omega) \tag{2.4}$$

Now assume that there exists a joint probability density function for x, $y(t_1)$, $\cdots$, $y(t_n)$. Then ε becomes

$$\varepsilon = \int [u_0 - G(u_1,\cdots,u_n)]^2 p_{x,y}(u_0,\cdots,u_n) \, d\mathbf{u} \tag{2.5}$$

Now let

$$p_{x,y}(u_0; u_1, t_1;\cdots; u_n, t_n) = p_{x/y}(u_0|u_1, t_1;\cdots; u_n, t_n)\, p_y(u_1, t_1;\cdots; u_n, t_n) \tag{2.6}$$

Then ε can be written as

$$\varepsilon = \int \left\{ \int [u_0 - G(u_1, \cdots, u_n)]^2 \, p_{x/y}(u_0 | u_1, t_1, \cdots, u_n, t_n) du_0 \right\} p_y(u_1, t_1, \cdots, u_n, t_n) du_1 \cdots du_n \tag{2.7}$$

But the inner expectation can be written as

$$E[(x - c)^2 | y(t_1) \cdots y(t_n)] \tag{2.8}$$

where c is any function of $y(t_1) \cdots y(t_n)$. Now

$$\begin{aligned} &E[(x - c)^2 | y(t_1) \cdots y(t_n)] \\ &\quad = E[(x - E[x | y(t_1) \cdots y(t_n)] + E[x | y(t_1) \cdots y(t_n)] - c)^2 | y(t_1) \cdots y(t_n)] \\ &\quad = E[(x - E[x | y(t_1) \cdots y(t_n)])^2 | y(t_1) \cdots y(t_n)] \\ &\quad + 2E[(x - E[x | y(t_1) \cdots y(t_n)]) | y(t_1) \cdots y(t_n)] \\ &\quad \cdot E[(E[x | y(t_1) \cdots y(t_n)] - c) | y(t_1) \cdots y(t_n)] \\ &\quad + E[(E[x | y(t_1) \cdots y(t_n)] - c)^2 | y(t_1) \cdots y(t_n)] \end{aligned} \tag{2.9}$$

Taking the expectation and noting that

$$E[(x - E[x | y(t_1) \cdots y(t_n)]) | y(t_1) \cdots y(t_n)] = 0 \tag{2.10}$$

and

$$(E[x | y(t_1) \cdots y(t_n)] - c)^2 \geqq 0 \tag{2.11}$$

we see that ε is minimized if and only if

$$G^*(y(t_1) \cdots y(t_n)) = c = E[x | y(t_1) \cdots y(t_n)] \tag{2.12}$$

Thus the function of the n data points that minimizes the mean square error is the conditional mean.

Now we want to generalize this problem to consider the following case. As before, let $x(\omega)$ be a measurable function from Ω to $\mathbf{R}$. But now let

$$y(s, \omega) = x(\omega) + n(s, \omega) \tag{2.13}$$

where now we are to use $y(s, \omega)$ for $s = t_0$ to $s = t$. Thus, we want an MMSE estimate of $x(\omega)$, given $y(s)$ for $s \in [t_0, t]$. This would imply using the preceding analysis, a conditioning on an infinite but countable set of $y(t_i, \omega)$ (because of the separability assumption of the process $y(t, \omega)$). This type of conditioning, however, is not definable as an extension of the simple density function. To analyze this problem, it is first necessary to return to the initial problem of a finite set of measurements. Let us begin by letting $n = 1$. The conditional expectation is a function of $y(t_1, \omega)$ but since $y(t_1, \omega)$ is also a function of ω we may instead consider the conditional expectation to be a function of ω. Thus, we can write $g(\omega)$ as

$$g(\omega) = E[x(\omega) | y(t_1, \omega)] \tag{2.14}$$

Therefore, the conditional expectation can be considered as a ω-function.

Now let B be a Borel set on $\mathbf{R}$; that is, B is any interval of the form $[a, b)$.

Then let $\mathscr{C}_1$, be the minimum σ-field generated by the Borel sets on $\mathbf{R}$; that is, $\mathscr{C}_1$ consists of ω-sets of the form

$$\{\omega: y(t_1, \omega) \in B\} \tag{2.15}$$

for all Borel sets B. Now let $\mathscr{P}_{\mathscr{C}_1}$ be $\mathscr{P}$ confined to sets in $\mathscr{C}_1$. Thus, $\mathscr{P}_{\mathscr{C}_1}$ is a probability measure on $\mathscr{C}_1$, while $\mathscr{P}$ is the measure an $\mathscr{A}$, the minimum σ-field of the underlying probability space. Now consider the following definition.

DEFINITION 2.1. Let $(\Omega, \mathscr{B})$ be a probability space and let P_1 and P_2 be two probability measures on $\mathscr{B}$. P_1 is *absolutely continuous* with respect to P_2 ($P_1 \ll P_2$) if and only if $P_1(B) = 0$ for all $B \in \mathscr{B}$ for which $P_2(B) = 0$.

Since $\mathscr{C}_1 \subset \mathscr{A}$, by definition, and $\mathscr{P}$ is defined on $\mathscr{A}$, then $\mathscr{P}$ is also defined on $\mathscr{C}_1$. Thus, $\mathscr{P}$ and $\mathscr{P}_{\mathscr{C}_1}$ are two measures defined on $\mathscr{C}_1$. Then it is clear that $\mathscr{P}$ is absolutely continuous with respect to $\mathscr{P}_{\mathscr{C}_1}$. Now define the measure μ on $\mathscr{C}_1$ such that

$$\mu = x\,\mathscr{P} \tag{2.16}$$

where x is a positive random variable. Then if $C \in \mathscr{C}_1$, $\mu(C)$ is

$$\mu(C) = \int_C x\, d\,\mathscr{P} \tag{2.17}$$

Now it is clear that μ is absolutely continuous with respect to $\mathscr{P}_{\mathscr{C}_1}$. We now present the following important theorem called the *Radon-Nikodym theorem.*

THEOREM 2.1

Let μ and $\mathscr{P}_{\mathscr{C}_1}$ be two measures on $(\Omega, \mathscr{C}_1)$ such that $\mu(C) = 0$ for all $C \in \mathscr{C}_1$ for which $\mathscr{P}_{\mathscr{C}_1}(C) = 0$. Then there exists a function f that is measurable with respect to $\mathscr{C}_1$ such that for all $C \in \mathscr{C}_1$

$$\mu(C) = \int_C f\, d\,\mathscr{P}_{\mathscr{C}_1} \tag{2.18}$$

The function f is unique in the sense that if there exists a g satisfying the result, then $f = g$ for all sets except possibly for sets of $\mathscr{P}$ measure zero.

Proof. The proof of this theorem may be found in Halmos [2,] pp. 128–129] or Neveu (p. 111). ∎

Extensions to the case where x is both positive and negative follow directly from measure theoretic arguments. We can explain the theorem in the following fashion: Let us define I_C as the indicator function on $\mathscr{A}$ where

$$I_C(\omega) = \begin{cases} 1; & \omega \in C \\ 0; & \omega \notin C \end{cases} \tag{2.19}$$

Then define the random variable $h(\omega)$ as

$$h(\omega) = x(\omega)\, I_C(\omega) \tag{2.20}$$

Let the probability space be $(\Omega, \mathscr{A}, \mathscr{P})$ so that the expectation of the random variable $h(\omega)$ is

$$E[h(\omega)] = \int_{\Omega} x(\omega)\, I_C(\omega) d\mathscr{P} \tag{2.21}$$

But we know from the previous theorem that there exists a function measurable with respect to $\mathscr{C}_1$ such that

$$\int x(\omega)\, I_C(\omega)\, d\mathscr{P} = \int_C f\, d\mathscr{P}_{\mathscr{C}_1} \tag{2.22}$$

where $\mathscr{P}_{\mathscr{C}_1}$ is the restriction of $\mathscr{P}$ to $\mathscr{C}_1$. We then define f as the conditional expectation of $h(\omega)$, given $\mathscr{C}_1$. That is,

$$f(\omega) = E[h(\omega) \mid \mathscr{C}_1] \tag{2.23}$$

Thus, $f(\omega)$ is a $\mathscr{C}_1$ measurable function, which means the $f(\omega)$ is a random variable on $\mathscr{C}_1$. But that means nothing more than that $f(\omega)$ is a function of $y(t_1, \omega)$, which generated $\mathscr{C}_1$. Thus, we may write

$$\int h(\omega)\, d\mathscr{P} = \int E[h(\omega) \mid \mathscr{C}_1]\, d\mathscr{P}_{\mathscr{C}_1} \tag{2.24}$$

The advantage of this notation is that it makes the conditional expectation a function of the underlying probability space and not of the measurements. Thus the meaning of the $E[h(\omega) \mid y(t_1, \omega) = Y_1]$ is the function $f(\omega)$, where ω belongs to the set C_1, where

$$C_1 = \{\omega : y(t_1, \omega) = Y_1\} \tag{2.25}$$

Thus $E[h(\omega) \mid y(t_1, \omega) = Y_1]$ is an ω-valued function that is constant on the set C_1. Likewise, the inverse image of the sets of $f(\omega)$ belong to $\mathscr{C}_1$ and thus are events. The sets of the form of C_1 are called *atoms* of the sub σ-field $\mathscr{C}_1$. An atom of a σ-field is a set of that σ-field that has no subset of it belonging to the σ-field. Wong [3, p. 27] shows that all sets of the form of $C_1 \in \mathscr{C}_1$ are atoms.

We can now explicitly define the conditional expectation.

DEFINITION 2.2. Let x be a positive random variable on $(\Omega, \mathscr{A}, \mathscr{P})$ and let $\mathscr{C}$ be a sub σ-field of $\mathscr{A}$. The conditional expectation of x with respect to the sub σ-field is the random variable on $(\Omega, \mathscr{C}, \mathscr{P}_{\mathscr{C}})$ such that

$$\int_C x\, d\mathscr{P} = \int_C E[x \mid \mathscr{C}]\, d\mathscr{P}_{\mathscr{C}}; \qquad C \in \mathscr{C} \tag{2.26}$$

where $E[x \mid \mathscr{C}]$ is the conditional expectation.

Now this theorem is valid for any sub σ-field $\mathscr{C}$. It is this fact that allows us to generalize our results from the single measurements, to multiple measurements, and then, finally, to a random variable measured over an interval.

Let $\mathscr{C}_i$ be the minimum σ-field generated by $y(t_1, \omega), \cdots, y(t_i, \omega)$. That is, $\mathscr{C}_i$ consists of the ω-sets

$$\{\omega : [y(t_1, \omega), \cdots, y(t_i, \omega)] \in B\}$$

where B are the Borel sets in $\mathbf{R}^i$. The atoms of $\mathscr{C}_i$ consists of the sets

$$C_i = \{\omega : y(t_1, \omega) = Y_1, \cdots, y(t_i, \omega) = Y_i\} \tag{2.27}$$

Then, from the preceding theorem, the conditional expectation is defined and is a random variable on $\mathscr{C}_i$ given by

$$E[x \mid \mathscr{C}_i]$$

Now let $\mathscr{C}_{i+1}$ be the sub σ-field generated by $y(t_1, \omega), \cdots, y(t_i, \omega), y(t_{i+1}, \omega)$. Then any set C_i in $\mathscr{C}_i$ can be represented by

$$\begin{aligned} C_i &= \{\omega : y(t_1, \omega) \in B_1, \cdots, y(t_i, \omega) \in B_i\} \\ &= \{\omega : y(t_1, \omega) \in B_1, \cdots, y(t_i, \omega) \in B_i, y(t_{i+1}, \omega) \in B_{i+1}\} \end{aligned} \tag{2.28}$$

Thus all sets C_i in $\mathscr{C}_i$ belong to $\mathscr{C}_{i+1}$. Therefore,

$$\mathscr{C}_i \subset \mathscr{C}_{i+1} \tag{2.29}$$

This means that all the events that are possible with i measurements are included in all the possible events with $i+1$ measurements. Now let $f(\omega)$ be a random variable on $\mathscr{A}$. Then let

$$g_k(\omega) = E[f(\omega) \mid \mathscr{C}_k] \tag{2.30}$$

Thus $g_k(\omega)$ is a $\mathscr{C}_k$ measurable function for each k.

The conditional expectation $E[x \mid \mathscr{C}]$ has several properties that we will use. We summarize them in the following theorem.

THEOREM 2.2

Let x be an integrable random variable on $(\Omega, \mathscr{A}, \mathscr{P})$. Let $\mathscr{B}$ and $\mathscr{B}'$ be sub σ-fields of $\mathscr{A}$. Then

1. $$E\left[\sum_{i=1}^{n} a_i x_i \mid \mathscr{B}\right] = \sum_{i=1}^{n} a_i E[x_i \mid \mathscr{B}] \tag{2.31}$$

2. Let $\mathscr{B}' \subset \mathscr{B} \subset \mathscr{A}$. For each random variable x

$$E[E[x \mid \mathscr{B}] \mid \mathscr{B}'] = E[x \mid \mathscr{B}'] \tag{2.32}$$

3. Let x be a random variable and let z be a random variable measurable with respect to $\mathscr{B}$. Then,

$$E[zx \mid \mathscr{B}] = zE[x \mid \mathscr{B}] \tag{2.33}$$

Proof. See Wong [3, pp. 30–31]. ∎

There are two special cases of interest. Namely, if x is $\mathscr{B}$ measurable, then

$$E[x \mid \mathscr{B}] = x \tag{2.34}$$

and if $\mathscr{B} = (\phi, \Omega)$, then

$$E[x|\mathscr{B}] = E[x] \tag{2.35}$$

Now from our previous result we know that since

$$\mathscr{C}_k \subset \mathscr{C}_{k+1} \tag{2.36}$$

then

$$E[g_{k+1}(\omega)|\mathscr{C}_k] = g_k(\omega) \tag{2.37}$$

which follows immediately from the definition of $g_k(\omega)$. We now introduce a generalized definition of a martingale, first discussed in Chapter 3. This definition is from Doob [2] (p. 294).

DEFINITION 2.3. Let $x(t, \omega)$ be a stochastic process with $E[|x(t, \omega)|] < \infty$ for all $t \in T$ and for each $t \in T$ let $\mathscr{B}_t$ be a sub σ-field of ω-sets such that

$$\mathscr{B}_s \subset \mathscr{B}_t$$

for all $s < t$. Then the process $x(t, \omega)$ is a martingale if

$$x(s, \omega) = E[x(t, \omega)|\mathscr{B}_s] \tag{2.38}$$

We denote the martingale by the triple $(x(t), \mathscr{B}_t, T)$.

It should be immediately clear that the definition of a martingale used in Chapter 3 is a special case of the above definition. Further discussions of martingales of this type can be found in Doob [2] (Chapter 7), Kushner [7, pp. 28–34], and Neveu (pp. 130–142). With this definition of a martingale we see that $(g_k(\omega), \mathscr{C}_k, I)$ is a martingale where T is now the set of integers. Now recall that $g_k(\omega)$ is a $\mathscr{C}_k$ measurable function that represents the conditional expectation of the random variable $f(\omega)$, given $y(t_1, \omega) \cdots y(t_k, \omega)$. Now the ultimate object is to obtain the expected value of some $f(\omega)$, given the process $y(t, \omega)$ with $t \in T$, some compact set. Since $y(t,\omega)$ is assumed to be separable, it is thus sufficient to condition on the separating set rather than on an uncountable number of random variables. Thus, by choosing the set $\{t_i\}$ to be the separating set, we then want to show that $g_k(\omega)$ as $k \to \infty$ has meaning. To do this we need the following theorem, called the *martingale convergence theorem*.

THEOREM 2.3

Let $z(\omega)$ be a random variable with $E[|z|] < \infty$ and let

$$\mathscr{B}_1 \subset \mathscr{B}_2 \subset \mathscr{B}_3 \subset \cdots$$

be sub σ-fields of ω-sets. Let $\mathscr{B}_\infty$ be the smallest sub σ-field of ω-sets with

$$\bigcup_{n=1}^{\infty} \mathscr{B}_n \subset \mathscr{B}_\infty \tag{2.39}$$

Then

$$\lim_{n\to\infty} E[z\,|\,\mathscr{B}_n] = E[z\,|\,\mathscr{B}_\infty] \tag{2.40}$$

exists and is unique with probability 1.

For the proof of this theorem, see Doob [2] (p. 331). Clearly, the $z(\omega)$ of the theorem is a martingale, and it is this fact that provides the limiting behavior. The relationship between this result and the choice of the separating set is in Wonham [3, pp. 165–167]. This result now leads us to the following conclusion. If

$$g_k(\omega) = E[x(\omega)\,|\,\mathscr{C}_k] \tag{2.41}$$

where $\mathscr{C}_k$ is the sub σ-field generated by $y(t_1, \omega)$ and so on; then

$$g_\infty(\omega) = \lim_{n\to\infty} g_n(\omega) = E[x(\omega)\,|\,\mathscr{C}_\infty] \tag{2.42}$$

exists and $\mathscr{C}_\infty$ is the sub σ-field generated by the process $y(s, \omega)$ for $s \in [t_0, t]$. Thus, $g_\infty(\omega)$ is a function measurable with respect to $\mathscr{C}_\infty$ and the atoms of $\mathscr{C}_\infty$ are ω-sets representing the individual trajectories that the process $y(s, \omega)$ may take from t_0 to t. To strengthen this fact, we let

$$\mathscr{C}_\infty \equiv O_{t_0,t} \tag{2.43}$$

where now $O_{t_0,t}$ represents the sub σ-field generated by the observation $y(s, \omega)$, $s \in [t_0, t]$. Thus, returning to the problem of estimating $x(\omega)$ from

$$y(s, \omega) = x(\omega) + n(s, \omega); \qquad s \in [t_0, t] \tag{2.44}$$

we now can say that the MMSE estimate is

$$\hat{x}(\omega) = E[x(\omega)\,|\,O_{t_0,t}] \tag{2.45}$$

We summarize this result in the following theorem, along with a detailed proof.

THEOREM 2.4

Let x be a random variable for the σ-field $\mathscr{A}$ on the probability space $(\Omega, \mathscr{A}, \mathscr{P})$. Let $O_{t_0,t}$ be a sub σ-field. Then for any function z that is measurable for $O_{t_0,t}$,

$$E[(x - z)^2] \geqq E[(x - E(x\,|\,O_{t_0,t}))^2] \tag{2.46}$$

or

$$E[x\,|\,O_{t_0,t}]$$

minimizes the MSE.

Proof.

$$\begin{aligned} E[(x - z)^2] &= E[([x - E[x|O_{t_0,t}]] + [E[x|O_{t_0,t}] - z])^2] \\ &= E[(x - E[x\,|\,O_{t_0,t}])^2] \\ &\quad + 2E[(x - E[x\,|\,O_{t_0,t}])(E[x\,|\,O_{t_0,t}] - z)] \\ &\quad + E[(E[x\,|\,O_{t_0,t}] - z)^2] \end{aligned} \tag{2.47}$$

Now since z is $O_{t_0,t}$ measurable and, by definition, $E[x|O_{t_0,t}]$ is also, then

$$f = E[x|O_{t_0,s}] - z \tag{2.48}$$

is $O_{t_0,t}$ measurable. Thus, it is sufficient to consider the expectation

$$E[f(x) - E[x|O_{t_0,t}]] \tag{2.49}$$

Since x is $\mathscr{A}$ measurable and $O_{t_0,t} \subset \mathscr{A}$, then x is also $O_{t_0,t}$ measurable. Therefore

$$E[f(x) - Ex|O_{t_0,t}] = E[E[f(x) - E[x|O_{t_0,t}]|O_{t_0,t}]] \tag{2.50}$$

which follows from the Radon-Nikodym theorem. Now, from the properties of the conditional expectation, we know that for functions f, $O_{t_0,t}$ measurable

$$E[f(x) - E[x|O_{t_0,t}]|O_{t_0,t}] = E[f(x) - E[x|O_{t_0,t}]] = 0 \tag{2.51}$$

which then proves the inequality and the theorem. ∎

The conditional expectation then generates a x-function measurable with respect to $O_{t_0,t}$ and that minimizes the mean square error. Thus, $E[x|O_{t_0,t}]$ is the MMSE estimate of the random variable x. This estimate carries over immediately to the problem where $x(t, \omega)$ is a random process and we wish to estimate $x(t, \omega)$ that is a random variable for all t. Thus, if $y(t, \omega)$ is given by

$$dy(s, \omega) = h(x(s, \omega), s)\, ds + dn(s, \omega) \tag{2.52}$$

for $s \in [t_0, t]$, the previous theorem shows that the MMSE estimate is

$$E[x|O_{t_0,t}]$$

To put this estimation in a Hilbert-space context we define three Hilbert spaces. The first space is the space generated by the random variables on Ω. The second is the space of all nonlinear functions measurable with respect to $O_{t_0,t}$. The third is the linear measurement space.

DEFINITION 2.4. Let $\mathscr{H}$ be the Hilbert space of all random processes from $\Omega \times \mathbf{T}$ into $\mathbf{R}$ with norm

$$\|x\|^2 = \int |x(\omega)|^2\, d\mathscr{P}(\omega) \tag{2.53}$$

for all $x(\omega)$ belonging to $\mathscr{H}$.

Clearly, $\hat{x}(\omega)$, the estimated random variable, belongs to H and so do all measurements $y(t_i, \omega)$. The fact that $\mathscr{H}$ is a Hilbert space follows directly from the fact that all L^2 spaces are Hilbert spaces. Note also that the inner product for two functions $x(\omega)$, $y(\omega) \in \mathscr{H}$ is

$$(x, y) = \int x(\omega) y(\omega)\, d\mathscr{P}(\omega) = E[xy] \tag{2.54}$$

DEFINITION 2.5. The space $\mathscr{H}(y)$ is the Hilbert space of all nonlinear functionals

of the process $y(s, \omega)$, where $s \in [t_0, t]$, and is defined as the space consisting of all random variables that are either finite combinations of random variables $G_n(y(t_0, \omega) \cdots y(t_n, \omega))$, where G_n is a Borel function measurable with respect to $O_{t_0,t}$ or are limits of such combinations. The inner product is defined as

$$(x, y) = E[xy]; \qquad x, y \in \mathcal{N}(y) \tag{2.55}$$

This means that in the case of estimation, $\mathcal{N}(y)$ is the space of all nonlinear but measurable functionals of countably many data points having finite second moments. Clearly, $\mathcal{N}(y)$ is a subspace of $\mathcal{H}$, the space of all random variables. A space of this sort was used by Masani and Wiener (pp. 193–194) in their initial work on nonlinear prediction. Thus, the fact that $\mathcal{N}(y)$ is a closed subspace of $\mathcal{H}$ consisting of all functionals measurable with respect to $O_{t_0,t}$ implies that there exists an orthogonal projection from $\mathcal{H}$ onto $\mathcal{N}(y)$ such that the projection x^* is measurable with respect to $O_{t_0,t}$ and

$$E[(x - x^*)^2] \tag{2.56}$$

is minimized. But we have already shown that $E[x \,|\, O_{t_0,t}]$ minimizes this expression, thus

$$x^* = E[x \,|\, O_{t_0,t}] \tag{2.57}$$

and $E[x \,|\, O_{t_0,t}]$ is the orthogonal projection of $x \in \mathcal{H}$ onto $\mathcal{N}(y)$, where $\mathcal{N}(y)$ is generated by $y(s, \omega)$, $s \in [t_0, t]$.

We can now define one further Hilbert space.

DEFINITION 2.6. The space $\mathcal{L}(y)$ is the Hilbert space of all random variables that are either finite linear combinations of $y(s, \omega)$, $s \in [t_0, t]$ or limits of such combinations. The inner product is defined by

$$(x, y) = E[x, y]; \qquad x, y \in \mathcal{L}(y) \tag{2.58}$$

Thus, it should be immediately obvious that $\mathcal{L}(y)$ generated by $y(s, \omega)$, $s \in [t_0, t]$, is measurable with respect to $O_{t_0,t}$ and that $\mathcal{L}(y)$ is a subspace of $\mathcal{N}(y)$. Therefore, we have

$$\mathcal{L}(y) \subset \mathcal{N}(y) \subset \mathcal{H} \tag{2.59}$$

As with $\mathcal{N}(y)$, we have a unique (up to equivalence classes in the norm) orthogonal projection $\hat{x}$ of x onto $\mathcal{L}(y)$. This is called the *linear estimate* of x, given measurements $y(s, \omega)$.

Since $\mathcal{L}(y)$ consists of all limits of linear combinations of $y(s, \omega)$, then the function

$$x'(t) = \lim_{n \to \infty} \sum_{i=11}^{n} a(t_1, t) y(t_i) \tag{2.60}$$

exists and belongs to $\mathcal{L}(y)$. We define this as

$$x'(t) = \int_{t_0}^{t} h(t, \tau)\, y(\tau)\, d\tau \tag{2.61}$$

if $y(\tau)$ as such exists. This is the nonstationary linear estimate of $x(t, \omega)$ for a given t.

From our results in the previous section we know that the error associated with linear estimates are greater than, or equal to, estimates associated with nonlinear estimates. For the case of Gaussian processes it can be shown that $E[x \mid O_{t_0,t}] \in \mathscr{L}(y)$.

We can now show how the structure of the conditional estimate and the generated Hilbert subspaces relate to the problem of estimation. Let $x(t, \omega)$ be a random process for $t \in T$, where T is some closed subset of the real line. Now let $dy(t, \omega)$ be a measurement given by

$$dy(t, \omega) = h(x(t), \omega)\, dt + dn(t, \omega) \tag{2.62}$$

where $n(t, \omega)$ is a Wiener process. Now we ask ourselves how to best obtain an estimate of $x(t, \omega)$ for a given t, given $y(s, \omega)$ from t_0 to t. To make the qualitative statement quantitative we must provide a cost criteroin that is deterministic. A useful criterion is the expectation of some positive function of the error experienced by obtaining x from the data. Let $\hat{x}(t, \omega)$ be a function measurable with respect to $O_{t_0,t}$, and let it be the estimate. Let $x(t, \omega)$ be the state to be estimated. The error $\tilde{x}(t, \omega)$ is defined by

$$\tilde{x}(t, \omega) = x(t, \omega) - \hat{x}(t, \omega) \tag{2.63}$$

Now let $f(\tilde{x})$ be a nonnegative function of $\tilde{x}$. A quantitative cost function is

$$E[f(\tilde{x})]$$

Now we know that if

$$f(\tilde{x}) = |\tilde{x}|^p; \qquad p \geq 1 \tag{2.64}$$

for $p \neq 2$, then we have an L^p space that is a Banach space but not a Hilbert space. Thus, we are not assured of the existence and uniqueness of a minimum $\tilde{x}$. Therefore, the only suitable choice is

$$E[|\tilde{x}|^2]$$

which provides the structure of a Hilbert space. Therefore, cost criteria that differ from the MSE criterion do not in general possess the properties of Hilbert-space norms. Thus, our object in the theory of estimation is to obtain $E[x(t) \mid O_{t_0,t}]$ or

$$E[x(t) \mid O_{t_0,t}] = \int u p_x(u, t \mid O_{t_0,t})\, du \tag{2.65}$$

the conditional probability density of the process x at time t, given the minimum σ-field generated by the process $y(t, \omega)$. Thus, it is sufficient to obtain $p_x(u, t \mid O_{t_0,t})$ to fully describe the process $\tilde{x}(t, \omega)$. We shall direct our

attention in Chapter 5 toward this effort. The following section is an application of the principle of orthogonal projections to discrete-time processes.

4.3 AN APPLICATION OF ORTHOGONAL PROJECTIONS

In this section we present all the theorems necessary for an understanding of, and ability to use, linear discrete-time filtering theory. This approach is an extension of the results of the orthogonal projection concept of Hilbert spaces generated by random processes. The presentation follows the work of Meditch [2] and Kalman [1].

The prediction problem is first presented with the introduction of an appropriate set of vector spaces. Following this is the discrete-time filtering problem. The techniques employed differ from those used in the next chapter, but this approach yields insight into the structure of the filter, particularly in the discrete-time case. Conversion to a continuous-time structure is performed in Meditch [2] and Problem. 4.10.

We shall assume the following model. Let $\mathbf{x}(k)$ be the random state vector at time kT, where T is an arbitrary sample time. $\mathbf{x}(k)$ is assumed to obey the following recursive relationship:

$$\mathbf{x}(k+1) = \boldsymbol{\Phi}(k+1, k)\,\mathbf{x}(k) + \mathbf{w}(k) \tag{3.1}$$

where $\mathbf{x}(k)$, $\mathbf{x}(k+1)$, $\mathbf{w}(k)$ are $n \times 1$ vectors and $\boldsymbol{\Phi}(k+1, k)$ is a $n \times n$ matrix. Thus, $\mathbf{x}(k)$ is a discrete-time Markov process. We further assume that $\mathbf{x}(t)$ has a finite second moment, so that $\mathbf{x}(k) \in L^2$.

The measurement at time $(k+1)T$ is denoted by

$$\mathbf{z}(k+1) = \mathbf{C}(k+1)\,\mathbf{x}(k+1) + \mathbf{v}(k+1) \tag{3.2}$$

Here $\mathbf{z}(k+1)$ and $\mathbf{v}(k+1)$ are $m \times 1$ vectors and $\mathbf{C}(k+1)$ is an $m \times n$ matrix. $\mathbf{w}(k)$ and $\mathbf{v}(k+1)$ are Gaussian random variables and are assumed to be independent. Furthermore, we assume the noises are zero mean random vectors

$$E[\mathbf{w}(k)] = E[\mathbf{v}(k)] = 0 \tag{3.3}$$

and

$$E[\mathbf{w}(k)\,\mathbf{w}^T(j)] = Q(k)\delta_{jk} \qquad (n \times n) \tag{3.4}$$

where

$$\delta_{jk} = \begin{cases} 0; & j \neq k \\ 1; & j = k \end{cases} \tag{3.5}$$

Also

$$E[\mathbf{v}(k)\,\mathbf{v}^T(j)] = \mathbf{R}(k)\delta_{jk}; \qquad (m \times m) \tag{3.6}$$

We also assume that $\mathbf{w}(k)$ and $\mathbf{v}(j)$ are independent for all k, j. Note also that we must assume that $\|\mathbf{Q}(k)\|$ and $\|\mathbf{R}(k)\|$ are finite. We let $H(k)$ be the Hilbert space of all finite second moment random variables from Ω onto $\mathbf{R}^n$.

We shall first develop the equation for obtaining an optimum predicted estimate.

We begin by defining a new subspace $y(k)$:

$$Y(k) = \{\mathbf{y} : \mathbf{y} = \sum_{i=1}^{k} \mathbf{A}(i)\,\mathbf{z}(i)\}\ \forall\{\mathbf{A}(i)\} \tag{3.7}$$

Where $\mathbf{z}(i)$ are $m \times 1$, and $\mathbf{A}(i)$ are arbitrary $n \times m$ matrices. Note that this subspace can have dimension at most equal to n.

Our new subspace contains all possible linear combinations of the system output from time 0. The estimate we seek is to be *linear*; that is, it is to be a linear combination of the observed system outputs. This estimator has already been shown to be sufficient and unique. Our estimate of the system state must therefore be contained in the space $Y(k)$.

LEMMA 3.1. $Y(k)$, as defined above, is closed under linear transformations:

$$\mathbf{A}\mathbf{y} \in Y(k) \qquad \forall \mathbf{y} \in Y(k)$$

where $\mathbf{A}$ is an arbitrary $n \times n$ matrix.

Proof. $\mathbf{y} \in Y(k) = \mathbf{y} = \sum_{i=1}^{k} \mathbf{A}(i)\,\mathbf{z}(i)$ for some $\{\mathbf{A}(i)\}$

$$\mathbf{A}\mathbf{y} = \mathbf{A} \sum_{i=1}^{k} \mathbf{A}(i)\,\mathbf{z}(i) = \sum_{i=1}^{k} \mathbf{A}'(i)\,\mathbf{z}(i) = \mathbf{y}^* \tag{3.8}$$

Where $\mathbf{A}'(i) = \mathbf{A} \cdot \mathbf{A}(i)$; but $\mathbf{y}^* \in Y(k)$ by definition. ∎

Recall from Section 4.1 (Theorem 1.3) that it is possible to decompose $\mathbf{x}(k)$ uniquely in the following manner†:

$$\mathbf{x}(k) = \mathbf{x}^*(k \mid j) + \mathbf{x}'(k \mid j) \tag{3.9}$$

where $\mathbf{X}^*(k \mid j) \in Y(j)$, and $\mathbf{x}'(k \mid j) \in Y^{\perp}(j)$

We shall now prove that $\mathbf{x}^*(k \mid j)$, which is the orthogonal projection of $\mathbf{x}(k)$ onto $Y(j)$, is indeed the linear estimate which minimizes the cost functional J.

THEOREM 3.1

The cost functional J,

$$J = E[[\mathbf{x}(k) - \hat{\mathbf{x}}(k \mid j)]^T[\mathbf{x}(k) - \hat{\mathbf{x}}(k \mid j)]] \tag{3.10}$$

is minimized for a linear estimate $\hat{\mathbf{x}}(k \mid j)$ when

† $Y(j) = \{\mathbf{y}: E[\mathbf{x}^T\mathbf{y}] = 0\ ;\ \forall \mathbf{x} \in Y(j)\}$. $Y^{\perp}(j)$ is called the *orthogonal complement* of $Y(j)$; if S is the whole space space, the result holds that

$$S = Y(j) \oplus Y^{\perp}(j)$$

$$\hat{\mathbf{x}}(k \mid j) = \text{o.p.}[\mathbf{x}(k);\ Y(j)] \tag{3.11}$$

Proof. Let $\mathbf{y} \notin Y(j)$; consider $\mathbf{y}$ to be an estimate of the state.

$$J = E[[\mathbf{x}(k) - \mathbf{y}]^T[\mathbf{x}(K) - \mathbf{y}]] \tag{3.12}$$

Now, using (3.9), substitute and expand.

$$J = E[\mathbf{x}^*(k \mid j)\,\mathbf{x}'(k \mid j)] + 2E[\mathbf{x}^*(k \mid j)\,(\mathbf{x}^*(k \mid j) - \mathbf{y}] + E[(\mathbf{x}^*(k \mid j) - \mathbf{y})^T(\mathbf{x}^*(k \mid j) - \mathbf{y})] \tag{3.13}$$

But the second term is zero, since $\mathbf{x}^*(k \mid j) - \mathbf{y} \in Y(j)$ and $\mathbf{x}'(k \mid j) \in Y(j)$. So we have

$$J = J_{x'(k|j)} + E\,[[\mathbf{x}^*(k \mid j) - \mathbf{y}]^T\,[\mathbf{x}^*(k \mid j) - \mathbf{y}]] \tag{3.14}$$

Obviously, we can do not better than to set the last term to 0 or to let

$$\mathbf{y} = \mathbf{x}^*(k \mid j) = \hat{\mathbf{x}}(k \mid j) = \text{o.p.}[\mathbf{x}(k);\ Y(j)] \tag{3.15}$$

Now we would like to derive a propagation expression for our estimate, in the form of the following theorem.

THEOREM 3.2

Assume that the best estimate $\hat{\mathbf{x}}(k|k)$, which will be called $\hat{\mathbf{x}}(k)$ for convenience, is given or obtained in some manner. This is the best estimate of the state at time k based upon k observations. Then the predicted optimal estimate of ~~$(k+1)$~~, given data to time k, is

$$\hat{\mathbf{x}}(k + 1 \mid k) = \Phi\,(k + 1, k)\,\hat{\mathbf{x}}(k) \tag{3.16}$$

Proof. We know that $\hat{\mathbf{x}}(k) \in Y(k)$; then $\Phi\,(k + 1, k)\hat{\mathbf{x}}(k) \in Y(k)$ also (see Lemma 3.1).
We want to show that

$$\hat{\mathbf{x}}(k + 1) - \Phi\,(k + 1, k)\,\hat{\mathbf{x}}(k) \perp Y(k) \tag{3.17}$$

which would prove that $\hat{\mathbf{x}}(k + 1 \mid k)$ is truly the orthogonal projection. Recall that $\hat{\mathbf{x}}(k + 1 \mid k)$ and $\tilde{\mathbf{x}}(k + 1 \mid k)$ exist uniquely such that

$$\mathbf{x}(k+1) = \mathbf{x}(k + 1 \mid k) + \tilde{\mathbf{x}}(k + 1 \mid k) \tag{3.18}$$

and $\tilde{\mathbf{x}}\,(k + 1 \mid k) \in Y(k)$ by definition. Thus, if (3.17) holds, the theorem is proven We now proceed to show this to be the case. Let us consider any $\mathbf{y} \notin Y(k)$. Define Γ as

$$\Gamma = E[[(\mathbf{x}(k + 1) - \Phi\,(k + 1, k)\,\hat{\mathbf{x}}(k))]^T\mathbf{y}];\ \mathbf{y} \notin Y(k) \tag{3.19}$$

Now

$$\mathbf{x}(k + 1) = \Phi(k + 1, k)\,\mathbf{x}(k) + \mathbf{w}(k) \tag{3.20}$$

$$\mathbf{x}(k) = \hat{\mathbf{x}}(k) + \tilde{\mathbf{x}}(k) \tag{3.21}$$

(as previously defined). So

$$\Gamma = E\left[[\Phi(k+1,k)\,\tilde{\mathbf{x}}(k)]^T\mathbf{y}\right] + E\left[\mathbf{w}^T(k)\mathbf{y}\right] \tag{3.22}$$

$$\Gamma = E[\tilde{\mathbf{x}}^T(k)\,\Phi^T(k+1,k)\mathbf{y}] + E[\mathbf{w}^T(k)\mathbf{y}] \tag{3.23}$$

In the first term, $\Phi^T(k+1,k)\,\mathbf{y} \in Y(k)$, but $\tilde{\mathbf{x}}^T(k) \perp \mathbf{y}\ \forall \mathbf{y} \in Y(k)$. Thus, the first term is zero, yielding

$$\Gamma = E[\mathbf{w}^T(k)\mathbf{y}] \tag{3.24}$$

But $\mathbf{y}$ is independent of $\mathbf{w}(k)$, as the most recent observation, $\mathbf{z}(k)$, does not depend upon $\mathbf{w}(k)$, since

$$\begin{aligned}\mathbf{z}(k) &= \mathbf{C}(k)\,\mathbf{x}(k) + \mathbf{v}(k)\\ &= \mathbf{C}(k)\,\Phi(k,k-1)\,\mathbf{x}(k-1) + \mathbf{w}(k-1) + \mathbf{v}(k)\end{aligned} \tag{3.25}$$

Therefore, we see $|\Gamma|$ must be zero, so (3.17) holds and $\hat{\mathbf{x}}(k+1|k)$ is truly the unique orthogonal projection. It is therefore the MMSE estimate. ∎

We would also like to define two related covariance matrices:

$$\mathbf{M}(k+1|k) = E[\tilde{\mathbf{x}}(k+1|k)\,\tilde{\mathbf{x}}^T(k+1|k)] \tag{3.26}$$

$$\mathbf{P}(k) = E[\tilde{\mathbf{x}}(k)\,\tilde{\mathbf{x}}^T(k)] \tag{3.27}$$

The relationship between the two can be determined by observing the propagation equation for $\tilde{\mathbf{x}}$:

$$\begin{aligned}\tilde{\mathbf{x}}(k+1|k) &= \mathbf{x}(k+1) - \hat{\mathbf{x}}(k+1|k)\\ &= \mathbf{x}(k+1) - \Phi(k+1,k)\,\hat{\mathbf{x}}(k)\\ &= \Phi(k+1,k)\,\mathbf{x}(k) + \mathbf{w}(k) - \Phi(k+1,k)\,\hat{\mathbf{x}}(k)\end{aligned} \tag{3.28}$$

So

$$\hat{\mathbf{x}}(k+1|k) = \Phi(k+1,k)\,\hat{\mathbf{x}}(k) + \mathbf{w}(k) \tag{3.29}$$

Let us solve for $\mathbf{M}(k+1|k)$:

$$\mathbf{M}(k+1|k) = E[\tilde{\mathbf{x}}(k+1|k)\,\tilde{\mathbf{x}}^T(k+1|k)] \tag{3.30}$$

$$\begin{aligned}\mathbf{M}(k+1|k) &= \Phi(k+1,k)\,E[\tilde{\mathbf{x}}(k)\,\tilde{\mathbf{x}}^T(k)]\,\Phi^T(k+1,k)\\ &\quad + E[\mathbf{w}(k)\,\mathbf{w}^T(k)] + E[\mathbf{w}(k)\,\tilde{\mathbf{x}}^T(k)\,\Phi^T(k+1,k)]\\ &\quad + E[\Phi(k+1,k)\,\tilde{\mathbf{x}}(k)\,\mathbf{w}^T(k)]\end{aligned} \tag{3.31}$$

But $\tilde{\mathbf{x}}(k)$ does not depend upon $\mathbf{w}(k)$, by an argument similar to that of the preceding theorem; thus the last two terms are zero, and using (3.27), we obtain

$$\mathbf{M}(k+1) = \Phi(k+1,k)\,\mathbf{P}(k)\,\Phi^T(k+1,k) + \mathbf{Q}(k) \tag{3.32}$$

Thus, a knowledge of the error covariance $\mathbf{P}(k)$ allows us to predict the error covariance of the estimate, $\mathbf{M}(k+1)$.

We will now derive the optimum filtered estimate $\hat{\mathbf{x}}(k|k)$, or $\hat{\mathbf{x}}(k)$. We shall first examine the behavior of the system output.

DEFINITION 3.1. Let $\hat{\mathbf{z}}(k \mid j)$ be the estimate of the output. It is given by

$$\hat{\mathbf{z}}(k \mid j) = \mathbf{C}(k)\,\hat{\mathbf{x}}(k \mid j) \qquad k > j \tag{3.33}$$

Following from this is the error associated with this estimate. Again define the error as

$$\tilde{\mathbf{z}}(k \mid j) = \mathbf{z}(k) - \hat{\mathbf{z}}(k \mid j) \tag{3.34}$$

Now, if we use $k + 1$ and k, we can then expand and simplify this expression:

$$\tilde{\mathbf{z}}(k + 1 \mid k) = \mathbf{z}(k + 1) - \mathbf{C}(k + 1)\,\hat{\mathbf{x}}(k + 1 \mid k) \tag{3.35}$$

But since

$$\mathbf{z}(k + 1) = \mathbf{C}(k + 1)\,\mathbf{x}(k + 1) + \mathbf{v}(k + 1) \tag{3.36}$$

this yields

$$\tilde{\mathbf{z}}(k + 1 \mid k) = \mathbf{C}(k + 1)\,\tilde{\mathbf{x}}(k + 1 \mid k) + \mathbf{v}(k + 1) \tag{3.37}$$

We can also write it in another fashion if we recall the prediction theorem:

$$\hat{\mathbf{x}}(k + 1 \mid k) = \boldsymbol{\Phi}(k + 1, k)\,\hat{\mathbf{x}}(k) \tag{3.38}$$

Therefore, substituting this yields

$$\tilde{\mathbf{z}}(k + 1 \mid k) = \mathbf{z}(k + 1) - \mathbf{C}(k + 1)\,\boldsymbol{\Phi}(k + 1, k)\,\hat{\mathbf{x}}(k) \tag{3.39}$$

Thus we see that $\tilde{\mathbf{z}}(k + 1 \mid k)$ is the sum of the most recent data $\mathbf{z}(k + 1)$ and of the past best estimate that we obtained from some unknown source. Using this, we shall define yet another vector space.

DEFINITION 3.2. A vector space $Z(k + 1)$ is defined as the set of all z, $n \times 1$ vectors such that

$$Z(k + 1) = \{\mathbf{z}: \mathbf{K}(k + 1)\,\tilde{\mathbf{z}}(k + 1 \mid k) = \mathbf{z}\} \tag{3.40}$$

where $\mathbf{K}(k + 1)$ is an arbitrary $n \times m$ matrix that maps the $\mathbf{z}(k + 1)$ vector and the weighted estimate of $\mathbf{x}(k)$ into an $n \times 1$ vector.

Note that $\hat{\mathbf{x}}(k)$ is also a linear combination of all $z(j)$ from $\mathbf{z}(1)$ to $\mathbf{z}(k)$. Thus, Z is a vector space that is a function of all the $\mathbf{z}(j)$ from $\mathbf{z}(1)$ through the present $\mathbf{z}(k + 1)$.

LEMMA 3.2. The vector spaces $Y(k)$ and $Z(k + 1)$ are orthogonal.

Proof. Choose any $\mathbf{y} \in Y(k)$. Now $\mathbf{z}(k + 1)$ is strictly defined. Thus, to show orthogonality, we must show that

$$\mathrm{E}[\{\mathbf{K}(k + 1)\,\tilde{\mathbf{z}}(k + 1 \mid k)\}^T \mathbf{y}] \tag{3.41}$$

vanishes. Taking the transpose yields

$$E[\tilde{\mathbf{z}}^T(k + 1 \mid k)\,\mathbf{K}^T(k + 1)\mathbf{y}] \tag{3.42}$$

But we showed that $\bar{\mathbf{z}}(k+1|k)$ could be written as a sum of $\bar{\mathbf{x}}(k+1|k)$ and $\mathbf{v}(k+1)$ (3.37). Therefore, (3.42) becomes

$$E[\bar{\mathbf{x}}^T(k+1|k)\,\mathbf{C}^T(k+1)\,\mathbf{K}^T(k+1)\mathbf{y}] + E[\mathbf{v}^T(k+1)\,\mathbf{K}^T(k+1)\mathbf{y}]$$

Now, since $\mathbf{y} \in Y(k)$ is defined as a set over all linear transformations and truly $\mathbf{C}^T(k+1)\,\mathbf{K}^T(k+1)$ is a linear transformation, then

$$\mathbf{C}^T(k+1)\,\mathbf{K}^T(k+1)\mathbf{y} \in Y(k)$$

But it was shown that all $\mathbf{y} \in Y(k)$ were orthogonal to $\bar{\mathbf{x}}(k+1|k)$, since $\bar{\mathbf{x}}(k+1|k) \in Y(k)$. Therefore, the first expectation vanishes. The second expectation vanishes because of a priori uncorrelatedness of the measurement noise. Thus, the total expectation vanishes, proving the lemma. ∎

Before continuing it is necessary to propose one further definition.

DEFINITION 3.3. Let $Y(k)$ and $Z(k+1)$ be two subspaces of a vector space $N(k+1)$. $N(k+1)$ is said to be the *direct sum* of $Y(k)$ and $Z(k+1)$, written

$$N(k+1) = Y(k) \oplus Z(k+1) \tag{3.43}$$

if any $\mathbf{n} \in N(k+1)$ may be written uniquely as $\mathbf{n} = \mathbf{y} + \mathbf{z}$, where $\mathbf{y} \in Y(k)$ and $\mathbf{z} \in Z(k+1)$.

The following lemma is an obvious consequence of this definition.

LEMMA 3.3. The direct sum space $N(k+1)$ is $Y(k+1)$ where

$$Y(k+1) = \{\mathbf{y} : \mathbf{y} = \sum_{i=1}^{k+1} \mathbf{A}(i)\,\mathbf{z}(i)\} \tag{3.44}$$

Proof. Consider any $\mathbf{n} \in N(k+1)$. From (3.43), we can know that

$$\mathbf{n} = \mathbf{y} + \mathbf{z} \qquad \mathbf{y} \in Y(k),\ \mathbf{z} \in Z(k+1) \tag{3.45}$$

From the definition of $Y(k)$ (3.7) and of $Z(k+1)$ (3.40), we can write this as

$$\mathbf{n} = \sum_{i=1}^{k} \mathbf{A}(i)\,\mathbf{z}(i) + \mathbf{K}(k+1)\,\bar{\mathbf{z}}(k+1|k) \tag{3.46}$$

Substituting for $\bar{\mathbf{z}}(k+1|k)$ from (3.35), we have

$$\begin{aligned}\mathbf{n} = &\sum_{i=1}^{k} \mathbf{A}(i)\,\mathbf{z}(i) + \mathbf{K}(k+1)\,\mathbf{z}(k+1) \\ &- \mathbf{K}(k+1)\,\mathbf{C}(k+1)\,\hat{\mathbf{x}}(k+1|k)\end{aligned} \tag{3.47}$$

But $\hat{\mathbf{x}}(k+1|k) = \text{o.p.}\,[\mathbf{x}(k+1);\, Y(k)]$; thus $\hat{\mathbf{x}}(k+1|k) \in Y(k)$. It is then clear that $\mathbf{n}$ can be written

$$\mathbf{n} = \sum_{i=1}^{k+1} \mathbf{A}'(i)\,\mathbf{z}(i) \tag{3.48}$$

Thus, $\mathbf{n} \in Y(k+1)\ \forall \mathbf{n}$, which implies that $N(k+1) \subset Y(k+1)$. Now take any $\mathbf{y} \in Y(k+1)$: it may be written

$$\mathbf{y} = \sum_{i=1}^{k+1} \mathbf{A}(i)\,\mathbf{z}(i) \tag{3.49}$$

We can now follow the preceding steps in reverse order, ending with

$$\mathbf{y} = \sum_{i=1}^{k} \mathbf{A}'(i)\,\mathbf{z}(i) + \mathbf{K}(k+1)\,\tilde{\mathbf{z}}(k+1 \mid k); \qquad \forall \mathbf{y} \in Y(k+1) \tag{3.50}$$

Clearly, then, $Y(k+1) \subset N(k+1)$, and we conclude

$$Y(k+1) = N(k+1) \blacksquare \tag{3.51}$$

Now we shall consider the effect of taking orthogonal projections.

LEMMA 3.4. If $Y(k)$ is orthogonal to $Z(k+1)$, then

$$\begin{aligned} &\text{o.p.}[\mathbf{x}(k+1);\ Y(k) \oplus Z(k+1)] \\ &= \text{o.p.}[\mathbf{x}(k+1);\ Y(k)] + \text{o.p.}[\mathbf{x}(k+1);\ Z(k+1)] \end{aligned} \tag{3.52}$$

Proof. Let

$$\begin{aligned} &\mathbf{y} \in Y(k) \\ &\mathbf{z} \in z(k+1) \end{aligned}$$

Then clearly $\mathbf{y} + \mathbf{z} \in Y(k+1)$. Now let

$$\mathbf{y}^* = \text{o.p.}[\mathbf{x}(k+1);\ Y(k)] \tag{3.53}$$

$$\mathbf{z}^* = \text{o.p.}[\mathbf{x}(k+1);\ Z(k+1)] \tag{3.54}$$

To prove the lemma, we must show that

$$E\{[\mathbf{x}(k+1) - (\mathbf{y}^* + \mathbf{z}^*)]^T (\mathbf{y} + \mathbf{z})\} = 0 \tag{3.55}$$

or that for all $\mathbf{y} + \mathbf{z} \in Y(k+1)$ the vector obtained by subtracting $\mathbf{y}^* + \mathbf{z}^*$ is orthogonal to $Y(k+1)$. Then since this is the definition of the orthogonal projection and since it is unique, then indeed the lemma is true. Recall that since $\mathbf{y}^* \in Y(k)$ and $\mathbf{z}^* \in Z(k+1)$ that $\mathbf{y}^* + \mathbf{z}^* \in Y(k+1)$, the direct sum space. We will now show that $\mathbf{x}(k+1) - (\mathbf{y}^* + \mathbf{z}^*)$ is orthogonal to all $\mathbf{y} \in Y(k+1)$:

$$\begin{aligned} &E[(\mathbf{x}(k+1) - (\mathbf{y}^* + \mathbf{z}^*)]^T (\mathbf{y} + \mathbf{z})] \\ &= E[(\mathbf{x}(k+1) - \mathbf{y}^*)^T \mathbf{y}] + E[(\mathbf{x}(k+1) - \mathbf{z}^*)^T \mathbf{z}] \\ &- E[\mathbf{y}^T \mathbf{z}] - E[\mathbf{z}^T \mathbf{y}] \end{aligned} \tag{3.56}$$

The first two terms are zero for all k since $\mathbf{y} \in Y(k)$, and so does $\mathbf{y}^*$ by the decomposition theorem. Likewise, for $\mathbf{z} \in Z(k+1)$. The last two terms are zero because they belong to mutually orthogonal vector spaces. Thus, the decomposition is obtained.

We are now prepared to present the orthogonality theorem in terms of filtering. It will develop a relationship that is recursive between the predicted

estimate and that unknown quantity $\mathbf{x}(k)$. In the prediction theorem we assumed somehow that $\hat{\mathbf{x}}(k|k)$ was available. Here we shall demonstrate how to obtain it.

THEOREM 3.3

The estimate of $\mathbf{x}(k+1)$ is

$$\hat{\mathbf{x}}(k+1) = \hat{\mathbf{x}}(k+1|k) + \mathbf{K}(k+1)\,\tilde{\mathbf{z}}(k+1|k) \tag{3.57}$$

Proof. $\hat{\mathbf{x}}(k+1) = \text{o.p.}[\mathbf{x}(k+1); Y(k+1)]$ (3.58)

but is also equal to

$$\hat{\mathbf{x}}(k+1) = \text{o.p.}[\mathbf{x}(k+1); Y(k)] + \text{o.p.}[\mathbf{x}(k+1); Z(k+1)] \tag{3.59}$$

which is by the definition of these quantities

$$\hat{\mathbf{x}}(k+1) = \hat{\mathbf{x}}(1+1\,|\,k) + \mathbf{K}(k+1)\,\tilde{\mathbf{z}}(k+1\,|\,k) \blacksquare \tag{3.60}$$

The only problem now is to obtain the value of $\mathbf{K}(k+1)$. This is given in the following theorem.

THEOREM 3.4

The gain matrix $\mathbf{K}(k+1)$ is given by

$$\mathbf{K}(k+1) = \mathbf{M}(k+1)\,\mathbf{C}^T(k+1)[\mathbf{C}(k+1)\,\mathbf{M}(k+1)\,\mathbf{C}^T(k+1) + \mathbf{R}(k+1)]^{-1} \tag{3.61}$$

Proof. Now we know that the $\mathbf{K}(k+1)$ matrix must satisfy the criterion that

$$\mathbf{K}(k+1)\,\tilde{\mathbf{z}}(k+1\,|\,k) = \text{o.p.}[\mathbf{x}(k+1); Z(k+1)] \tag{3.62}$$

Therefore,

$$\mathbf{x}(k+1) - \mathbf{K}(k+1)\,\tilde{\mathbf{z}}(k+1\,|\,k) \tag{3.63}$$

must be orthogonal to all vectors in $Z(k+1)$. Thus, choose any $\mathbf{z} \in Z(k+1)$, for example,

$$\mathbf{z} = \mathbf{B}\tilde{\mathbf{z}}(k+1\,|\,k) \tag{3.64}$$

where $\mathbf{B}$ is any nonzero $n \times m$ matrix.

Then, substituting into sufficiency argument for the orthogonality condition, one obtains*

$$E[\mathbf{x}(k+1)\,\tilde{\mathbf{z}}^T(k+1\,|\,k) - \mathbf{K}(k+1)\,\tilde{\mathbf{z}}(k+1\,|\,k)\,\tilde{\mathbf{z}}^T(k+1\,|\,k)]\,[\mathbf{B}^T] = 0 \tag{3.65}$$

But this must be true for all B matrices, which implies the expectation must be zero.**

*This is a sufficiency argument since all we require is that the trace of the expression be zero. Yet by uniqueness this more stringent requirement will not change the estimate.

**$\mathbf{B}$ may be 0 matrix, but that will not help, since it is evident that this would be a trivial orthogonality.

estimate and that unknown quantity $\mathbf{x}(k)$. In the prediction theorem we assumed somehow that $\hat{\mathbf{x}}(k|k)$ was available. Here we shall demonstrate how to obtain it.

THEOREM 3.3

The estimate of $\mathbf{x}(k+1)$ is

$$\hat{\mathbf{x}}(k+1) = \hat{\mathbf{x}}(k+1|k) + \mathbf{K}(k+1)\,\tilde{\mathbf{z}}(k+1|k) \tag{3.57}$$

Proof. $\hat{\mathbf{x}}(k+1) = \text{o.p.}[\mathbf{x}(k+1);\, Y(k+1)]$ (3.58)

but is also equal to

$$\hat{\mathbf{x}}(k+1) = \text{o.p.}[\mathbf{x}(k+1);\, Y(k)] + \text{o.p.}[\mathbf{x}(k+1);\, Z(k+1)] \tag{3.59}$$

which is by the definition of these quantities

$$\hat{\mathbf{x}}(k+1) = \hat{\mathbf{x}}(1+1\,|\,k) + \mathbf{K}(k+1)\,\tilde{\mathbf{z}}(k+1\,|\,k)\,\blacksquare \tag{3.60}$$

The only problem now is to obtain the value of $\mathbf{K}(k+1)$. This is given in the following theorem.

THEOREM 3.4

The gain matrix $\mathbf{K}(k+1)$ is given by

$$\mathbf{K}(k+1) = \mathbf{M}(k+1)\,\mathbf{C}^T(k+1)[\mathbf{C}(k+1)\,\mathbf{M}(k+1)\,\mathbf{C}^T(k+1) + \mathbf{R}(k+1)]^{-1} \tag{3.61}$$

Proof. Now we know that the $\mathbf{K}(k+1)$ matrix must satisfy the criterion that

$$\mathbf{K}(k+1)\,\tilde{\mathbf{z}}(k+1\,|\,k) = \text{o.p.}[\mathbf{x}(k+1);\, Z(k+1)] \tag{3.62}$$

Therefore,

$$\mathbf{x}(k+1) - \mathbf{K}(k+1)\,\tilde{\mathbf{z}}(k+1\,|\,k) \tag{3.63}$$

must be orthogonal to all vectors in $Z(k+1)$. Thus, choose any $\mathbf{z} \in Z(k+1)$, for example,

$$\mathbf{z} = \mathbf{B}\tilde{\mathbf{z}}(k+1\,|\,k) \tag{3.64}$$

where $\mathbf{B}$ is any nonzero $n \times m$ matrix.

Then, substituting into sufficiency argument for the orthogonality condition, one obtains*

$$E[\mathbf{x}(k+1)\,\tilde{\mathbf{z}}^T(k+1\,|\,k) - \mathbf{K}(k+1)\,\tilde{\mathbf{z}}(k+1\,|\,k)\,\tilde{\mathbf{z}}^T(k+1\,|\,k)]\,[\mathbf{B}^T] = 0 \tag{3.65}$$

But this must be true for all B matrices, which implies the expectation must be zero.**

*This is a sufficiency argument since all we require is that the trace of the expression be zero. Yet by uniqueness this more stringent requirement will not change the estimate.

**$\mathbf{B}$ may be 0 matrix, but that will not help, since it is evident that this would be a trivial orthogonality.

Substitute into the second term

$$\tilde{\mathbf{z}}(k+1\,|\,k) = \mathbf{C}(k+1)\,\tilde{\mathbf{x}}(k+1\,|\,k) + \mathbf{v}(k+1) \tag{3.66}$$

It is clear, then, that

$$\begin{aligned} &E[\tilde{\mathbf{z}}(k+1\,|\,k)\,\tilde{\mathbf{z}}^T(k+1\,|\,k)] \\ &= \mathbf{C}(k+1)\,\mathbf{M}(k+1)\,\mathbf{C}^T(k+1) + \mathbf{R}(k+1) \end{aligned} \tag{3.67}$$

The vector $\mathbf{x}(k+1)$ can be decomposed as

$$\mathbf{x}(k+1) = \hat{\mathbf{x}}(k+1\,|\,k) + \tilde{\mathbf{x}}(k+1\,|\,k) \tag{3.68}$$

but $\mathbf{z}(k+1\,|\,k)$ is orthogonal to $\mathbf{x}(k+1\,|\,k)$, so we can ignore the term containing their product. Therefore, it suffices to evaluate

$$E[\tilde{\mathbf{x}}(k+1\,|\,k)\,\tilde{\mathbf{z}}^T(k+1\,|\,k)] \tag{3.69}$$

which equals

$$\mathbf{M}(k+1)\,\mathbf{C}^T(k+1)$$

Thus, $\mathbf{K}(k+1)$ is obtained by manipulation of these results since

$$\begin{aligned} &\mathbf{M}(k+1)\,\mathbf{C}^T(k+1) - \mathbf{K}(k+1)\,[\mathbf{C}(k+1) \\ &\mathbf{M}(k+1)\,\mathbf{C}^T(k+1) + \mathbf{R}(k+1)] = 0 \end{aligned} \tag{3.70}$$

Manipulation proves the theorem. ∎

The last problem is obtaining $\mathbf{P}(k+1)$ so that we can calculate the propagation of $\mathbf{M}(k)$. Recall we have shown (3.32) that

$$\mathbf{M}(k+1) = \boldsymbol{\Phi}(k+1,k)\,\mathbf{P}(k)\,\boldsymbol{\Phi}^T(k+1,k) + \mathbf{Q}(k) \tag{3.71}$$

We now wish to evaluate

$$\mathbf{P}(k+1) = E[\tilde{\mathbf{x}}(k+1)\,\tilde{\mathbf{x}}^T(k+1)] \tag{3.72}$$

Now we will compute $\tilde{\mathbf{x}}(k+1)$

$$\tilde{\mathbf{x}}(k+1) = \mathbf{x}(k+1) - \hat{\mathbf{x}}(k+1) \tag{3.73}$$

But we know that $\hat{\mathbf{x}}(k+1)$ is given by the Kalman filter

$$\begin{aligned} \hat{\mathbf{x}}(k+1) = \mathbf{x}(k+1) &- \hat{\mathbf{x}}(k+1\,|\,k) - \mathbf{K}(k+1)\,\mathbf{C}(k+1)\,\tilde{\mathbf{x}}(k+1\,|\,k) \\ &- \mathbf{K}(k+1)\,\mathbf{v}(k+1) \end{aligned} \tag{3.74}$$

Combining and recalling the definition of $\hat{\mathbf{x}}(k+1\,|\,k)$ yields

$$\begin{aligned} \tilde{\mathbf{x}}(k+1) &= [\mathbf{I} - \mathbf{K}(k+1)\,\mathbf{C}(k+1)]\,\tilde{\mathbf{x}}(k+1\,|\,k) \\ &- \mathbf{K}(k+1)\,\mathbf{v}(k+1) \end{aligned} \tag{3.75}$$

Using (3.75) and taking the expectation in (3.72), it is obvious that

$$\begin{aligned} \mathbf{P}(k+1) = [\mathbf{I} - \mathbf{K}(k+1)\,\mathbf{C}(k+1)]\,\mathbf{M}(k+1)\,[\mathbf{I} - \mathbf{K}(k+1)\,\mathbf{C}(k+1)]^T \\ + \mathbf{K}(k+1)\,\mathbf{R}(k+1)\,\mathbf{K}^T(k+1) \end{aligned} \tag{3.76}$$

Figure 4.4 Computational algorithm for Discrete Time Kalman filter.

This then completes all that is necessary for computation of the Kalman filter. The only remaining question is how to implement it. We start with $\hat{\mathbf{x}}(0)$ as the mean of $\mathbf{x}(0)$. We also have a prior $\mathbf{P}(0)$. This yields $\mathbf{M}(1)$, which in turn yields $\mathbf{K}(1)$, and so on. Thus, we have developed a recursive relationship. The order of calculations is shown in Figure 4.4. and the filter implementation in Figure 4.5.

In conclusion, we have in this section structured the recursive nature of obtaining estimates of $\mathbf{x}(k+1)$ given data up to, and including, $\mathbf{z}(k+1)$. One last point would be to rephrase (3.57) in terms of the previous estimate and the incoming data. Using (3.38) and (3.39) in (3.57), we obtain

$$\begin{aligned}\hat{\mathbf{x}}(k+1) = {} & [\mathbf{I} - \mathbf{K}(k+1)\,\mathbf{C}(k+1)]\,\boldsymbol{\Phi}(k+1,k)\,\hat{\mathbf{x}}(k) \\ & + \mathbf{K}(k+1)\,\mathbf{z}(k+1)\end{aligned} \tag{3.77}$$

This then completes the discrete case of the filtering problem.

4.4. CONCLUSIONS

In the preceding three sections we have emphasized the MMSE criterion and have shown that its choice allows us to use the structure of the Hilbert space to prove its properties. Thus, the first section went to great lengths to

Figure 4.5 The discrete system and discrete filter.

define the Hilbert space and develop its properties. The Hilbert space used in estimation theory is the L^2 space of random variables with bounded second moments. In the second section it was shown that the conditional mean was the MMSE estimate of a random variable $x(\omega)$ given measurements. Specifically, the interpretation of the conditional mean as an ω function measurable with respect to $O_{t_0,t}$, the minimum σ-field generated by the observation set, provided the basis for estimation, given a random process. We then proceeded to define two important Hilbert subspaces generated by the measurements and related the results of the first section to the MMSE estimator.

The third section was an exposition of the usefulness of the Hilbert space results to the problem of obtaining estimates of discrete-time Markov processes. The main tool in this analysis was the orthogonal projection lemma, which states that the error must be orthogonal to all elements in the Hilbert subspace generated by the measurements.

There are several other areas worth mentioning that rest upon the Hilbert-space interpretation. The first of these is the reproducing kernel Hilbert space (RKHS) analysis. The RKHS were first introduced by Aronszajn and were later used extensively by Parzen [2–4] to obtain estimation structures. They have been expanded upon by Dutweiler and by Kailath [7], who used them to obtain results both in estimation and detection of random processes.

The properties of RKHS depend upon the properties of covariance matrices and the Hilbert spaces generated thereupon.

An interesting and sometimes useful analysis uses the concept of *pseudo-inverses*, also called *generalized inverses*. These quantities can be thought of as follows: Let $\mathbf{y}$ represent a vector in some Hilbert subspace M and let $\mathbf{y}$ represent a measurement. The quantity to be estimated is $\mathbf{x}$, and it is an element of the Hilbert space $H \supset M$. Now, in many cases there exists a linear functional $\mathbf{A}$ that maps $\mathbf{x}$ into $\mathbf{y}$.

$$\mathbf{y} = \mathbf{A}\mathbf{x} \tag{4.1}$$

If the functional $\mathbf{A}$ had an inverse, then

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{y} \tag{4.2}$$

and the estimation problem could be solved directly. Unfortunately, this is not the case. Yet we can define an operator $\mathbf{A}^{+}$ called the pseudoinverse that will yield an estimate $\hat{\mathbf{x}}$

$$\hat{\mathbf{x}} = \mathbf{A}^{+}\mathbf{y} \tag{4.3}$$

such that the projection $\hat{\mathbf{x}}$ has minimum error $\mathbf{x} - \hat{\mathbf{x}}$. The concept of such inverses was introduced by Penrose [1,2] and were applied by Greville [1,2]. Foster applied these to the filtering problem. Zadeh and Desoer, and Desoer and Whalen, interpret the pseudoinverse $\mathbf{A}^{+}$ as an operator within the context of Hilbert spaces. Ben-Israel and Charnes provide further extensions of Desoer's work, while Deutsch (pp. 82–89) and Kalman [3] apply this directly to the problem of estimation. Pseudoinverses provide a valuable intuitive basis for orthogonal projections and estimation but do not yield a general enough structure for nonlinear estimation.

Some of the original work in estimation was based upon linear estimates, that is, estimates based upon linear functions of the data. The concept of the Wold decomposition theorem has played an important role in this theory (see Cramer and Leadbetter). To describe this theorem, consider the problem of estimating a random variable $x(\omega)$ from measurements $y(s, \omega)$. Let $H(y)$ be the Hilbert subspace generated by all linear combinations of $y(s, \omega)$, $s < \infty$. Let $H(y,t)$ be the Hilbert subspace generated by all linear combinations of $y(s, \omega)$, $s \leqq t$. Then we have

$$H(y, -\infty) \subset H(y, t) \subset H(y) \tag{4.4}$$

Now a process is deterministic if

$$H(y, -\infty) = H(y) \tag{4.5}$$

and is purely nondeterministic if

$$H(y, -\infty) = \phi \tag{4.6}$$

where ϕ is the null set. We know from our results of this chapter that the linear estimates of x, $E[x \mid H(y, t)]$ are the linear MMSE estimates (e.g., $\hat{x}$). Furthermore, we know that the errors are decreasing as $H(y, t)$ increases. If, however, y is deterministic, then we can never hope to learn anything from such a process beyond our a priori knowledge. The Wold decomposition states that any process $y(t)$ can be decomposed as

$$y(t) = \eta(t) + \nu(t) \tag{4.7}$$

where $\eta(t)$ is deterministic and $\nu(t)$ is nondeterministic. Thus, knowing only $\nu(t)$ is sufficient to improve our knowledge: $\eta(t)$ is irrelevant. The process $\nu(t)$ is called the *innovations*, because it provides the new knowledge. This interpretation has been used by Kailath [2] and Kailath and Frost [1,2] to obtain both the linear and nonlinear continuous-time estimates. More advanced results on finite past nonstationary estimation along these lines is presented by Dudley [1] and Dym and McKean.

As a final comment we mention the results concerning other cost criteria. As we mentioned, a cost criterion must be something that weights the error $x - \hat{x}$, so that a measure of performance can be obtained. The mean square error

$$E[(x - \hat{x})^2] \tag{4.8}$$

is useful because of the L^2 properties. However, other criteria such as

$$E[|x - \hat{x}|] \tag{4.9}$$

$$E[|x - \hat{x}|^3] \tag{4.10}$$

may be used. Yet L^1 and L^3 are Banach spaces but not Hilbert spaces, and existence and uniqueness of orthogonal projections cannot be proved (they do not exist). Yet there are some results concerning other criteria subject to constraints on the probability measures. These are based on the work of Anderson and Sherman [1,2]. They are discussed at length in the texts of Deutsch (pp. 19–23), Van Trees [1, pp. 54–63], Jaswinski [2, pp. 145–150], and Sage and Melsa (pp. 180–182).

We shall now use the results of this chapter in the next chapter to obtain $p_x(u, t \mid \phi_{t_0,t})$ and thus the MMSE.

4.5 PROBLEMS

4.1. Show that L^1, L^3, and L^∞ are Banach Spaces. Recall that the L^∞ norm is the sup norm, where

$$\|f\|_\infty = \sup_{t \in T} f(t)$$

where T is the set on which the random process $f(t)$ is defined.

4.2. A scalar Gaussian random process $x(t)$ is of zero mean and has covariance $K(t,s)$ where

$$E[x(t)x(s)] = K(t, s) \qquad (t, s \in [0, T])$$

is defined on a separable Hilbert space H.

(a) Let $\{\varphi_i(t)\}$ be a set of orthonormal functions such that

$$\int_0^T \varphi_i(t)\,\varphi_j(t)\,dt = \delta_{ij}$$

Show that the random variables $\{x_n\}$ are independent if and only if

$$x(t) = \sum_{i=1}^{\infty} x_i \varphi_i(t)$$

and

$$\varphi_i(t) = \lambda_i \int_0^T K(t, s)\,\varphi_i(s)\,ds$$

(this is called the Karhunen-Loeve expansion).

(b) Show that under the L^2 norm

$$\|x(t) - x_n(t)\| \to 0$$

where

$$x_n(t) = \sum_{i=1}^{n} x_i \varphi_i(t)$$

(c) Prove that

$$K(s, t) = \sum_{i=1}^{\infty} \lambda_i \varphi_i(t) \varphi_i(s)$$

4.3.* Let x be a random variable taking on three values, 0, 4, and 16 with probabilities $\frac{1}{2}$, $\frac{1}{4}$, and $\frac{1}{4}$, respectively. Find the value of c that minimizes $E[|x - c|]$. Find $E[x]$ and show that $E[x]$ does not minimize this quantity.

4.4. [Dutweiler] Let H_1 and H_2 be two separable Hilbert spaces and let $\{f_i\}_{i=1}^{\infty}$ and $\{g_i\}_{i=1}^{\infty}$ be dense sets in H_1 and H_2, respectively. Let $< , >_{H_i}$ denote the inner production on H_1. If for all i,j

$$< f_i, f_j >_{H_1} = < g_i, g_j >_{H_2}$$

Show that there exists a one-to-one linear and onto mapping T: $H_1 \to H_2$ that is,

$$T(f_i) = g_i \qquad (\forall i \in N^+)$$

4.5. Consider the estimation problem discussed in Section 4.2, where

$$y(s, \omega) = x(\omega) + n(s, \omega) \qquad (s \in [t_0, t])$$

*Suggested by Prof. R. M. Dudley.

and the MMSE estimate is

$$\hat{x}(\omega) = E[x(\omega) \,|\, \mathcal{O}_{t_0,t}]$$

Show that $\hat{x}(\omega)$ is an unbiased estimate of $x(\omega)$.

4.6. Consider a set of random variables $\{r_i\}$ taking on one of two possible forms:

$$H_0: r_i = n_i$$
$$H_1: r_i = s + n_i$$

where H_j signifies the jth hypothesis and $\{n_i\}$ is a sequence of random variables. Let $i = 1, \cdots, n$, where n is finite, and let

$$p_{r|H_i}(u_i \cdots u_n \,|\, H_i)$$

be the joint conditional probability of $\{r_i\}_{i=1}^n$, given hypothesis H_i. Define the random variable x_n as

$$x_n = \frac{p_{r|H_1}(r_1 \cdots r_n \,|\, H_1)}{p_{r|H_0}(r_1 \cdots r_n \,|\, H_0)}$$

Clearly, x_n is a function of $\{r_i\}_{i=1}^n$. x_n is called the *likelihood ratio*.

(a) Show that x_n is a martingale; that is, show that

$$E[x_{n+1} \,|\, r_1 \cdots r_n] = x_n$$

(b) Now let

$$H_0: r(t) = n(t)$$
$$H_1: r(t) = s(t) + n(t)$$

where $n(t)$ is a random process on some set T and $s(t)$ is a known function on T. Let $n(t)$ be expanded—assume $n(t) \in L^2(\Omega, \mathcal{A}, P)$ and $L^2(\Omega, \mathcal{A}, P)$ is separable—in a series

$$n(t) = \sum_{i=1}^{\infty} n_i \varphi_i(t) \text{ (l.i.m.)}$$

where $\{\varphi_i(t)\}$ is a complete orthonormal set, (e.g., a Karhunen-Loeve expansion). Let

$$n_K(t) = \sum_{i=1}^{K} n_i \varphi_i(t)$$

$$s_K(t) = \sum_{i=1}^{K} s_i \varphi_i(t)$$

and let

$$x_K = \frac{p_{r_K|H_1}(r_1 \cdots r_K \,|\, H_1)}{p_{r_K|H_0}(r_1 \cdots r_K \,|\, H_0)}$$

where

$$r_K(t) = \sum_{i=1}^{K} r_i \varphi_i(t)$$

Using the martingale convergence theorem, show that

$$\lim_{K \to \infty} x_K = x$$

exists and is unique.

(c) Find the limiting form of x (see Problem 7.6).

4.7. Let $\mathbf{x}(k)$ be a discrete Markov process given by

$$\mathbf{x}(k+1) = \boldsymbol{\Phi}(k+1, k)\,\mathbf{x}(k) + \mathbf{u}(k)$$

and let $\mathbf{y}(k+1)$ be an observation given by

$$\mathbf{y}(k+1) = \mathbf{C}(k+1)\,\mathbf{x}(k+1) + \mathbf{v}(k+1)$$

Assume $\mathbf{u}(k)$, $\mathbf{v}(k)$ are zero mean Gaussian processes with

$$E[\mathbf{u}(k)\,\mathbf{u}^T(j)] = \mathbf{Q}(k)\,\delta_{jk}$$
$$E[\mathbf{v}(k)\,\mathbf{v}^T(j)] = \mathbf{R}(k)\,\delta_{jk}$$

(a) Show that $p(\mathbf{x}(k+1) \mid \mathbf{y}(k+1) \cdots \mathbf{y}(0))$ is a linear functional of $\mathbf{y}(k+1) \cdots \mathbf{y}(0)$.

(b) Is this true for all Gaussian processes of this form? Does $\mathbf{x}(k)$ need to be Markov?

4.8. Let $y(i)$ be a set of measurements of the form

$$y(i) = x + w(i) \qquad (i = 1, \cdots, N)$$

where the $w(i)$ are independent Gaussian random variables with mean 0 and variance σ_i^2.

(a) Find the minimum mean square estimate of x, given $y(i)$, $i = 1, \cdots, N$.

(b) Find the variance of the estimate of x as a function of N.

Use the Kalman filter approach developed in Section 4.3.

4.9. Consider the problem of estimating a scalar parameter x, given n measurements of the form

$$z(i) = x + w(i)$$

where the $w(i)$ are zero mean independent Gaussian random variables with covariance $R(i)$.

(a) Show that $x(n) = E[x \mid z(1) \cdots z(n)]$ is a linear function of the $z(i)$'s.

(b) Generalize this to the case where

$$x(i+1) = \Phi(i+1, i)x(i) + u(i)$$

where $u(i)$ is a zero mean scalar random variable with

$$E[u(i)u(j)] = Q(i)\,\delta_{ij}$$

(c) Use the results of (a) and (b) to show that for any estimation problem where $x(k)$ is Gaussian and the measurement are Gaussian and linearly related to $x(k)$,

$$x(n) = E[x(n) \mid z(1) \cdots z(n)]$$

is a linear function of the $z(i)$'s.

4.10. Consider the discrete estimation problem in Section 4.3. Let T, the sample time, be Δt and let

$$\mathbf{R}(k) = \frac{\mathbf{R}(t)}{\Delta t} \qquad (t = kT)$$

$$\mathbf{Q}(k) = \mathbf{Q}(t)\,\Delta t$$

$$\Phi(k+1, k) = \mathbf{I} + \mathbf{A}(t)\Delta t$$

Show that as $\Delta t \to 0$, the estimation equations become

$$\frac{d\hat{\mathbf{x}}}{dt} = \mathbf{A}(t)\hat{\mathbf{x}}(t) + \mathbf{P}(t)\mathbf{C}^T(t)\mathbf{R}^{-1}(t)[\mathbf{z}(t) - \mathbf{C}(t)\hat{\mathbf{x}}(t)]$$

$$\frac{d\mathbf{P}}{dt} = \mathbf{P}(t)\mathbf{A}^T(t) + \mathbf{A}(t)\mathbf{P}(t) + \mathbf{Q}(t) - \mathbf{P}(t)\mathbf{C}^T(t)\mathbf{R}^{-1}(t)\mathbf{C}(t)\mathbf{P}(t)$$

4.11. Show that $\mathbf{K}(k+1)$ can be given by

$$\mathbf{K}(k+1) = \mathbf{P}(k+1)\mathbf{C}^T(k+1)\mathbf{R}^{-1}(k+1)$$

and rewrite the equation for $\hat{\mathbf{x}}(k+1)$ using this substitution.

4.12. Let the model be as in (3.1), (3.2). But now assume that

$$E[\mathbf{w}(k)] = \mathbf{u}(k)$$
$$E[\mathbf{v}(k)] = \mathbf{m}(k)$$

Evaluate the discrete-time Kalman filter.

4.13. Consider the model of Section 4.2 with equations (3.1) and (3.2). Now assume that $\mathbf{w}(k)$ and $\mathbf{v}(k)$ are not independent but that

$$E[\mathbf{w}(k)\mathbf{v}(j)] = \mathbf{S}(k)\,\delta_{jk}$$

Evaluate the discrete-time Kalman filter for this case.

4.14. Consider the discrete-time model

$$\mathbf{x}(k+1) = \Phi(k+1, k)\mathbf{x}(k) + \mathbf{B}(k)\mathbf{w}(k) + \mathbf{E}(k)\mathbf{u}(k)$$

where $\mathbf{w}(k)$ is a $g \times 1$ Gaussian sequence with

$$E[\mathbf{w}(k)\mathbf{w}^T(j)] = \mathbf{Q}(k)\,\delta_{jk}$$

and $\mathbf{u}(k)$ is a $r \times 1$ deterministic vector. Assume

$$\mathbf{z}(k+1) = \mathbf{C}(k+1)\,\mathbf{x}(k+1) + \mathbf{v}(k+1)$$

is as in Section 4.3. Determine the discrete-time Kalman filter for this problem.

4.15. An $n \times m$ matrix $\mathbf{M}$ of rank r can be written as

$$\mathbf{M} = \mathbf{BC}$$

where $\mathbf{B}$ is $n \times r$ and $\mathbf{C}$ is $r \times m$. The Moore-Penrose generalized inverse is defined as

$$\mathbf{M}^{+} = \mathbf{C}^T(\mathbf{CC}^T)^{-1}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T$$

(a) A generalized inverse of a matrix $\mathbf{M}$ of rank r is an $m \times n$ matrix $\mathbf{M}^S$ such that

$$\mathbf{M}\mathbf{M}^S\mathbf{M} = \mathbf{M}$$

Show that the Moore-Penrose generalized inverse is a generalized inverse.

(b) Show that $(\mathbf{M}^{+})^{+} = \mathbf{M}$.

(c) Let $\mathbf{y}$ be an $m \times 1$ vector and let

$$\mathbf{y} = \mathbf{Ax} + \eta$$

where $\mathbf{x}$ is $n \times 1$, η a zero mean Gaussian random vector and $\mathbf{A}$, $m \times n$. Show that the MMSE estimate of $\mathbf{x}$ is

$$\hat{\mathbf{x}} = \mathbf{By} + (\bar{\mathbf{x}} - \mathbf{B}\mathbf{A}\bar{\mathbf{x}})$$

where $\bar{\mathbf{x}}$ is $E[\mathbf{x}]$ and

$$\mathbf{B} = \mathbf{S}\mathbf{A}^T(\mathbf{A}\mathbf{S}\mathbf{A}^T + \mathbf{N})^{+}$$

with

$$\mathbf{S} = E[(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^T]$$
$$\mathbf{N} = E[\eta\eta^T]$$

(d) If $\mathbf{N} = 0$ show that

$$\mathbf{B} = \sqrt{\mathbf{S}}\,(\mathbf{A}\sqrt{\mathbf{S}})^{+}$$

4.16. Prove the Wold decomposition theorem.

4.17. (Kailath [7]) A reproducing kernel Hilbert space (RKHS) $H(R, I)$ associated with a covariance function $R(\cdot,\cdot)$ on $I \times I$ is a Hilbert space of functions on I with an inner product $<\cdot,\cdot>_H$, which has the property that for all $m(t) \in H(R,I)$, $t \in I$,

$$<m(s), R(s, t)>_H = m(t)$$

Clearly, this assumes $R(s,t) \in H(R, I)$

(a) Let $\|m\|_H$ be a norm on H. Show that for $m, n \in H(R,I)$

$$4 <m, n>_H = \|m + n\|_H^2 - \|m - n\|_H^2$$

(b) Let $\{m_n\}$ be a Cauchy sequence in $H(R, I)$. Show that

$$|m_n(t) - m_m(t)| \to 0$$

CHAPTER 5

PROPAGATION EQUATIONS

The concept of state was introduced in Chapter 2 as a quantity that provided the analyst with a quantitative means of describing the temporal evolution of some systems. For deterministic systems the state at each instant of time was deterministic and could be evaluated and written down. Once the system became perturbed by some random force, such a clear description of its state became quite nebulous. To say that the state vector has a given value at some future time, given only knowledge of the present, would be merely an educated guess and could not be predicted with the certitude of deterministic dynamics. It is therefore necessary to consider different quantities to analyze stochastic systems. These quantities are most usefully expressed in terms of probability density functions. These density functions depict in a deterministic fashion how the state of the stochastic system progresses with time. Thus, to a great degree they represent the state of stochastic systems.

A second reason for wanting to study the probability densities of dynamic systems is that these densities are the basis of the estimation results developed in Chapter 4. There we found that for a state system given by

$$d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}(t), t)\, dt + d\mathbf{n}(t)$$

and a measurement system given by

$$d\mathbf{y}(t) = \mathbf{h}(\mathbf{x}(t), t)\, dt + d\mathbf{w}(t)$$

the MMSE estimate of $\mathbf{x}(t)$, given measurements from t_0 to t, was

$$\hat{\mathbf{x}}(t) = E[\mathbf{x}(t) \,|\, \mathcal{Y}_{t_0 t}]$$

where $\mathcal{Y}_{t_0 t}$ was the minimum σ-field generated by $\mathbf{y}(s)$, $s \in [t_0, t]$. This expectation can be obtained if we know $p_{\mathbf{x}}(\mathbf{u}, t \,|\, \mathcal{Y}_{t_0 t})$. Thus, if we know this conditional density, we know $\hat{\mathbf{x}}(t)$. Furthermore, we can also obtain the performance of the estimator, namely, the matrix

$$\mathbf{P}(t) = E[(\mathbf{x}(t) - \hat{\mathbf{x}}(t))((\mathbf{x}(t) - \hat{\mathbf{x}}(t))^T]$$

Thus, the main objective of this chapter is to learn how the conditional and unconditional probability densities of the process $\mathbf{x}(t)$ evolve with time. To do this we first review the stochastic system and measurement model discussing its structure and its relevance to realistic representations of actual systems. The next section considers only the system model and evaluates the evolution of the density $p_{\mathbf{x}}(\mathbf{u}, t \mid \mathbf{x}(s))$, called the *transition density*. This transition probability represents the state of the system $\mathbf{x}(t)$ and is used later in the evolutions of other conditional densities.

In Section 5.3 we develop the most important equations in this book. They are the propagation equations for $\mathbf{x}(t)$ given, the measurements observed—for example, $p_{\mathbf{x}}(\mathbf{u}, t \mid O_{t_0,t})$. We consider two classes of measurements:

1. Gaussian additive measurements: Here the measurement equation is

$$d\mathbf{y}(t) = \mathbf{h}(\mathbf{x}(t), t)\, dt + d\mathbf{w}(t)$$

where $\mathbf{w}(t)$ is a Wiener process, thus Gaussian. The equation for $p_{\mathbf{x}}(\mathbf{u}, t \mid \phi_{t_0,t})$ is the *Kushner-Stratonovich equation* (KSE).

2. Poisson step measurements: The measurements are a Poisson step process $\mathbf{N}(t)$, where $\mathbf{N}(t)$ has only unit jumps with rate parameter $\boldsymbol{\lambda}(\mathbf{x}(t), t)$. The equations for $p_{\mathbf{x}}(\mathbf{u}, t \mid \phi_{t_0,t})$ in this case are called *Snyder's equations* (SE). We discuss these equations in detail, presenting two examples of the propagation of the density.

The last section, 5.4, discusses Bucy's representation theorem. This is an integral approach to the evaluation of the conditional densities.

5.1 THE MODEL

In order to develop nonlinear estimation theory it is first necessary to define the model that will be used in the analysis. The model must have two fundamental properties. It first must provide a means for the analysis by possessing a suitable structure. Second and most important, it represents a mathematical description of a physical problem. The first requirement of our model, that of analytical tractability, will be met by a wide class of Markov-process descriptions of dynamical systems. The second requirement of being a suitable embodiment of a realistic physical process can also be met by choosing Markov processes as the building blocks. Thus, our primary aim in this section is to develop models that are Markovian and describe their properties sufficiently well so that later analysis can be performed directly.

The model is usually divided into two parts, the system and the measurement. The system equation is a suitably chosen state variable expression that represents the behavior of the quantities sought. In general, the state vector will itself be a random quantity. From Chapter 2 we found that a general

description for the propagation of a state vector in a deterministic environment can be represented by the system equation

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}, t) \tag{1.1}$$

with $\mathbf{x}(t_0) = \mathbf{x}_0$. There is clearly no randomness associated with this equation, and if $\mathbf{f}(\mathbf{x}, t)$ satisfies the Lipschitz conditions, we know from Appendix A that there exists a unique solution to this vector equation. Thus, the state of this system is known for all time. This is the basis of the Lagrangian description of the universe, wherein with the proper equations the course of all mankind could be perfectly predicted. Unfortunately, nature is not that generous, and descriptions like (1.1) are valid only for a small class of realistic problems. What actually does occur is that (1.1) is driven by a random forcing function that makes an exact determination of $\mathbf{x}(t)$ impossible. Thus,

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}, t) + \mathbf{u}(t) \tag{1.2}$$

represents the true state of affairs, where $\mathbf{u}(t)$ is a stochastic process; then the state $\mathbf{x}(t)$ is also a random process devoid of the determinism initially proposed. Now (1.2) represents the mathematical description of some physical system, and it may represent it quite well, thus satisfying the second requirement of the model development. The difficulty arises when one attempts to perform analysis on (1.2) for $\mathbf{u}(t)$ being an arbitrary random process. Thus, in order to satisfy the requirement of analytical tractability, we must specialize the form of the additive random disturbance. As was stated at the outset, what is desired is that $\mathbf{x}(t)$ be a Markov process, because Markov processes are most analytically tractable. In order to insure this, the system model is given in the following form:

$$d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}, t)\, dt + d\mathbf{n}(t) \tag{1.3}$$

where now $\mathbf{n}(t)$ is an independent increment process. This insures that $\mathbf{x}(t)$ is Markov. This can be seen as follows: Since $\mathbf{n}(t)$ is an independent increment process the value of $\mathbf{n}(t)$ over the interval (t_i, t_{i+1}) is independent of the process over any nonoverlapping interval. Furthermore, if we are asked for the statistics of the process at time t_{i+1} and are given the process value at times $t_i, t_{i-1}, \cdots, t_{i-k}$, the t's being ordered with respect to the subscripts, then clearly it depends solely on the most recent state and the forcing function over the interval (see Problem 5.1).

The choice of the $(n \times 1)$-vector independent increment process is totally arbitrary, although two are most frequently used. Specifically the most common choices are the Wiener process and the generalized Poisson process. Thus, for the sake of generality, we shall assume that

$$d\mathbf{n}(t) = d\mathbf{n}_p(t) + d\mathbf{n}_g(t) \tag{1.4}$$

where the $n \times 1$ noise process is the sum of $\mathbf{n}_p(t)$, an $n \times 1$ generalzed Poisson process and $\mathbf{n}_g(t)$ an $n \times 1$ Wiener process.

The system equation can be generalized to read

$$d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}, t)dt + \mathbf{g}(\mathbf{x}, t)\, d\mathbf{n}(t) \tag{1.5}$$

where $\mathbf{f}(\mathbf{x}, t)$ is an $n \times 1$ vector; $\mathbf{g}(\mathbf{x}, t)$, an $n \times q$ matrix; and $\mathbf{n}(t)$, a $q \times 1$ independent increment process. This representation is also a Markov process. To avoid the problems of stochastic controllability, $\mathbf{g}(\mathbf{x}, t)$ is usually chosen to be $n \times n$ and is also Holder-continuous in t, Lipschitz-continuous in $\mathbf{x}$, and globally bounded. Furthermore, the matrix $\mathbf{g}^T(\mathbf{x}, t)\mathbf{g}(\mathbf{x}, t)$ must have similar properties (see Horowitz and Section 5.2). For our purposes, descriptions of the form in (1.3) will be sufficient and results concerning (1.5) will be left to the problems. We will also further clarify the restrictions in the appropriate theorems.

Another reason for using a Wiener process and a generalized Poisson process can be obtained if we recall from Chapter 3 that both $d\mathbf{n}_g(t)/dt$ and $d\mathbf{n}_p(t)/dt$ were stationary white noise processes that excited all modes uniformly. $d\mathbf{n}_g(t)$ accounts for continuous fluctuations in the state, while $d\mathbf{n}_p(t)$ accounts for the discontinuities.

The statistics of the two noise processes follow directly. We shall first assume that both $d\mathbf{n}_g(t)$ and $d\mathbf{n}_p(t)$ are $n \times 1$ vectors of zero mean:

$$E[d\mathbf{n}_g(t)] = E[d\mathbf{n}_p(t)] = 0 \tag{1.6}$$

The Wiener process has a positive definite $n \times n$ covariance matrix $\mathbf{Q}(t)$ given by

$$E[d\mathbf{n}_g(t)d\mathbf{n}_g^T(t)] = \mathbf{Q}(t)\, dt \tag{1.7}$$

Clearly, if $d\mathbf{n}_g(t)/dt$ can be formally written, we would obtain by first defining

$$\mathbf{w}(t) = \frac{d\mathbf{n}_g(t)}{dt} \tag{1.8}$$

the covariance matrix of the Gaussian white noise

$$E[\mathbf{w}(t)\mathbf{w}^T(s)] = \mathbf{Q}(t)\, \delta(t - s) \tag{1.9}$$

where $\delta(t - s)$ is the scalar delta function.

Similarly, for Poisson noise we can evaluate a covariance matrix. To make the Poisson process as general as possible, it is assumed that each component of $d\mathbf{n}_p(t)$, $dn_{p_i}(t)$, is governed by rate parameter $\lambda_i(t)$ and has amplitude probability-density function $p_{a_i}(\alpha_i)$. To evaluate the covariance matrix, we first consider the covariance of two single components $dn_{p_i}(t)$ and $dn_{p_j}(t)$, assuming that a_i and a_j are zero mean. Now, for $i \neq j$ (suppressing the p subscript)

$$
\begin{aligned}
E[dn_i(t)\, dn_j(t)] &= 0\ P[\text{no change in } n_i(t) \text{ in } t, t + dt] \\
&\quad P[\text{no change in } n_j(t) \text{ in } t, t + dt] \\
&\quad + 2E[a_i]0\ P[\text{one change } n_i(t) \text{ in } t, t + dt] \\
&\quad P[\text{no change } n_j(t) \text{ in } t, t + dt] \\
&\quad + E[a_i a_j]\, P[\text{one change } n_i(t) \text{ in } t, t + dt] \\
&\quad P[\text{one change } n_j(t) \text{ in } t, t + dt] \\
&= E[a_i a_j]\lambda_i(t)\, dt \lambda_j(t)\, dt + o(dt) \qquad (1.10)
\end{aligned}
$$

where $o(dt)$ is the sum of terms of order dt. Yet clearly for $i \neq j$ the above covariance term is itself $o(dt)$. Yet for $i = j$ we easily have

$$E[dn_i(t)\, dn_j(t)] = E[a_i^2]\lambda_j(t)\, dt + o(dt) \qquad (1.11)$$

Thus, we obtain

$$E\left[d\mathbf{n}_p(t)\, d\mathbf{n}_p^T(t)\right] = \begin{bmatrix} \sigma_{a_1}^2 \lambda_1(t) & 0 \\ 0 & \sigma_{a_n}^2 \lambda_n(t) \end{bmatrix} dt \qquad (1.12)$$

where $\sigma_{a_i}^2$ is the variance of a_i (assuming zero mean) and $\lambda_i(t)$ is the arrival rate of the process $dn_{p_i}(t)$.

The state variable model is a natural model of systems in many instances. The following three examples consider specific cases. The first considers the linear time-invariant state variable realization of a process with a desired power spectral density. The second case considers a simple electrical circuit where the resistor is modeled as a generator of thermal noise. The third case demonstrates how the state variable formulation can be used for a problem that at the outset appears to have no connection at all with dynamical systems.

Example. Let $\mathbf{x}(t)$ be given by the solution of the following differential equation:

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{u}(t) \qquad (1.13)$$

where $\mathbf{A}$ is an $n \times n$ constant matrix and $\mathbf{u}(t)$ is an $n \times 1$ zero mean white noise process with

$$E[\mathbf{u}(t)\mathbf{u}^T(s)] = \mathbf{Q}\delta(t - s) \qquad (1.14)$$

The correlation functions of the process, assuming zero mean (see Problem 5.2), is defined as

$$\mathbf{K}(t, s) = E[\mathbf{x}(t)\mathbf{x}^T(s)] \qquad (1.15)$$

Recall from Chapter 3 that if $s > t$, then

$$\mathbf{x}(s) = \boldsymbol{\Phi}(s, t)\mathbf{x}(t) + \int_t^s \boldsymbol{\Phi}(s, \zeta)\mathbf{u}(\zeta)\, d\zeta \qquad (1.16)$$

where $\Phi(s, t)$ is the transition matrix given by

$$\Phi(s, t) = \exp(\mathbf{A}(s - t)) \tag{1.17}$$

Now, since $\mathbf{u}(\zeta)$ is a white noise process, it is easily seen that

$$E\left[\mathbf{x}(t)\int_t^s \mathbf{u}^T(\zeta)\Phi^T(s, \zeta)\, d\zeta\right] = 0 \tag{1.18}$$

Thus.

$$\mathbf{K}(t, s) = E[\mathbf{x}(t)\mathbf{x}^T(t)]\Phi^T(s, t) \tag{1.19}$$

Likewise for the case where $s < t$

$$\mathbf{K}(t, s) = \Phi(t, s)E[\mathbf{x}(s)\mathbf{x}^T(s)] \tag{1.20}$$

Let $\mathbf{P}(t)$ be the covariance matrix

$$\mathbf{P}(t) = E[\mathbf{x}(t)\mathbf{x}^T(t)] \tag{1.21}$$

Differentiating $\mathbf{P}(t)$, we obtain

$$\dot{\mathbf{P}}(t) = E[\dot{\mathbf{x}}(t)\mathbf{x}^T(t)] + E[\mathbf{x}(t)\dot{\mathbf{x}}^T(t)] \tag{1.22}$$

Now, using (1.13) in the above, we readily obtain

$$\dot{\mathbf{P}}(t) = \mathbf{A}\mathbf{P}(t) + \mathbf{P}(t)\mathbf{A}^T + E[\mathbf{u}(t)\mathbf{x}^T(t)] + E[\mathbf{x}(t)\mathbf{u}^T(t)] \tag{1.23}$$

Using the transition matrix and the covariance of $\mathbf{u}(t)$, it can be shown that (1.23) reduces to

$$\dot{\mathbf{P}}(t) = \mathbf{A}\mathbf{P}(t) + \mathbf{P}(t)\mathbf{A}^T + \mathbf{Q} \tag{1.24}$$

This is called the degenerate *Riccati equation*, and for our purposes the steady-state solution is required, since t_0 is assumed to be $-\infty$ and the system in (1.13) is u.a.s.i.l. Thus, let $\mathbf{P}$ be the solution to

$$\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A} + \mathbf{Q} = \mathbf{0} \tag{1.25}$$

Then

$$\mathbf{K}(t, s) = \begin{cases} \exp(-\mathbf{A}\tau)\mathbf{P}; & t > s \\ \mathbf{P}[\exp(-\mathbf{A}\tau)]^T; & t < s \end{cases} \tag{1.26}$$

where $\tau = |t - s|$. The spectral matrix of the process is $\mathbf{S}(f)$, which is the Fourier transform of (1.26). For stationary processes, $\mathbf{S}(f)$ is usually given, and thus, there exists a state variable realization that yields that characterization. The process of going from $\mathbf{S}(f)$ to the state variable realization is called *spectral factorization* and is discussed by Davis. The usefulness of state variable realizations for describing processes of given spectra has been shown to be quite extensive, especially in the field of communications (see Van Trees [1,3]). Problems 5.3 and 5.4 discuss these properties in more detail.

Example. Consider the $R - C$ series circuit. The resistor generates a white noise current $i_n(t)$ where

$$E[i_n(t)] = 0 \tag{1.27}$$

$$E[i_n(t)i_n(t+\tau)] = \frac{4kT}{R}\,\delta(\tau) \tag{1.28}$$

where k is Boltzman's constant, T is the temperature in degrees Kelvin, and R is the resistance in ohms. The equation governing the circuit voltage $e(t)$ is

$$C\dot{e}(t) + \frac{1}{R}\,e(t) = i_n(t) \tag{1.29}$$

Letting

$$x(t) = \frac{1}{C}\,e(t) \tag{1.30}$$

and

$$u(t) = \frac{i_n(t)}{C} \tag{1.31}$$

we have in state variable form

$$\dot{x}(t) = -ax(t) + u(t) \tag{1.32}$$

where a is $1/RC$. This example is typical of many problems where the driving function is a white noise process. What is interesting in this example is that the common interpretation of $i_n(t)$ is that of the derivative of a generalized Poisson process and not the derivative of a Wiener process.

Example. A common problem in estimation theory is that of estimating a random variable and not a time-varying process. To represent this problem in state variable form follows quite simply. Let $\mathbf{x}_0$ be the $n \times 1$ random variable. Then let

$$\dot{\mathbf{x}}(t) = \mathbf{0}; \qquad \mathbf{x}(t_0) = \mathbf{x}_0 \tag{1.33}$$

Clearly, $\mathbf{x}(t) = \mathbf{x}_0$ for all t. Gaussian white noise could also be present by letting $\mathbf{Q}$ be identically $\mathbf{0}$.

Thus, what we conclude about the model is that it represents a vehicle through which the statistical dynamics of a wide class of processes can be represented.

The second part of a model is a formulation for the measurement. In this chapter we shall consider two formulations. The first is the now classical form of an arbitrary function of the state in additive white Gaussian noise. The second form of measurement is that of a counting process whose rate parameter depends upon the state vector. The former representation follows from classical measurements where the measurements were usually embedded in noise. The latter case is a result of considering a wider class of measurement processes where such things as quantum effects must be considered. The Poisson model was first proposed by Snyder for biomedical applications

of radioactive tracer interpretation. It has been used by Clark in the field of optical communications, by McGarty [2] in the field of upper atmospheric research, and by Evans [2] in the area of extra-low-frequency communications.

The additive Gaussian model is given by

$$d\mathbf{y}(t) = \mathbf{h}(\mathbf{x}(t), t)\, dt + d\mathbf{w}(t) \tag{1.34}$$

where $d\mathbf{w}(t)$ is an $m \times 1$ Wiener process with

$$E[d\mathbf{w}(t)\, d\mathbf{w}^T(t)] = \mathbf{R}(t)\, dt \tag{1.35}$$

where $\mathbf{R}(t)$ is an $m \times m$ matrix and $\mathbf{h}(x(t), t)$ is an $m \times 1$ nonlinear function of the state.

The linearized white Gaussian noise model is given by

$$\mathbf{z}(t) = \mathbf{C}(t)\mathbf{x}(t) + \mathbf{v}(t) \tag{1.36}$$

where

$$\frac{d\mathbf{y}(t)}{dt} = \mathbf{z}(t) \tag{1.37}$$

and

$$\frac{d\mathbf{w}(t)}{dt} = \mathbf{v}(t) \tag{1.38}$$

Again we must note that (1.36) is a formal expression of (1.34), since (1.38) does not exist mathematically.

Examples of measurement processes modeled by equations (1.34) or (1.36) can be found in many areas. For example, in communication systems where phase modulation is employed, the received signal is given by

$$z(t) = \sin(2\pi f_0 t + \mathbf{C}^T\mathbf{x}(t)) + v(t) \tag{1.39}$$

where $\mathbf{C}$ is an $m \times 1$ vector; $\mathbf{x}(t)$, $n \times 1$ process with a suitable state variable description; and $v(t)$, additive white noise. The frequency f_0 is the carrier frequency. This formulation has been used by Snyder [1] in the analysis of both phase and frequency modulation schemes.

The second measurement is described by a Poisson counting process $dN(t)$, where $N(t)$ has the following properties:

$$N(0) = 0 \tag{1.40}$$

$$P[N(t) = k \mid \lambda(x(s), s);\ s \in [0, t]] = \frac{\left[\int_0^t \lambda(x(\zeta), \zeta)\, d\zeta\right]^k}{k!} \exp\left(-\int_0^t \lambda(x(\zeta), \zeta)\, d\zeta\right) \tag{1.41}$$

and $dN(t)$ is an independent increment process. The conditioning in (1.41) is

as $m, n \to \infty$ for all $t \in I$. Note that for RKHS norm convergence implies pointwise convergence.

(c) Show that if $R(t, s)$ is continuous for all $(t, s) \in I \times I$, then the functions in $H(R, I)$ are continuous on I. Further, show that if

$$\frac{\partial^{2m}}{\partial t^m \partial s^m} R(t, s)$$

exist then functions in $H(R, I)$ are n times differential.

(d) If an RKHS has two kernels R_1 and R_2, show that $R_1 = R_2$.

4.18. Consider the discrete-time filter developed in Section 4.3. Show that it can be written as

$$\begin{aligned}
\hat{\mathbf{x}}(k+1) &= \mathbf{P}(k+1)\mathbf{M}^{-1}(k+1)\hat{\mathbf{x}}(k+1) \\
&\quad + \mathbf{P}(k+1)\mathbf{C}^T(k+1)\mathbf{R}^{-1}(k+1)\,\mathbf{z}(k+1) \\
\mathbf{M}(k+1) &= \boldsymbol{\Phi}(k+1, k)\,\mathbf{P}(k)\boldsymbol{\Phi}^T(k+1, k) + \mathbf{Q}(k) \\
\mathbf{P}(k+1) &= [\mathbf{M}^{-1}(k+1) + \mathbf{C}^T(k+1)\mathbf{R}^{-1}(k+1)\mathbf{C}(k+1)]^{-1}
\end{aligned}$$

where the gain matrix $\mathbf{K}(k+1)$ has been eliminated.

Hint. Use the matrix identities

$$\begin{aligned}
(\mathbf{A}^{-1} + \mathbf{B}^T\mathbf{C}^{-1}\mathbf{B})^{-1} &= \mathbf{A} - \mathbf{A}\mathbf{B}^T(\mathbf{B}\mathbf{A}\mathbf{B}^T + \mathbf{C})^{-1}\mathbf{B}\mathbf{A} \\
(\mathbf{A}^{-1} + \mathbf{B}^T\mathbf{C}^{-1}\mathbf{B})^{-1}\mathbf{B}^T\mathbf{C}^{-1} &= \mathbf{A}\mathbf{B}^T(\mathbf{B}\mathbf{A}\mathbf{B}^T + \mathbf{C})^{-1}
\end{aligned}$$

necessary, since $P[N(t) = k]$ unconditioned can only be determined by averaging over the random variable $x(s)$ (see Problem 5.4).

In both measurements the role of an independent increment process is clearly evident. It ensures us that the increments of knowledge concerning the process given to us by a measurement and conditioned on the process at a given time are independent of past measurements. This will become a central fact when dealing with the issue of evaluating the propagation of the conditional density of the process, given the measurements.

A common representation of these processes is by the block diagrams shown in Figures 5.1 and 5.2 for the additive Gaussian noise measurement and the Poisson measurement respectively.

Figure 5.1 Additive Gaussian noise measurement model.

Figure 5.2 Poisson measurement model.

What we have done in this section is to justify the use of the system model and the measurement model as adequate representations of physical systems. We shall further extend this development in the next section when we discuss the Fokker-Planck equation in terms of a more generalized development.

5.2 SYSTEM PROPAGATION EQUATIONS

The state equation for the system represents the dynamics of a random process that propagates as a function of time. Since it is a random process, an exact determination of the state is impossible. Yet an adequate representation of the system can be given as one knows the transition probability den-

sity function for the process. This follows upon recalling that for a Markov process the transition density acts as the transition matrix for linear dynamical systems.

To develop the machinery to obtain the transition density of the process $p_{\mathbf{x}}(\mathbf{u}, t \mid \mathbf{x}(s) = \mathbf{v})$, it is first necessary to develop a more exact definition of the system and to define clearly the conditions that such a model must meet.

DEFINITION 2.1. Let K be a set in an m dimensional euclidean space. Let $\mathbf{k}(\mathbf{x}(t),t)$ be an $m \times 1$ vector such that $\mathbf{k}(\mathbf{x}(t),t) \in K$. Let $\mathbf{x}(t)$ be an $n \times 1$ vector. Let $\| \ \|$ represent a suitable metric for both $\mathbf{k}(\mathbf{x}(t), t)$ and $\mathbf{x}(t)$. Then $\mathbf{k}(\mathbf{x}(t),t)$ is Holder-continuous on k if for some constants $C, \lambda > 0$

$$\|\mathbf{k}(\mathbf{x}_1, t) - \mathbf{k}(\mathbf{x}_2, t)\| < C\|\mathbf{x}_1 - \mathbf{x}_2\|^{\lambda} \tag{2.1}$$

Clearly, Holder continuity is a generalization to the Lipschitz conditions.

DEFINITION 2.2. A state equation written as a stochastic differential equation

$$d\mathbf{x} = + \mathbf{f}(\mathbf{x}, t)\, dt + \mathbf{g}(\mathbf{x}, t)\, d\mathbf{n}(t) \tag{2.2}$$

where $t \in T = [t_0, t_1]$, and

$$d\mathbf{n}(t) = d\mathbf{n}_p(t) + d\mathbf{n}_g(t) \tag{2.3}$$

where $\mathbf{n}_g(t)$ is a Wiener process of dimension m and $\mathbf{n}_p(t)$ is a generalized Poisson process of dimension m and (1) $\mathbf{x}(t_0) = \mathbf{x}_0$ where $\mathbf{x}_0$ is independent of $d\mathbf{n}(t)$; (2) the $n \times m$ matrix $\mathbf{g}(\mathbf{x}, t)$ is Holder-continuous in t and Lipschitz-continuous in $\mathbf{x}$. Also the matrix

$$\mathbf{G} = \mathbf{g}(\mathbf{x}, t)\mathbf{g}^T(\mathbf{x}, t) \tag{2.4}$$

is strictly positive definite and the terms

$$\frac{\partial G_{ij}}{\partial x_i}\ ;\ \frac{\partial^2 G_{ij}}{\partial x_i \partial x_j} \qquad (i, j = 1,2,\cdots,n)$$

are globally Lipschitz-continuous in x, continuous in t, and globally bounded; (3) the vector $\mathbf{f}(\mathbf{x}, t)$ is continuous in t and globally Lipschitz-continuous in $\mathbf{x}$ and $\partial f_i / \partial x_i$ are globally Lipschitz continuous in x and continuous in t; then (2.2) is called the *standard stochastic state realization* (SSSR).

Condition (1) insures that the process is Markov, and conditions (2) and (3) will be necessary to insure the existence of a solution to the conditional density equation. These conditions as applied to the solution of the Fokker-Planck equation have been discussed by Elliot.

The SSSR model provides one with a representation suitably broad in nature that can be used to model a Markov process. To fully describe a Markov process, we recall from Chapter 3 that it is sufficient to obtain the transition density of the process $\mathbf{x}(t)$, given $\mathbf{x}(s)$ for some $s < t$. With this and the Chapmann-Kolmogrov equation and the density of $\mathbf{x}$ at some arbitrary time $t_0 < s < t$, a complete statistical representation of the process

is possible. We also recall that the characteristic function, which is the Fourier transform of the probability density function, would also be sufficient to describe the system. Thus:

DEFINITION 2.3. Let $p_{\mathbf{x}}(\mathbf{u}, t\,|\,\mathbf{x}(s) = \mathbf{v})$ be the conditional probability-density function of the random process $\mathbf{x}$ at time t, given that $\mathbf{x}$ at time s is equal to $\mathbf{v}$. The characteristic function is defined by $M_{\mathbf{x}}(\mathbf{u}, t\,|\,\mathbf{x}(s))$ where

$$M_{\mathbf{x}}(\mathbf{u}, t\,|\,\mathbf{x}(s)) = E[\exp(j\mathbf{u}^T\mathbf{x}(t))|\mathbf{x}(s) = \mathbf{v}] = \int_{R^n} \exp(j\mathbf{u}^T\zeta) p_{\mathbf{x}}(\zeta, t\,|\,\mathbf{x}(s) = \mathbf{v})\, d\zeta \tag{2.5}$$

where $\mathbf{u}$ is an $n \times 1$ vector, ζ is an $n \times 1$ vector, and the integration is over all $\mathbf{R}_n$.

If we obtain an equation for the temporal evaluation of $M_{\mathbf{x}}(\mathbf{u}, t\,|\,\mathbf{x}(s))$, then equivalently we have obtained the temporal evolution of $p_{\mathbf{x}}(\mathbf{u}, t\,|\,\mathbf{x}(s))$.

The solution to this problem was first presented by Moyal (pp. 195–202), who credited it to Bartlett and is given in the following theorem.

THEOREM 2.1

(Bartlett-Moyal) Let $M_{\mathbf{x}}(\mathbf{u},t\,|\,\mathbf{x}(s))$ be the characteristic function of the Markov process $\mathbf{x}(t)$, $t \in T$, where T is some interval. Assume the following:

1. $M_{\mathbf{x}}(\mathbf{u}, t\,|\,\mathbf{x}(s))$ is continuously differentiable in t, $t \in T$.
2.

$$\frac{1}{\Delta t} \left| E[(\exp\{j\mathbf{u}^T[\mathbf{x}(t + \Delta t) - \mathbf{x}(t)]\} - 1)|\mathbf{x}(t)] \right| \leqq g(\mathbf{u}; t, \mathbf{x}) \tag{2.6}$$

where $E[|g|]$ is bounded on T.

3.

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} E[(\exp\{j\mathbf{u}^T[\mathbf{x}(t + \Delta t) - \mathbf{x}(t)]\} - 1)|\mathbf{x}(t)] = \phi(\mathbf{u}, t, \mathbf{x}(t)) \tag{2.7}$$

Then

$$\frac{\partial M_{\mathbf{x}}(\mathbf{u}, t\,|\,\mathbf{x}(s))}{\partial t} = E[\exp[j\mathbf{u}^T\mathbf{x}(t)]\, \phi(\mathbf{u}, t, \mathbf{x}(t))|\mathbf{x}(s)] \tag{2.8}$$

where the expectation in (2.8) is over $\mathbf{x}(t)$

Proof. The proof to this theorem is quite straightforward and follows directly from the definition of the derivative of the conditional density function. Recall that by definition

$$\frac{\partial M_{\mathbf{x}}(\mathbf{u}, t\,|\,\mathbf{x}(s))}{\partial t} = \lim_{\Delta t \to 0} \frac{1}{\Delta t} [M_{\mathbf{x}}(\mathbf{u}, t + \Delta t\,|\,\mathbf{x}(s)) - M_{\mathbf{x}}(\mathbf{u}, t\,|\,\mathbf{x}(s))] \tag{2.9}$$

But also by definition

$$M_{\mathbf{x}}(\mathbf{u}, t + \Delta t \,|\, \mathbf{x}(s)) = E[\exp [j\mathbf{u}^T \mathbf{x}(t + \Delta t)] \,|\, \mathbf{x}(s)] \tag{2.10}$$

From the definition of the conditional characteristic function, we have

$$M_{\mathbf{x}}(\mathbf{u}, t + \Delta t \,|\, \mathbf{x}(s)) = \int p_{\mathbf{x}}(\mathbf{v}, t + \Delta t \,|\, \mathbf{x}(s)) \exp (j\mathbf{u}^T \mathbf{v})\, d\mathbf{v} \tag{2.11}$$

But recall that since $\mathbf{x}(t)$ is a Markov process, we can use the Chapman-Kolmogorov equation to yield

$$p_{\mathbf{x}}(\mathbf{v}, t + \Delta t \,|\, \mathbf{x}(s)) = \int p_{\mathbf{x}}(\mathbf{v}, t + \Delta t \,|\, \mathbf{r}, t) p_{\mathbf{x}}(\mathbf{r}, t \,|\, \mathbf{x}(s))\, d\mathbf{r} \tag{2.12}$$

Using this in (2.11) yields

$$\begin{aligned} M_{\mathbf{x}}(\mathbf{u}, t + \Delta t \,|\, \mathbf{x}(s)) &= \iint p_{\mathbf{x}}(\mathbf{v}, t + \Delta t \,|\, \mathbf{r}, t) p_{\mathbf{x}}(\mathbf{r}, t \,|\, \mathbf{x}(s)) \exp (j\mathbf{u}^T \mathbf{v})\, d\mathbf{v}\, d\mathbf{r} \\ &= \int p_{\mathbf{x}}(\mathbf{r}, t \,|\, \mathbf{x}(s)) \exp (j\mathbf{u}^T \mathbf{r}) \\ &\quad \left[\int p_{\mathbf{x}}(\mathbf{v}, t + \Delta t \,|\, \mathbf{r}) \exp (j\mathbf{u}^T \mathbf{v} - j\mathbf{u}^T \mathbf{r})\, d\mathbf{v}\, d\mathbf{r}\right] \end{aligned} \tag{2.13}$$

which by definition

$$\begin{aligned} M_{\mathbf{x}}(\mathbf{u}, t + \Delta t \,|\, \mathbf{x}(s)) = E[&\exp (j\mathbf{u}^T \mathbf{x}(t)) \\ &E[\exp (j\mathbf{u}^T (\mathbf{x}(t + \Delta t) - \mathbf{x}(t)) \,|\, \mathbf{x}(t)] \,|\, \mathbf{x}(s)] \end{aligned} \tag{2.14}$$

We can now use this notation to write

$$\begin{aligned} &\frac{M_{\mathbf{x}}(\mathbf{u}, t + \Delta t \,|\, \mathbf{x}(s)) - M_{\mathbf{x}}(\mathbf{u}, t \,|\, \mathbf{x}(s))}{\Delta t} \\ &= E[\exp (j\mathbf{u}^T \mathbf{x}(t)) \frac{1}{\Delta t} \\ &\quad E[(\exp (j\mathbf{u}^T (\mathbf{x}(t + \Delta t) - \mathbf{x}(t))) - 1) \,|\, \mathbf{x}(t)] \,|\, \mathbf{x}(s)] \end{aligned} \tag{2.15}$$

Taking the limit as $\Delta t \to 0$ and using (2.7) and (2.9) yields (2.8) ∎

The first important fact to note about this theorem is that the assumption that $\mathbf{x}(t)$ was a Markov process was essential to the derivation. This was employed in (2.13) and allowed us to write (2.14) in the factorable form. The function $\phi(\mathbf{u}, t, \mathbf{x}(t))$ is also called the *Ito differential of the Markov process* (see Frost, p. 36) and is also termed the *infinitesimal generator* of the Markov semigroup (see Dynkin [1 Chapter 2] or Wong [2, Chapter 5]. In many ways $\phi(\mathbf{u}, t, \mathbf{x}(t))$ plays the role of the transition matrix that we developed in Chapter 2 for linear time-varying systems. This analogy can be carried quite far in defining a stochastic system, just as $\Phi(t, t_0)$ is used in defining a dynamic system. Once $\phi(\mathbf{u}, t, \mathbf{x}(t))$ is evaluated, all that is necessary to define fully the state of a stochastic system has been given.

To define fully the SSSR system, it is thus sufficient to obtain $\phi(\mathbf{u}, t, \mathbf{x}(t))$ This is done in the following Lemma.

LEMMA 2.1. Let $\mathbf{x}$ be an $(n \times 1)$-vector Markov process generated by

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\, dt + d\mathbf{n} \tag{2.16}$$

where

$$d\mathbf{n} = d\mathbf{n}_g + d\mathbf{n}_p \tag{2.17}$$

and $d\mathbf{n}_g$ is an $n \times 1$ Wiener process with covariance matrix

$$E\left[d\mathbf{n}_g(t)\, d\mathbf{n}_g^T(t)\right] = \mathbf{Q}(t)\, dt \tag{2.18}$$

and $d\mathbf{n}_p$ is an $n \times 1$ generalized Poisson process with rate vector $\boldsymbol{\lambda}(t)$ and jump probability density $p_\alpha(\alpha)$. Then

$$\psi(\mathbf{u}, t, \mathbf{x}(t)) = j\mathbf{u}^T\mathbf{f}(\mathbf{x}, t) - \tfrac{1}{2}\mathbf{u}^T\mathbf{Q}\mathbf{u} - \sum_{i=1}^{n} \lambda_i [1 - M_{\alpha_i}(u_i)] \tag{2.19}$$

where λ_i is the ith component of $\boldsymbol{\lambda}(t)$ and $M_{\alpha_i}(u_i)$ is the characteristic function of the ith jump.

Proof. Recall that

$$\psi(\mathbf{u}, t, \mathbf{x}(t)) = \frac{E\left[e^{j\mathbf{u}^T d\mathbf{x}} - 1 \,\middle|\, \mathbf{x}(t)\right]}{dt} \tag{2.20}$$

From (2.16) we have

$$\psi(\mathbf{u}, t, \mathbf{x}(t)) = \frac{E\left[e^{j\mathbf{u}^T\mathbf{f}(\mathbf{x},t)\,dt + j\mathbf{u}^T d\mathbf{n}_g + j\mathbf{u}^T d\mathbf{n}_p} - 1 \,\middle|\, \mathbf{x}(t)\right]}{dt} \tag{2.21}$$

Clearly,

$$E\left[e^{j\mathbf{u}^T\mathbf{f}(\mathbf{x},t)\,dt + j\mathbf{u}^T d\mathbf{n}_g + j\mathbf{u}^T d\mathbf{n}_p}\right] = e^{j\mathbf{u}^T\mathbf{f}(\mathbf{x},t)\,dt} E\left[e^{j\mathbf{u}^T d\mathbf{n}_g}\right] E\left[e^{j\mathbf{u}^T d\mathbf{n}_p}\right] \tag{2.22}$$

which follows directly from the conditioning on $\mathbf{x}(t)$ and the independence of the two noise processes. From Chapter 3 we know that since $d\mathbf{n}g$ is an $n \times 1$ Gaussian process with zero mean and known covariance, we have

$$E\left[e^{j\mathbf{u}^T d\mathbf{n}_g}\right] = \exp\left(-\tfrac{1}{2}\mathbf{u}^T\mathbf{Q}\mathbf{u}\, dt\right) \tag{2.23}$$

Likewise, the characteristic function of the generalized Poisson process can also be evaluated. Let us first note that the probability of two or more jumps occurring in dt is $o(dt)$. Thus.

$$E\left[e^{j\mathbf{u}^T d\mathbf{n}_p}\right] = 1\,P[\text{no jumps}] + \sum_{i=1}^{n} E\left[e^{ju_i\alpha_i}\right] P[\text{only one jump in } dn_i] \tag{2.24}$$

But

$$P[\text{no jumps}] = \prod_{i=1}^{n} (1 - \lambda_i\, dt) = 1 - \sum_{i=1}^{n} \lambda_i\, dt + o(dt) \tag{2.25}$$

Also

$$P[\text{only one jump in } dn_{p_i}] = \lambda_i dt \prod_{j \neq i}^{n} (1 - \lambda_j dt) = \lambda_i dt + o(dt) \tag{2.26}$$

Thus

$$E\left[e^{j\mathbf{u}^T d\mathbf{n}_p}\right] = 1 - \sum_{i=1}^{n} \lambda_i \, dt \left[1 - M_{a_i}(u_i)\right] \tag{2.27}$$

where

$$M_{a_i}(u_i) = E[e^{ju_i a_i}] \tag{2.28}$$

Using (2.28) and (2.23) in (2.21) and deleting terms of $o(dt)$ proves the lemma. ∎

With this lemma we now have all the necessary tools to evaluate the equation for the temporal evolution of the transition probability density for the system defined. This is done in the following theorem.

THEOREM 2.2

Let $\mathbf{x}(t)$ be a Markov process generated by

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\, dt + d\mathbf{n}_g + d\mathbf{n}_p \tag{2.29}$$

Let $p_{\mathbf{x}}(\mathbf{u}, t \,|\, \mathbf{x}(s)) = p$ be the transition probability density function for the process $\mathbf{x}(t)$. Then p satisfies the partial differential equation

$$\frac{\partial p}{\partial t} = -\sum_{i=1}^{n} \frac{\partial (f_i p)}{\partial u_i} + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 Q_{ij} p}{\partial u_i \, \partial u_j} + \sum_{i=1}^{n} \lambda_i \left[-p + p * p_{a_i} \right] \tag{2.30}$$

where the convolution (*) is defined by

$$p * p_{a_i} = \int p_{a_i}(u_i - v_i) p_{\mathbf{x}}(u_1, \cdots v_i \cdots u_n, t \,|\, x(s))\, dv_i \tag{2.31}$$

Proof. From the previous lemma and theorem we know that

$$\frac{\partial M_{\mathbf{x}}(\mathbf{u}, t \,|\, \mathbf{x}(s))}{\partial t} = E\left[e^{i\mathbf{u}^T \mathbf{x}(t)} \left[j\mathbf{u}^T \mathbf{f}(\mathbf{x}, t) - \frac{1}{2} \mathbf{u}^T \mathbf{Q} \mathbf{u} + \sum_{i=1}^{n} \lambda_i [M_{a_i}(u_i) - 1] \right] \Big| \mathbf{x}(s) \right] \tag{2.32}$$

Now take (2.32) and inverse Fourier transform it. Clearly,

$$\frac{1}{(2\pi)^n} \int \frac{\partial M_{\mathbf{x}}(\xi, t \,|\, \mathbf{x}(s))}{\partial t} \exp\left(-j\xi^T \mathbf{u} \right) d\xi = \frac{\partial p_{\mathbf{x}}(\mathbf{u}, t \,|\, \mathbf{x}(s))}{\partial t} \tag{2.33}$$

Likewise, for the terms on the right in (2.32) we have

$$\frac{1}{(2\pi)^n} \int \exp\left(-j\xi^T \mathbf{u} \right) j\xi^T \mathbf{f}(\mathbf{v}, t) \exp\left(j\xi^T \mathbf{v} \right) p_{\mathbf{x}}(\mathbf{v}, t \,|\, \mathbf{x}(s))\, d\xi\, d\mathbf{v} = -\sum_{i=1}^{n} \frac{\partial f_i(u, t) p_{\mathbf{x}}(\mathbf{u}, t \,|\, \mathbf{x}(s))}{\partial u} \tag{2.34}$$

The other terms in (2.30) follow directly by using the properties of Fourier transforms, which in turn prove the theorem. Note that the convolution results from the product of the characteristic functions. ∎

The existence and uniqueness of the partial-differential equations of the form shown in the above theorem have been discussed in Elliot. Extensions to the case where the Poisson-process rate depends on the process $\mathbf{x}(t)$ and where the Gaussian noise process entails a process-dependent multiplicative term are discussed in Problem 5.5.

There are two important special cases of (2.30) that historically were independently developed by other techniques. The first equation is termed the Fokker-Planck equation and is used when $d\mathbf{n}_p$ is identically zero. It is discussed and derived by Uhlenbeck and Orenstein and has been utilized in the analysis of fluctuation phenomena. Moyal has an extensive discussion and examples for this case. We present this result in the following corollary.

Corollary 2.1. (Fokker-Planck equation). Let $\mathbf{x}(t)$ be an $(n \times 1)$-vector Markov process given by

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\, dt + d\mathbf{n}_g(t) \tag{2.35}$$

where

$$E\left[d\mathbf{n}_g\, d\mathbf{n}_g^T\right] = \mathbf{Q}dt \tag{2.36}$$

Let $p = p_{\mathbf{x}}(\mathbf{u}, t \,|\, \mathbf{x}(s))$ be the transition density for this process. Then p satisfies the Fokker-Planck equation

$$\frac{\partial p}{\partial t} = L^+ p \tag{2.37}$$

where L^+ is the forward Fokker-Planck operator defined by

$$L^+ = -\sum_{i=1}^{n} \frac{\partial}{\partial u_i}(f_i \cdot) + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} Q_{ij} \frac{\partial^2(\cdot)}{\partial u_i\, \partial u_j} \tag{2.38}$$

Proof. The proof is immediate by letting

$$p_{a_i}(\alpha_i) = \delta(\alpha_i): \qquad \forall i = 1, \cdots, n \tag{2.39}$$

This implies that the generalized Poisson process has no jumps and thus that a_i is everywhere zero or λ is identically zero for all time. ∎

The derivation of the Fokker-Planck equation dates back to Einstein, who in May 1905 submitted his famous paper on Brownian motion. Einstein's interest was in the one-dimensional motion of a free particle with no restoring force. The equation he obtained for the probability density of the process conditioned on knowledge of an initial position satisfied the equation:

$$\frac{\partial p}{\partial t} = \frac{D\,\partial^2 p}{\partial x^2} \tag{2.40}$$

where D is the diffusion coefficient. Einstein in the same paper went on to obtain the diffusion constant in terms of the physically measurable properties of the gas. He further noted that the solution to (2.41) was that of the diffusion equation of heat transfer, namely,

$$p_x(u, t \mid x(t_0) = 0) = \frac{1}{\sqrt{2\pi 2Dt}} \exp\left(- \frac{u^2}{2(2Dt)} \right) \tag{2.41}$$

or as one should expect, Gaussian with variance $2Dt$ and mean zero (see Einstein, pp. 1–18).

The complete derivation of (2.38) by methods differing from those above were carried out by Uhlenbeck and Orenstein (1930) and Wang and Uhlenbeck (1945). Their method of approach will be discussed in further detail when we develop the generalized Fokker-Planck equation. The historical development of Brownian motion (e.g., the Wiener process) acting on dynamical systems has been sketched by Nelson in detail, with particular emphasis on how quantum theory may be interpreted by means of suitably defined stochastic systems (see Nelson, Chapter 13–16). Similar analysis in the field of statistical mechanics is discussed by Kac.

We can now consider two examples of how the Fokker-Planck equation can be used to evaluate physical systems. The first example is used to show that for a simple physical system a complete statistical description can be obtained by means of the equation. The second example considers its use in the problem of prediction. Specifically, if we are given the value of a process $\mathbf{x}$ at time $s < t$—that is, $\mathbf{x}(s)$—how does one best predict $x(t)$?

Example. Let $x_1(t)$ represent the position of a one-dimensional particle as a function of time. Let the velocity be $x_2(t)$, where

$$x_2(t) = \frac{dx_1(t)}{dt} \tag{2.42}$$

The particle is in a viscous medium where the restoring force due to viscous drag is proportional to the velocity, the proportionality constant being a. The particle is also acted on by an external Brownian motion force $\dot{u}(t)$—$u(t)$ is a Wiener process—such that by writing a force balance, one obtains

$$\frac{dx_2(t)}{dt} = - \frac{ax_2(t)}{m}(t) + \dot{u}(t) \tag{2.43}$$

where m is the mass of the particle and $\dot{u}(t)$ is white noise of spectral height Q. The Fokker-Planck equation for the velocity becomes

$$\frac{\partial p}{\partial t} = - \frac{\partial(- (a/m)up)}{\partial u} + \frac{Q\partial^2 p}{\partial u^2} \tag{2.45}$$

We further assume that the initial position at t_0 is $x_2(t_0)$ and is deterministic. Formally this means

$$p_{x_2}(u_2, t_0) = \delta(u_2 - x_2(t_0)) \tag{2.45}$$

We can now proceed to solve (2.44). It can be shown (see Problem 5.6) for this problem that

$$p_{x_2}(u_2, t \mid x_2(t_0)) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left[-\frac{1}{2} \frac{(u_2 - \bar{x}_2)^2}{\sigma^2} \right] \tag{2.46}$$

where

$$\bar{x}^2 = x_2(t_0) \exp\left(-\frac{a}{m} t \right) \tag{2.47}$$

and

$$\sigma^2 = Q\left[1 - \exp\left(-2\frac{a}{m} t \right) \right] \tag{2.48}$$

The time dependence of p is shown in Figure 5.3. Clearly, at $t = t_0$, p should be impulsive. As t increases, both the mean and variance change with time.

Figure 5.3 Transition probability of dynamic system with white noise excitation.

the mean decaying exponentially to zero and the variance increasing exponentially to Q. For $t \gg t_0 + m/a$ we find the density of x_2 to be independent of $x_2(t_0)$ with a constant variance.

An immediate consequence of the fact that $\dot{u}(t)$ has infinite energy can be seen by a closer look at the results we obtain. The particle at t_0 has velocity $x_2(t_0)$ which is clearly finite. At $t = t_0 + \varepsilon$ for some $\varepsilon > 0$ such that $x_2(t) > Kx_2(t_0)$, where $K > 0$. This implies that there will always be a finite probability that the particle undergoes exceptionally great acceleration. It can also be shown that there is also a finite probability that $x_2(t)$ can exceed the velocity of light for physical systems. This clearly violates the laws of relativity and is a result of the unphysical as well as unmathematical nature of white noise. Thus, care should be taken in accepting the results quite literally. In general, though, these results are representative and useful for the analysis of such systems. The complete solution for the joint probability density is discussed in Problem 5.7.

Example. Let $x(t)$ be governed by the eqaution

$$\dot{x}(t) = -x(t) + \dot{u}(t) \tag{2.49}$$

where

$$E[\dot{u}(t)\dot{u}(s)] = U\delta(t - s) \tag{2.50}$$

Assume that we know that x at time s is $x(s)$. We now wish to obtain the minimum mean square estimate of x at time t, $\hat{x}(t)$, given $x(s)$. Clearly, since x is Gaussian, the conditional mean is easily obtained. From the last problem

$$p_x(u, t \mid x(s)) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left[-\frac{1}{2}\frac{(u - \bar{x})^2}{\sigma^2}\right] \tag{2.51}$$

where as before

$$\bar{x} = x(s)\exp(-\lambda) \tag{2.52}$$

and

$$\sigma^2 = U[1 - \exp(-2\lambda)] \tag{2.53}$$

Thus, $\hat{x}(t)$, given $x(s)$, is

$$\hat{x}(t) = x(s)\exp(-\lambda) \tag{2.54}$$

Furthermore, the actual minimum mean square error is given by (2.53). This problem is classically called the prediction problem. That is, we try to predict x at some future time, given x at some prior time. It shows the utility of the Fokker-Planck equation in the estimation problem as well as in the characterization problem.

A second form of the propagation equation results if we assume that the system is driven by a generalized Poisson process with no Gaussian white

noise. This leads to the set of propagation equations called the *Feller-Kolmogorov equations.*

COROLLARY 2.2 (Feller-Kolmogorov equation). Let $\mathbf{x}(t)$ be an $(n \times 1)$-vector Markov process given by

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\, dt + d\mathbf{n}_p(t) \tag{2.56}$$

where $d\mathbf{n}_p(t)$ is a generalized Poisson process with rate vector $\lambda(t)$, and let the jump vectors a_i be independent and identically distributed with density function

$$p_a(\alpha) \tag{2.56}$$

Let $p = p_{\mathbf{x}}(\mathbf{u}, t \,|\, \mathbf{x}(s))$ be the transition density for this process. Then p satisfies the Feller-Kolmogorov equation

$$\frac{\partial p}{\partial t} = -\sum_{i=1}^{n} \frac{\partial}{\partial u_i}(f_i p) + \sum_{i=1}^{n} \lambda_i(t)\,[p * p_{a_i} - p] \tag{2.57}$$

The proof of this theorem follows immediately by setting all Q_{ij} identically to zero in Theorem 2.2. The simplest form of the Feller-Kolmogorov equation is for a scalar case ($n = 1$) with $f \equiv 0$, the jump having unit amplitude only; that is,

$$p_{a_1} = \delta(a_1 - 1) \tag{2.58}$$

We can also assume that $\lambda_1(t)$ is time independent. Thus, the Feller-Kolmogorov equation is

$$\frac{\partial p}{\partial t} = \lambda\,[p_x(u - 1, t \,|\, x(t_0)) - p_x(u, t \,|\, x(t_0))] \tag{2.59}$$

Let us assume that $x(t_0) = 0$ so that x can only take on positive integer values. Thus, $x(t)$ can equal only $0, 1, 2, \cdots, n, n + 1, \cdots$. This raises the problem that p is then impulsive. To avoid this difficulty we first note that the impulses of p occur only at the integers, so that the probability that $x(t)$ equals n is

$$P_n(t) = \int_{n-\varepsilon}^{n+\varepsilon} p_x(u, t \,|\, x(t_0) = 0)\, du \tag{2.60}$$

where $0 < \varepsilon < 1$. Now (2.61) can be easily transformed by integrating over u from $n - \varepsilon$ to $n + \varepsilon$ and noting that

$$\begin{aligned} &\int_{n-\varepsilon}^{n+\varepsilon} p_x(u - 1, t \,|\, x(t_0) = 0)\, du \\ &= \int_{n-1-\varepsilon}^{n-1+\varepsilon} p_x(u, t \,|\, x(t_0) = 0)\, du = P_{n-1}(t) \end{aligned} \tag{2.61}$$

Thus, (2.61) becomes

$$\frac{\partial P_n(t)}{\partial t} = \lambda P_{n-1}(t) - \lambda P_n(t) \tag{2.62}$$

for all $n \geqq 1$. Clearly, for $n = 0$ we have

$$\frac{\partial P_0(t)}{\partial t} = -\lambda P_0(t) \tag{2.63}$$

The solution to these sets of equations can be obtained recursively and are discussed in Feller [1, Chapter 17]. They are

$$P_n(t) = \frac{[\lambda t]^n}{n!} \exp(-\lambda t) \tag{2.64}$$

the Poisson distribution for $P_n(t)$, the probability that $x(t)$ equals n. This simple example is the analogue of the solution Einstein obtained for the Fokker-Planck equation for the case of simple Brownian motion. The interesting issue here is that the generator of a probability density function can be used to define a process and vice versa. This aspect is more fully explored in Nelson for Brownian motion and by Dynkin [1] for general Markov processes. A complete discussion of these forms of Poisson processes is contained in Feller [1, Chapter 17].

The preceding analysis depended upon the fact that $\mathbf{x}(t)$ was a Markov process. Similar analysis can be made if this assumption is not made and the resulting equations are called the *generalized Fokker-Planck equations*. The reason for presenting such an analysis is not only to provide the necessary completeness to the Fokker-Planck equation but to introduce the moment approach to the development of transition densities.

Let us begin by considering an arbitrary random process $\mathbf{x}(t)$. As before, we are seeking the probability density function of the process $\mathbf{x}(t)$, given a set of past values of the process. The change in emphasis now is on the conditioning; for a Markov process it was sufficient to consider only the process at a single point, whereas for an arbitrary process an arbitrary set of past values must be considered. This probability density suffices to describe the process, because it allows for a complete statistical characterization. Thus, it acts as a generalized transition probability density function. To see more fully what is needed, recall that to completely characterize the process we must have the joint probability density function of the process for any number of times t belonging to some set T. Specifically, we need

$$p_{\mathbf{x}_1,\cdots,\mathbf{x}_n}(\mathbf{u}_1, t_1; \mathbf{u}_2, t_2; \cdots; \mathbf{u}_n, t_n) \tag{2.65}$$

for any set of $\{t_i\}$. For Markov processes it is clear that

$$\begin{aligned} &p_{\mathbf{x}_1,\cdots,\mathbf{x}_n}(\mathbf{u}_1, t_1; \cdots; \mathbf{u}_n, t_n) \\ &= p_{\mathbf{x}_n}(\mathbf{u}_n, t_n \,|\, \mathbf{x}(t_{n-1})) \cdots p_{\mathbf{x}_2}(\mathbf{u}_2, t_2 \,|\, \mathbf{x}(t_1))\, p_{\mathbf{x}_1}(\mathbf{u}_1, t_1) \end{aligned} \tag{2.66}$$

where it is thus sufficient to have the transition density $p_{\mathbf{x}}(\mathbf{u}, t \,|\, \mathbf{x}(s))$ for all t and $\mathbf{x}(s)$. But for non-Markov processes we have

$$p_{\mathbf{x}_n,\cdots,\mathbf{x}_1}(\mathbf{u}_1, t_1;\cdots;\mathbf{u}_n, t_n) = p_{\mathbf{x}_n}(\mathbf{u}_n, t_n \,|\, \mathbf{x}(t_{n-1})\cdots\mathbf{x}(t_1))$$
$$p_{\mathbf{x}_{n-1}}(\mathbf{u}_{n-1}, t_{n-1} \,|\, \mathbf{x}(t_{n-2})\cdots\mathbf{x}(t_1))\cdots p_{\mathbf{x}_1}(\mathbf{u}_1, t_1) \tag{2.67}$$

Thus, what is needed for the non-Markov case is

$$p_{\mathbf{x}}(\mathbf{u}, t \,|\, (X, T)) \tag{2.68}$$

where X, T represents the sets

$$(X, T) = \{\mathbf{x}(t_1);\cdots;\mathbf{x}(t_i)\} \tag{2.69}$$

for some sets of times $T = \{t_1,\cdots, t_i\}$. With this function (2.70) a complete statistical characterization can be obtained.

One immediate fact is that the rule for total probability follows:

$$p_{\mathbf{x}}(\mathbf{u}, t + \Delta t \,|\, (X, T))$$
$$= \int p_{\mathbf{x}}(\mathbf{u}, t + \Delta t \,|\, \mathbf{x}(t) = \mathbf{v}:(X, T)) p_{\mathbf{x}}(\mathbf{v}, t \,|\, (X, T))\, d\mathbf{v} \tag{2.70}$$

where it is assumed that $\mathbf{x}(t) \notin (X, T)$. With this formulation of total probability it can be argued that $p_{\mathbf{x}}(\mathbf{u}, t \,|\, (X, T))$ and $\partial/\partial t\, p_{\mathbf{x}}(\mathbf{u}, t \,|\, (X, T))$ determine $p_{\mathbf{x}}(\mathbf{u}, t \,|\, (X, T))$ for all t under suitable conditions (see Wong [2], p. 181). Thus, as before, our objective is to obtain the time variation of the probability density function (2.70).

To begin the analysis, we define the conditional characteristic function of this increment of the process. It is given by

$$M_{d\mathbf{x}}(\mathbf{u}, t + \Delta t \,|\, \mathbf{x}(t), X, T)$$
$$= E[\exp\{j\mathbf{u}^T[\mathbf{x}(t + \Delta t) - \mathbf{x}(t)]\} \,|\, \mathbf{x}(t), X, T] \tag{2.71}$$

Now $M_{d\mathbf{x}}$ can be expanded in a Taylor series about the vector $\mathbf{u} = \mathbf{0}$. This yields

$$M_{d\mathbf{x}}(\mathbf{u}, t + \Delta t \,|\, \mathbf{x}(t), X, T)$$
$$= 1 + \sum_{i_1=1}^{n} \frac{\partial M_{d\mathbf{x}}}{\partial u_{i_1}}\bigg|_{\mathbf{u}=0} u_{i_1} + \frac{1}{2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n} \frac{\partial^2 M_{d\mathbf{x}}}{\partial u_{i_1}\partial u_{i_2}}\bigg|_{\mathbf{u}=0} u_{i_1} u_{i_2} + \cdots \tag{2.72}$$

From the above we note that

$$\frac{\partial M_{d\mathbf{x}}}{\partial u_{i_1}}\bigg|_{\mathbf{u}=0} = E[j(x_{i_1}(t + \Delta t) - x_{i_1}(t)) \,|\, \mathbf{x}(t), X, T] \tag{2.73}$$

and that

$$\frac{\partial^2 M_{d\mathbf{x}}}{\partial u_{i_1}\partial u_{i_2}}\bigg|_{\mathbf{u}=0} = E[(j)^2(x_{i_1}(t + \Delta t) - x_{i_1}(t))$$
$$(x_{i_2}(t + \Delta t) - x_{i_2}(t)) \,|\, \mathbf{x}(t), X, T] \tag{2.74}$$

where the higher-order derivatives are similarly defined. Now define the quantities

$$a_{i_1} = E\left[x_{i_1}(t + \Delta t) - x_{i_1}(t) \middle| \mathbf{x}(t), X, T\right] \tag{2.75}$$

$$a_{i_1 i_2} = E\left[(x_{i_1}(t + \Delta t) - x_{i_1}(t))(x_{i_2}(t + \Delta t) - x_{i_2}(t)) \middle| \mathbf{x}(t), X, T\right] \tag{2.76}$$

These represent the conditional moments of the increments of the process. With these definitions and the expansion of the Taylor series, we can write $M_{d\mathbf{x}}$ as

$$M_{d\mathbf{x}} = 1 + \sum_{i_1=1}^{n} a_{i_1}(ju_{i_1}) + \frac{1}{2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n} a_{i_1 i_2}(ju_{i_1} ju_{i_2}) + \cdots \tag{2.77}$$

Now recall from the law of conditional probability that we have the relationship given in (2.71). Also note that

$$p_{\mathbf{x}}(\mathbf{u}, t + \Delta t \,|\mathbf{x}(t) = \mathbf{v}, X, T) = \frac{1}{(2\pi)^n}\int M_{d\mathbf{x}}(\mathbf{v}, t + \Delta t \,|\mathbf{x}(t), X, T) \exp\left[-j\mathbf{s}^T(\mathbf{u} - \mathbf{v})\right] d\mathbf{s} \tag{2.78}$$

which follows directly from the inverse Fourier transform relationship. Using the Taylor-series expansion of $M_{d\mathbf{x}}$ in this inverse transform, we obtain

$$p_{\mathbf{x}}(\mathbf{u}, t + \Delta t \,|\mathbf{x}(t) = \mathbf{v}, X, T) = \delta(\mathbf{u} - \mathbf{v}) + \sum_{i_1=1}^{n} a_{i_1}(-1)\frac{\partial\delta(\mathbf{u} - \mathbf{y})}{\partial u_{i_1}} + \frac{1}{2}\sum_{i_1=1}^{n}\sum_{i_2=2}^{n} a_{i_1 i_2}(-1)^2 \frac{\partial^2\delta(\mathbf{u} - \mathbf{v})}{\partial u_{i_1}\partial u_{i_2}} + \cdots \tag{2.79}$$

which follows directly from the identity

$$\delta(\mathbf{u} - \mathbf{v}) = \frac{1}{(2\pi)^n}\int \exp\left[-j\mathbf{s}^T(\mathbf{u} - \mathbf{v})\right] d\mathbf{s} \tag{2.80}$$

Likewise,

$$\begin{aligned}
&\frac{1}{(2\pi)^n}\int js_{i_1} \exp\left[-j\mathbf{s}^T(\mathbf{u} - \mathbf{v})\right] d\mathbf{s} \\
&= (-1)\frac{\partial}{\partial u_{i_1}}\frac{1}{(2\pi)^n}\int \exp\left[-j\mathbf{s}^T(\mathbf{u} - \mathbf{v})\right] d\mathbf{s} \\
&= (-1)\frac{\partial\delta(\mathbf{u} - \mathbf{v})}{\partial u_{i_1}}
\end{aligned} \tag{2.81}$$

and similarly for higher-order integrals. Thus, using (2.81) in the law of total probability (2.70), we obtain after integrating by parts

$$\begin{aligned}
p_{\mathbf{x}}(\mathbf{u}, t + \Delta t \,|X, T) &= p_{\mathbf{x}}(\mathbf{u}, t \,|X, T) \\
&+ \sum_{i_1=1}^{n}(-1)\frac{\partial(a_{i_1} p_{\mathbf{x}}(\mathbf{u}, t \,|X, T))}{\partial u_{i_1}} \\
&+ \frac{1}{2}\sum_{i_1=1}^{n}(-1)^2\frac{\partial^2(a_{i_1 i_1} p_{\mathbf{x}}(\mathbf{u}, t \,|X, T))}{\partial u_{i_1}\partial u_{i_2}} \\
&+ \cdots
\end{aligned} \tag{2.82}$$

which follows from the integrations

$$\int_{-\infty}^{\infty} a_{i_1}(\mathbf{v}) p_{\mathbf{x}}(\mathbf{v}, t \mid X, T) \frac{\partial \delta(\mathbf{u} - \mathbf{v})}{\partial u_{i_1}} \, d\mathbf{v}$$
$$= \frac{\partial}{\partial u_{i_1}} \int_{-\infty}^{\infty} a_{i_1}(\mathbf{v}) p_{\mathbf{x}}(\mathbf{v}, t \mid X, T) \delta(\mathbf{u} - \mathbf{v}) \, d\mathbf{v}$$
$$= \frac{\partial (a_{i_1}(\mathbf{u}) p_{\mathbf{x}}(\mathbf{v}, t \mid X, T))}{\partial u_{i_1}} \tag{2.83}$$

All higher-order derivatives appearing in (2.82) are obtained in a similar fashion. A propagation for the conditional density can be obtained by rearranging and dividing by Δt and taking the limit. To do this, we first define the quantity

$$A_{m_1 \cdots m_n}(\mathbf{u}) \tag{2.84}$$
$$= \lim_{\Delta t \to 0} \frac{E[(x_1(t + \Delta t) - x_1(t))^{m_1} \cdots (x_n(t + \Delta t) - x_n(t))^{m_n} \mid \mathbf{x}(t) = \mathbf{u}, X, T]}{\Delta t}$$

Also, define

$$\frac{\partial p_{\mathbf{x}}(\mathbf{u}, t \mid X, T)}{\partial t} = \lim_{\Delta t \to 0} \frac{p_{\mathbf{x}}(\mathbf{u}, t + \Delta t \mid X, T) - p_{\mathbf{x}}(\mathbf{u}, t \mid X, T)}{\Delta t} \tag{2.85}$$

Then (2.85) can be written as

$$\frac{\partial p}{\partial t} = \sum_{m_1=0}^{\infty} \cdots \sum_{m_n=0}^{\infty} \left(\prod_{i=1}^{n} \frac{(-1)^{m_i} D_i^{m_i}}{(m_i)!} A_{m_1 \cdots m_n} p \right) \tag{2.86}$$

where $p = p_{\mathbf{x}}(\mathbf{u}, t \mid X, T)$ and

$$D_i^{m_i} = \frac{\partial^{m_i}}{\partial u_i^{m_i}} \tag{2.87}$$

This is the generalized form of the Fokker-Planck equation. It is clearly much more complex than the Fokker-Planck equation, because of the presence of the infinite number of derivatives of the density function. Such a form also makes severe restrictions on the types of acceptable density functions, infinitely differentiable, as well as requiring knowledge of all the coefficients $A_{m_1, \ldots, m_n}$. To avoid these difficulties, we seek conditions on the process for which $\partial p / \partial t$ is determined by a finite set of derivatives. Specifically, we seek conditions that reduce it to a form similar to that of the Fokker-Planck equation. The following two lemmas provide those conditions and are due to Pawula.

LEMMA 2.2. Let $A_{m_1,0,\ldots,0}$ be given by

$$A_{m_1,0,\ldots,0} = \lim_{\Delta t \to 0} \frac{E[(x_1(t + \Delta t) - x_1(t))^{m_1} \mid \mathbf{x}(t), X, T]}{\Delta t} \tag{2.88}$$

If $A_{m_1,0,\ldots,0}$ is zero for some even m_1, then

$$A_{m_1,0,\cdots,0} = 0; \qquad \forall m_1 \geqq 3 \tag{2.89}$$

Proof. For m_1 odd and $m_1 \geqq 3$, we have

$$\begin{aligned} A_{m_1,0,\cdots,0} &= \lim_{\Delta t \to 0} \frac{E[(x_1(t+\Delta t) - x_1(t))^{m_1} | \mathbf{x}(t), X, T]}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{1}{\Delta t} E_c\left[(x_1(t+\Delta t) - x_1(t))^{(m_1-1)/2}\right. \\ &\qquad \left.(x_1(t+\Delta t) - x_1(t))^{(m_1-1)/2}\right] \end{aligned} \tag{2.90}$$

where $E_c[\ ]$ is the expectation operation with the appropriate conditioning supplied. Now, using the Schwarz inequality, we find that

$$\begin{aligned} A^2_{m_1,0,\cdots,0} \leqq \lim_{\Delta t \to 0} &\frac{1}{\Delta t} E_c[(x_1(t+\Delta t) - x_1(t))^{m_1-1}] \\ &\frac{1}{\Delta t} E_c\left[(x_1(t+\Delta t) - x_1(t))^{m_1+1}\right] \end{aligned} \tag{2.91}$$

or

$$A^2_{m_1,0,\cdots,0} \leqq A_{m_1-1,0,\cdots,0}\, A_{m_1+1,0,\cdots,0} \tag{2.92}$$

Likewise, for $m_1 \geqq 3$ we can write

$$\begin{aligned} A_{m_1,0,\cdots,0} &= \frac{1}{\Delta t} E_c\left[(x_1(t+\Delta t) - x_1(t))^{m_1}\right] \\ &= \frac{1}{\Delta t} E_c[(x_1(t+\Delta t) - x_1(t))^{(m_1-2)/2} \\ &\qquad (x_1(t+\Delta t) - x_1(t))^{(m_1-2)/2}] \end{aligned} \tag{2.93}$$

Following the previous result, we obtain

$$A^2_{m_1,0,\cdots,0} \leqq A_{m_1-2,0,\cdots,0}\, A_{m_1+2,0,\cdots,0} \tag{2.94}$$

for $n \geqq 4$ and n even. Thus, for $m_1 = r$, where r is an even number, if $A_{r,0,\cdots,0}$ is zero then

$$A^2_{r-2,0,\cdots,0} \leqq A_{r-4,0,\cdots,0}\, A_{r,0,\cdots,0}; \qquad r \geqq 0 \tag{2.95}$$

$$A^2_{r-1,0,\cdots,0} \leqq A_{r-2,0,\cdots,0}\, A_{r,0,\cdots,0}; \qquad r \geqq 2 \tag{2.96}$$

$$A^2_{r+1,0,\cdots,0} \leqq A_{r,0,\cdots,0}\, A_{r+2,0,\cdots,0}; \qquad r \geqq 2 \tag{2.97}$$

$$A^2_{r+2,0,\cdots,0} \leqq A_{r+4,0,\cdots,0}\, A_{r,0,\cdots,0}; \qquad r \geqq 2 \tag{2.98}$$

Hence, $A_{r-2,0,\cdots,0}$, $A_{r-1,0,\cdots,0}$, $A_{r+1,0,\cdots,0}$, $A_{r+2,0,\cdots,0}$ must all be identically zero if $A_{r,0,\cdots,0}$ is zero and if all A are bounded. Hence, $A_{m_1,0,\cdots,0}$ must all be zero if $A_{m_1,0,\cdots,0}$ is zero for any m_1 satisfying the hypothesis. ∎

We can now proceed to give the conditions under which this will hold for arbitrary values of A.

LEMMA 2.3. If each of the moments $A_{m_1,0,\cdots,0}$, $A_{0,m_2,0,\cdots,0}$, $\cdots$, $A_{0,\cdots,0,m_n}$ is finite and vanishes for some even m_i, then

$$A_{m_1,\cdots,m_n} = 0 \tag{2.99}$$

for all m_i such that

$$\sum_{i=1}^{n} m_i \geqq 0 \tag{2.100}$$

This clearly will allow us to simplify the propagation equation to the extent that it resembles the Fokker-Planck equation developed previously. The result will imply that only the variables $A_{1,0,\cdots,0}$, $\cdots$, $A_{0,\cdots,0,1}$, and the coefficients $A_{2,0,\cdots,0}$ and so on will be possibly nonzero if the higher moments can be shown to be zero. This assertion on higher moments is usually easy to justify for systems of the SSSR form but in general is not easily ascertained.

Proof. We shall prove this lemma by induction, by doing it for the case of $n = 3$. For higher-order systems (i.e., $n \geqq 3$) the induction is obvious. Let A_{m_1,m_2,m_3} be defined as before. Now,

$$A^4_{m_1,m_2,m_3} = \left[\lim_{\Delta t\to 0} \frac{1}{\Delta t} E_c[(x_1(t+\Delta t) - x_1(t))^{m_1}(x_2(t+\Delta t) - x_2(t))^{m_2}\right.$$

$$\left.(x_3(t+\Delta t) - x_3(t))^{m_3}]^4\right] \leqq \lim_{\Delta t\to 0}\left[\frac{1}{\Delta t} E_c[(x_1(t+\Delta t) - x_1(t))^{2m_1}]\right]^2$$

$$\left[\frac{1}{\Delta t} E_c[(x_2(t+\Delta t) - x_2(t))^{2m_2}(x_3(t+\Delta t) - x_3(t))^{2m_3}]\right]^2 \tag{2.101}$$

which follows from the Schwarz inequality. Using this inequality once more, we show that

$$A^4_{m_1,m_2,m_3} \leqq A^2_{2m_1,0,0}\, A_{0,4m_2,0}\, A_{0,0,4m_3} \tag{2.102}$$

Thus, if as assumed the last two moments on the right in the above inequality vanish, then the left-hand side vanishes. Thus,

$$A_{m_1,m_2,m_3} = 0 \tag{2.103}$$

for all m_1, m_2, m_3. Now consider the case

$$A^2_{0,m_2,m_3} \leqq A_{0,2m_2,0}\, A_{0,0,2m_3} \tag{2.104}$$

But again this vanishes for $m_2, m_3 > 0$ and $m_2 + m_3 \geqq 3$. This then completes the lemma. ∎

If we now assume that the higher-order moments vanish, then it is sufficient to consider only first and second moments. It then becomes convenient to define them separately. Let

$$B_i(\mathbf{u}, \mathbf{t}) = \lim_{\Delta t\to 0} \frac{1}{\Delta t} E\left[x_i(t+\Delta t) - x_i(t) \middle| \mathbf{x}(t) = \mathbf{u}, X, T\right] \tag{2.105}$$

and

$$C_{ij}(\mathbf{u}, t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} E\left[(x_i(t + \Delta t) - x_i(t))\right.$$
$$\left.(x_j(t + \Delta t) - x_j(t))\middle|\mathbf{x}(t) = \mathbf{u}, X, T\right] \tag{2.106}$$

With these identifications we can now state the following theorem.

THEOREM 2.3

Let $p = p_{\mathbf{x}}(\mathbf{u}, t\,|X, T)$ for some set X, T and let each of the moments $A_{m_i,0,\cdots,0}$, $\cdots$, $A_{0,\cdots,m_n}$ vanish for some even m_i. Then the transition density satisfies the equation

$$\frac{\partial p}{\partial t} = -\sum_{i=1}^{n} \frac{\partial(B_i p)}{\partial u_i} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2(C_{ij} p)}{\partial u_i\, \partial u_j} \tag{2.107}$$

The similarity of the above equation to the Fokker-Planck equation is less than coincidental. The above assumptions are historically those used by Wang and Uhlenbeck in their classic paper on Brownian motion. Clearly, if the process is Markovian, then

$$p_{\mathbf{x}}(\mathbf{u}, t\,|X, T) = p_{\mathbf{x}}((\mathbf{u}, t\,|\mathbf{x}(s)) \tag{2.108}$$

where $s = \sup\{t_i : t_i \in T\}$. The assumption made by Wang and Uhlenbeck was that all moments above the second must vanish. But Pawula (1967) showed, as was presented here, that it was sufficient for only a finite set of moments to vanish, in order to satisfy the conditions. Applications of these techniques are discussed in Middleton (Chapter 10) and in the problems. A rigorous approach is discussed and developed by Dynkin [1, Chapter 5, para. 6].

The propagation equations for the conditional density are important in our study of estimation for several reasons. The first is that they play an integral role in the analysis of the propagation of the conditional density of the nonlinear estimates. This is because the conditional density evaluated in this section governs the state propagation subject to no measurements after some initial estimate of the state at time t_0. Intuitively, therefore, we should expect the conditional density to follow the path prescribed by the dynamics of the system, with any later measurements acting only as perturbations to the given trajectory. We shall in the next section build upon this concept and evaluate the propagation of the conditional density when measurements are present. The second important role that the conditional densities play is in obtaining prediction values of the state in the absence of measurements. We discussed this in a previous example. Exact analytical extensions have been made in this area by Dym and McKean and by Dudley [1].

5.3 PROPAGATION OF CONDITIONAL DENSITY

In the previous section we considered a model for a dynamical system and then obtained a propagation equation for the transition probability (conditional density function) of the given system. As we also stated, the estimation problem contains not only a state that is to be estimated but also a measurement from which this estimate is to be obtained. The two types of measurements we shall consider are the Gaussian additive disturbances and the Poisson step process measurements. As already discussed, both encompass a wide class of actual measurements that are found or easily approximated in actual practice.

As we recall, the state was given by the solution to the following differential equations:

$$d\mathbf{x}(t) = \mathbf{f}(x, t)\, dt + d\mathbf{n}(t) \tag{3.1}$$

where $d\mathbf{n}$ was a zero mean independent increment process that was at most the sum of a Wiener process and an independent generalized Poisson process. The measurements, however, fall into two classes. The $(m \times 1)$-vector Gaussian measurement $d\mathbf{y}(t)$ is given by

$$d\mathbf{y}(t) = \mathbf{h}(\mathbf{x}(t), t) + d\mathbf{w}(t) \tag{3.2}$$

where $\mathbf{w}(t)$ is a Wiener process with zero mean and covariance

$$E[d\mathbf{w}(t)\, d\mathbf{w}^T(t)] = \mathbf{R}(t)\, dt \tag{3.3}$$

where $\mathbf{R}(t)$ is an $m \times m$ nonsingular covariance matrix. Clearly the nonlinearity must possess certain properties that will allow us to obtain useful estimates of the state. We shall discuss these matters in Chapter 7

The Poisson model is more simply stated. What is observed is a Poisson counting process $d\mathbf{N}(t)$ that is an $(m \times 1)$-vector process with arrival rate $\boldsymbol{\lambda}(\mathbf{x}(t), t)$; where $\boldsymbol{\lambda}(\mathbf{x}(t), t)$ is also an $m \times 1$ vector.

As was discussed in Chapter 4, $O_{t_0,t}$ will be the minimum σ-field generated by the observation process. Thus, given $O_{t_0,t}$ as generated by either $d\mathbf{y}(t)$ or $d\mathbf{N}(t)$ we are then asked to obtain the MMSE estimate of the state at time t. If we let $\hat{\mathbf{x}}(t)$ be that estimate, then

$$\hat{\mathbf{x}}(t) = E[\mathbf{x}(t) | O_{t_0,t}] \tag{3.4}$$

yields such an estimate. It is also sufficient to have $p_{\mathbf{x}}(\mathbf{u}, t | O_{t_0,t})$, the conditional probability density of $\mathbf{x}(t)$, given $O_{t_0,t}$, since

$$\hat{\mathbf{x}}(t) = \int \mathbf{u} p_{\mathbf{x}}(\mathbf{u}, t | O_{t_0,t})\, d\mathbf{u} \tag{3.5}$$

Thus, the object of this section is to obtain $p_{\mathbf{x}}(\mathbf{u}, t | O_{t_0,t})$ for both measurement processes. To do so, we first present a theorem that, assuming certain

properties of the conditional probability density, obtains a propagation equation closely related to those obtained in the last section. Once this is done, we proceed to show that the hypothesis holds for both cases and evaluate the necessary functional forms. This then leads to partial-differential integral equations for the conditional density in both cases.

The first one to propose a solution to this problem was Stratonovich [1] in 1959. His solution was based upon a different interpretation of stochastic integrals than is now accepted and thus is considered in error, because of the omission of a first-order term. The true breakthrough came in 1964 when Kushner [1, 2] used the correct Ito formulations. Striebel in 1965 used a different formulation than Kushner and arrived at the result for the linear case. Indirectly, this result was already known quite extensively through the work of Kalman [1] in 1959 and Kalman and Bucy in 1960. In 1967, Kushner [3] rederived his results rigorously and provided a stronger mathematical foundation to the technique. Since then other approaches using the representation theorem (see the next section) have been proposed as more elegant mathematically. This theorem was first proposed in its present form by Bucy in 1967, and it has been extensively elaborated on by Kallianpur and Striebel [1–3] in 1968 and 1969 and by Fujisaki, Kallianpur, and Kunita in 1971. Almost all of the above approaches rely upon measure theoretic concepts except the earlier works of Kushner [1, 2]. We shall therefore follow these two references to some degree, leaving the other considerations to the reader's interest. Similarly, for the step Poisson measurements we shall follow the non-measure theoretic approaches of Snyder [4, 5], who first solved this problem.

We now present the basic theorem on which all propagations equations rest.

THEOREM 3.1

Let $\mathbf{x}(t)$ be an $(n \times 1)$-vector Markov process given by

$$d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}, t)\, dt + d\mathbf{n}(t) \tag{3.6}$$

and let $d\mathbf{y}(t)$ be an $(m \times 1)$-vector Markov measurement process that depends pointwise on $\mathbf{x}(t)$. Let $p_{\mathbf{x}}(\mathbf{u}, t \,|\, O_{t_0,t})$ be the conditional probability density function of the process $\mathbf{x}$ at time t given $O_{t_0,t}$ the minimum σ-field generated by the measurement set $\mathbf{y}(s)$, $s \in [t_0, t]$. If there exists a function $q(d\mathbf{y}, dt, \mathbf{u})$ such that

$$p_{\mathbf{x}}(\mathbf{u}, t \,|\, O_{t_0,t+dt}) = p_{\mathbf{x}}(\mathbf{u}, t \,|\, O_{t_0,t})(1 + q(d\mathbf{y}, dt, \mathbf{u})) + o(dt) \tag{3.7}$$

which is $O(dt)$,* then,

$$\frac{\partial p}{\partial t} = L^{+} p + q(\mathrm{dy}, \mathrm{dt}, \mathbf{u})p \tag{3.8}$$

*A function $O(dt)$ is one that is such that $\lim_{dt \to 0} \frac{q(d\mathbf{y}, dt, \mathbf{u})}{dt} = \text{constant}$.

where $p = p_{\mathbf{x}}(\mathbf{u}, t \,|\, O_{t_0,t})$ and L^+ is the generator associated with the Markov process $\mathbf{x}(t)$.

The L^+ operator was the operator that was obtained in the last sectoin for several types of processes. The object in that section was to obtain L^+ from the Bartlett-Moyal theorems. In this section the object will be to obtain the function $q(d\mathbf{y},dt,\mathbf{u})$, which represents how the information affects the propagation of the conditional probability density function.

Proof. Let $M_{\mathbf{x}}(\mathbf{v}, t + dt \,|\, O_{t_0,t+dt})$ be the conditional characteristic function. Let

$$\frac{\partial M_{\mathbf{x}}(\mathbf{v}, t \,|\, O_{t_0,t})}{\partial t} = \lim_{\Delta t \to 0} \frac{M_{\mathbf{x}}(\mathbf{v}, t + \Delta t \,|\, O_{t_0,t+\Delta t}) - M_{\mathbf{x}}(\mathbf{v}, t \,|\, O_{t_0,t})}{\Delta t} \tag{3.9}$$

The object of the proof is to obtain this derivative. Now

$$M_{\mathbf{x}}(\mathbf{v}, t + \Delta t \,|\, O_{t_0,t+\Delta t}) = \int \exp(j\mathbf{v}^T\mathbf{u}) p_{\mathbf{x}}(\mathbf{u}, t + \Delta t \,|\, O_{t_0,t+\Delta t})\, d\mathbf{u} \tag{3.10}$$

Now using the law of total probability this can be writetn as

$$\begin{aligned} &M_{\mathbf{x}}(\mathbf{v}, t + \Delta t \,|\, O_{t_0,t+\Delta t}) \\ &= \int\int \exp(j\mathbf{v}^T\mathbf{u}) p_{\mathbf{x}}(\mathbf{u}, t + \Delta t \,|\, \mathbf{x}(t) = s, O_{t_0,t+\Delta t}) p_{\mathbf{x}}(\mathbf{s}, t \,|\, O_{t_0,t+\Delta t})\, d\mathbf{u}\, d\mathbf{s} \end{aligned} \tag{3.11}$$

But

$$p_{\mathbf{x}}(\mathbf{u}, t + \Delta t \,|\, \mathbf{x}(t) = \mathbf{s}, O_{t_0,t+\Delta t}) = p_{\mathbf{x}}(\mathbf{u}, t + \Delta t \,|\, \mathbf{x}(t) = \mathbf{s}) \tag{3.12}$$

that is, it is independent of the observation. Add and subtract $\exp(j\mathbf{v}^T\mathbf{s})$ to the integral to obtain

$$\begin{aligned} M_{\mathbf{x}}(\mathbf{v}, t + \Delta t \,|\, O_{t_0,t+\Delta t}) &= \int \exp(j\mathbf{v}^T\mathbf{s}) p_{\mathbf{x}}(\mathbf{s}, t \,|\, O_{t_0,t+\Delta t}) \\ &\quad \int \exp[j\mathbf{v}^T(\mathbf{u} - \mathbf{s})] p_{\mathbf{x}}(\mathbf{u}, t + \Delta t \,|\, \mathbf{x}(t))\, d\mathbf{u}\, d\mathbf{s} \\ &= \int \exp(j\mathbf{v}^T\mathbf{s}) p_{\mathbf{x}}(\mathbf{s}, t \,|\, O_{t_0,t+\Delta t}) \\ &\quad E[e^{j\mathbf{v}^T d\mathbf{x}(t)} \,|\, \mathbf{x}(t) = \mathbf{s}]\, d\mathbf{s} \end{aligned} \tag{3.13}$$

Now by hypothesis we can write

$$p_{\mathbf{x}}(\mathbf{s}, t \,|\, O_{t_0,t+\Delta t}) = p_{\mathbf{x}}(\mathbf{s}, t \,|\, O_{t_0,t}) + q(d\mathbf{y}, dt, \mathbf{s}) p_{\mathbf{x}}(\mathbf{s}, t \,|\, O_{t_0,t}) \tag{3.14}$$

Thus, using this in the characteristic function, we can write for the difference

$$\begin{aligned} &M_{\mathbf{x}}(\mathbf{v}, t + \Delta t \,|\, O_{t_0,t+\Delta t}) - M_{\mathbf{x}}(\mathbf{v}, t \,|\, O_{t_0,t}) \\ &= \int p_{\mathbf{x}}(\mathbf{s}, t \,|\, O_{t_0,t}) \exp(j\mathbf{v}^T\mathbf{s}) E[\exp(j\mathbf{v}^T d\mathbf{x}(t)) - 1 \,|\, \mathbf{x}(t) = \mathbf{s}]\, d\mathbf{s} \\ &+ \int p_{\mathbf{x}}(\mathbf{s}, t \,|\, O_{t_0,t}) \exp(j\mathbf{v}^T\mathbf{s}) q(d\mathbf{y}, dt, \mathbf{s}) \\ &\quad E[\exp[j\mathbf{v}^T d\mathbf{x}(t)] \,|\, \mathbf{x}(t) = \mathbf{s}]\, d\mathbf{s} \end{aligned} \tag{3.15}$$

The first term on the right is what we obtained from the Bartlet-Moyal theorem and under inverse transforms leads to the L^+ operator. The second term contains the expression

$$E[\exp[jv^T d\mathbf{x}(t)]|\mathbf{x}(t) = \mathbf{s}] = 1 + O(dt) \tag{3.16}$$

But, since $q(d\mathbf{y}, dt, \mathbf{s})$ is already $O(dt)$ and two $O(dt)$ terms are $o(dt)$, then

$$\begin{aligned} &\int p_{\mathbf{x}}(\mathbf{s}, t\,|O_{t_0,t}) \exp(jv^T\mathbf{s}) q(d\mathbf{y}, dt, \mathbf{s}) \\ &E[\exp(jv^T d\mathbf{x}(t))|\mathbf{x}(t) = \mathbf{s}]\, d\mathbf{s} \\ &= \int p_{\mathbf{x}}(\mathbf{s}, t\,|O_{t_0,t}) \exp(jv^T\mathbf{s}) q(d\mathbf{y}, dt, \mathbf{s})\, d\mathbf{s} + o(dt) \end{aligned} \tag{3.17}$$

Thus, by dividing through by Δt, taking the limit, inverse transforming and properly identifying L^+, we obtain the desired result. ∎

Inherent in the preceding theorem was the fact that the function $q(d\mathbf{y}, dt, \mathbf{u})$ could be evaluated for any $d\mathbf{y}(t)$ or $d\mathbf{N}(t)$. We now proceed to evaluate the function for the case of additive Gaussian measurements. The proof is long and may be omitted upon first reading.

LEMMA 3.1. Let $\mathbf{x}(t)$ be an $(n \times 1)$-vector Markov process generated by

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\, dt + d\mathbf{n}(t) \tag{3.18}$$

and let $d\mathbf{y}$ be a continuous $(m \times 1)$-vector Markov process given by

$$d\mathbf{y} = \mathbf{h}(\mathbf{x}, t)\, dt + d\mathbf{w}(t) \tag{3.19}$$

where $d\mathbf{w}$ is an $(m \times 1)$ -vector Wiener process with covariance

$$E[d\mathbf{w}\, d\mathbf{w}^T] = \mathbf{R}(t)\, dt \tag{3.20}$$

where $\mathbf{R}$ is an $m \times m$ nonsingular matrix. Let $p_{\mathbf{x}}(\mathbf{u}, t\,|O_{t_0,t+dt})$ be the conditional probability density function of the process $\mathbf{x}(t)$, given $O_{t_0,t+dt}$ the minimum σ-field generated by $\mathbf{y}(s)$, $s \in [t_0, t + dt)$. Then, to order dt,

$$p_{\mathbf{x}}(\mathbf{u}, t\,|O_{t_0,t+dt}) = p_{\mathbf{x}}(\mathbf{u}, t\,|O_{t_0,t})[1 + q(d\mathbf{y}, dt, \mathbf{u})] \tag{3.21}$$

where

$$\begin{aligned} q(d\mathbf{y}, dt, \mathbf{u}) = [&d\mathbf{y} - E[\mathbf{h}(\mathbf{x}(t), t)]dt]^T \mathbf{R}^{-1}(t) \\ &[\mathbf{h}(\mathbf{u}, t) - E[\mathbf{h}(\mathbf{x}(t), t)]] \end{aligned} \tag{3.22}$$

and

$$E[\mathbf{h}(\mathbf{x}(t), t)] = \int \mathbf{h}(\mathbf{u}, t) p_{\mathbf{x}}(\mathbf{u}, t\,|O_{t_0,t})\, d\mathbf{u} \tag{3.23}$$

Proof. Let $O_{t_0,t+dt} = O_{t_0,t} \bigcup d\mathbf{y}$, where $d\mathbf{y}$ is the amount of information obtained in the interval dt. Note that this is a heuristic argument and can be more rigorously stated by discretizing the intervals, using the fact that over

this discretization we have martingales and then using the martingale convergence results (see Doob) to show the resulting equivalence. This has been done in Kushner [3], and we have used it in Section 4.2 to define the conditional expectation. Thus, we use this fact to write

$$p_{\mathbf{x}}(\mathbf{u}, t \,|O_{t_0, t+dt}) = p_{\mathbf{x}}(\mathbf{u}, t \,|O_{t_0,t}, d\mathbf{y}) \tag{3.24}$$

Now, using Bayes's rule for total probability, we have

$$p_{\mathbf{x}}(\mathbf{u}, t \,|\, O_{t_0,t+dt}) = \frac{p_{d\mathbf{y}}(d\mathbf{y}|\mathbf{x}(t) = \mathbf{u}, O_{t_0,t})}{p_{d\mathbf{y}}(d\mathbf{y}|O_{t_0,t})}\, p_{\mathbf{x}}(\mathbf{u}, t \,|O_{t_0,t}) \tag{3.25}$$

Now recall that

$$p_{d\mathbf{y}}(d\mathbf{y}|O_{t_0,t}) = \int p_{\mathbf{x}}(\mathbf{v}, t \,|O_{t_0,t}) p_{d\mathbf{y}}(d\mathbf{y}|\mathbf{x}(t) = \mathbf{v}, O_{t_0,t})\, d\mathbf{v} \tag{3.26}$$

Further note that $d\mathbf{y}$ depends only on $\mathbf{x}(t)$ and is independent of $O_{t_0,t}$. Thus,

$$p_{d\mathbf{y}}(d\mathbf{y}|\mathbf{x}(t) = \mathbf{v}, O_{t_0,t}) = p_{d\mathbf{y}}(d\mathbf{y}|\mathbf{x}(t) = \mathbf{v}) \tag{3.27}$$

Then define the function $R(d\mathbf{y}, dt, \mathbf{u})$ as

$$R(d\mathbf{y}, dt, \mathbf{u}) = \frac{p_{d\mathbf{y}}(d\mathbf{y}|\mathbf{x}(t) = \mathbf{u})}{\int p_{d\mathbf{y}}(d\mathbf{y}|\mathbf{x}(t) = \mathbf{v}) p_{\mathbf{x}}(\mathbf{v}, t \,|O_{t_0,t})\, d\mathbf{v}} \tag{3.28}$$

Therefore, the conditional probability can be written as

$$p_{\mathbf{x}}(\mathbf{u}, t \,|O_{t_0,t+dt}) = R(d\mathbf{y}, dt, \mathbf{u}) p_{\mathbf{x}}(\mathbf{u}, t \,|O_{t_0,t}) \tag{3.29}$$

The desired result then depends upon expanding $R(d\mathbf{y},dt,\mathbf{u})$ in terms of the infinitesimal variables $d\mathbf{y}$ and dt. Before doing so, we note that the expansion should be of order dt. That is, we should include all terms such that they are not $o(dt)$. But there are clearly dt terms, dy_i terms, and $dy_i\, dy_j$ terms, since from Chapter 3 we found that

$$dy_i\, dy_j = R_{ij}\, dt + o(dt); \qquad \text{w.p.1} \tag{3.30}$$

where R_{ij} is the ijth entry of $\mathbf{R}(t)$. This fact follows directly from

$$\begin{aligned} dy_i\, dy_j = {} & h_i(\mathbf{x}, t)\, dt\, dy_j + h_j(\mathbf{x}, t)\, dt\, dy_i \\ & + h_i(\mathbf{x}, t) h_j(\mathbf{x}, t)\, dt\, dt + dw_i\, dw_j \end{aligned} \tag{3.31}$$

The first three terms are $o(dt)$ terms, but the last term is $R_{ij}\, dt$ with probability one. Thus, as stated, it is not $o(dt)$. Therefore, the expansion of $R(d\mathbf{y},dt,\mathbf{u})$ must include these terms. $R(d\mathbf{y},dt,\mathbf{u})$ can be written as

$$\begin{aligned} R(d\mathbf{y}, dt, \mathbf{u}) = {} & R\Big|_{d\mathbf{y},dt=0} + \sum_{i=1}^{m} \frac{\partial R}{dy_i}\Big|_{d\mathbf{y},dt=0} dy_i + \frac{\partial R}{\partial(dt)}\, dt \\ & + \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \frac{\partial^2 R}{\partial(dy_i)(dy_j)}\Big|_{d\mathbf{y}=dt=0} R_{ij} dt + o(dt) \end{aligned} \tag{3.32}$$

Thus, to $o(dt)$ we must evaluate the $m^2 + m + 1$ partial derivatives of R. Before doing so, we make one further simplification. Note that

$$p_{d\mathbf{y}}(d\mathbf{y}|\mathbf{x}(t) = \mathbf{u}) = \frac{1}{(2\pi)^{m/2}|\mathbf{R}|^{1/2}} \exp\left\{ -\frac{1}{2}[d\mathbf{y} - \mathbf{h}(\mathbf{u}, t)\,dt]^T \frac{\mathbf{R}^{-1}}{dt}[d\mathbf{y} - \mathbf{h}(\mathbf{u}, t)dt]\right\} \quad (3.33)$$

This follows from the fact that $d\mathbf{y}$ conditioned on $\mathbf{x}(t) = \mathbf{u}$ is a Wiener process (i.e., Gaussian) with mean $\mathbf{h}(\mathbf{u}, t)dt$ and covariance $\mathbf{R}\,dt$. This $R(d\mathbf{y}, dt, \mathbf{u})$ can be written as

$$R(d\mathbf{y}, dt, \mathbf{u}) = \frac{\exp\left\{\frac{1}{2}[2d\mathbf{y}^T\mathbf{R}^{-1}\mathbf{h}(\mathbf{u}, t) - \mathbf{h}^T(\mathbf{u}, t)\mathbf{R}^{-1}\mathbf{h}(\mathbf{u}, t)\,dt]\right\}}{E\left[\exp\left\{\frac{1}{2}[2d\mathbf{y}^T\mathbf{R}^{-1}\mathbf{h}(\mathbf{x}, t) - \mathbf{h}^T(\mathbf{x}, t)\mathbf{R}^{-1}\mathbf{h}(\mathbf{x}, t)\,dt]\right\}\right]} \quad (3.34)$$

where the expectation operator in the denominator $E[\ ]$ is the expectation of the expression with respect to the density $p_{\mathbf{x}}(\mathbf{u}, t|O_{t_0,t})$. With this expression we immediately note that

$$R(d\mathbf{y}, dt, \mathbf{u})\big|_{d\mathbf{y}, dt = 0} = 1 \quad (3.35)$$

To simplify the following analysis, introduce the function

$$T(d\mathbf{y}, dt, \mathbf{u}) = \exp\left\{\frac{1}{2}[2d\mathbf{y}^T\mathbf{R}^{-1}\mathbf{h}(\mathbf{u}, t) - \mathbf{h}^T(\mathbf{u}, t)\mathbf{R}^{-1}\mathbf{h}(\mathbf{u}, t)\,dt]\right\} \quad (3.36)$$

Then

$$R(d\mathbf{y}, dt, \mathbf{u}) = \frac{T(d\mathbf{y}, dt, \mathbf{u})}{E[T(d\mathbf{y}, dt, \mathbf{x}(t))]} \quad (3.37)$$

The first partial derivation can be obtained easily using this substitution. That is

$$\frac{\partial R(d\mathbf{y}, dt, \mathbf{u})}{\partial(dt)} = \frac{\partial T}{\partial(dt)}\frac{1}{E[T]} - \frac{T}{[E[T]]^2}\frac{\partial E[T]}{\partial(dt)} \quad (3.38)$$

But

$$\frac{\partial T}{\partial(dt)} = -\frac{1}{2}\mathbf{h}^T(\mathbf{u}, t)\mathbf{R}^{-1}\mathbf{h}(\mathbf{u}, t)T \quad (3.39)$$

Thus,

$$\frac{\partial R(d\mathbf{y}, dt, \mathbf{u})}{\partial(dt)} = -\frac{1}{2}\mathbf{h}^T(\mathbf{u}, t)\mathbf{R}^{-1}\mathbf{h}(\mathbf{u}, t)\frac{T}{E[T]} + \frac{T}{[E[T]]^2}E\left[\frac{1}{2}\mathbf{h}^T(\mathbf{x}, t)\mathbf{R}^{-1}\mathbf{h}(\mathbf{x}, t)T\right] \quad (3.40)$$

Taking the limit dt, $d\mathbf{y}$ equal to zero yields

$$\frac{\partial R(d\mathbf{y}, dt, \mathbf{u})}{\partial(dt)} = -\frac{1}{2}\mathbf{h}^T(\mathbf{u}, t)\mathbf{R}^{-1}\mathbf{h}(\mathbf{u}, t) + \frac{1}{2}E[\mathbf{h}^T(\mathbf{x}, t)\mathbf{R}^{-1}\mathbf{h}(\mathbf{x}, t)] \quad (3.41)$$

Do the same for $\partial R(d\mathbf{y}, dt, \mathbf{u})/\partial(dy_i)$. Thus, if we let

$$\mathbf{R}^{-1} = \begin{bmatrix} r^{-1}_{1,1} & \cdots & r^{-1}_{1,m} \\ r^{-1}_{m,1} & \cdots & r^{-1}_{m,m} \end{bmatrix} \quad (3.42)$$

be the inverse of $\mathbf{R}$, we have upon taking the prescribed limits

$$\left.\frac{\partial R(d\mathbf{y}, dt, \mathbf{u})}{\partial(dy_i)}\right|_{dt, d\mathbf{y}=0} = \sum_{j=1}^{m} r^{-1}_{j,i} h_j(\mathbf{u}, t) - E\left[\sum_{j=1}^{m} r^{-1}_{j,i} h_j(\mathbf{x}, t)\right] \quad (3.43)$$

But what was sought was this summed over the dy_i. This yields

$$\sum_{i=1}^{m}\sum_{i=1}^{m} dy_i r^{-1}_{j,i}[h_j(\mathbf{u}, t) - E[h_j(\mathbf{x}, t)]]$$
$$= d\mathbf{y}^T\mathbf{R}^{-1}[\mathbf{h}(\mathbf{u}, t) - E[\mathbf{h}(\mathbf{x}, t)]] \quad (3.44)$$

A similar procedure can now be applied to the derivative

$$\frac{\partial^2 R(d\mathbf{y}, dt, \mathbf{u})}{\partial(dy_i)\,\partial(dy_j)}$$

From the previous analysis we note that

$$\frac{\partial R(d\mathbf{y}, dt, \mathbf{u})}{\partial(dy_i)} = \sum_{j=1}^{m} r^{-1}_{j,i} h_j(\mathbf{u}, t)\,\frac{T}{E[T]} - E\left[\sum_{j=1}^{m} r^{-1}_{j,i} h_j(\mathbf{u}, t)T\right]\frac{T}{[E[T]]^2} \quad (3.45)$$

To simplify the analysis let us identify r_i by

$$r_i = \sum_{j=1}^{m} r^{-1}_{j,i} h_j(\mathbf{u}, t) \quad (3.46)$$

Taking the remaining derivative of the expression yields

$$\frac{\partial^2 R(d\mathbf{y}, dt, \mathbf{u})}{\partial(dy_j)\,\partial(dy_i)} = r_i r_j \frac{T}{E[T]} - r_i E[r_j T]\frac{T}{[E[T]]^2} - E[r_i r_j T]\frac{T}{E[T]^2} - E[r_i T] r_j \frac{T}{[E[T]]^2} + \frac{2E[r_i T]TE[r_j T]}{E[[T]]^3} \quad (3.47)$$

In the limit as dt, $d\mathbf{y} \to 0$, we obtain

$$\frac{\partial^2 R(d\mathbf{y}, dt, \mathbf{u})}{\partial(dy_i)\,\partial(dy_i)} = r_i r_j - r_i E[r_j] - r_j E[r_i] - E[r_j r_i] + 2E[r_i]E[r_j] \quad (3.48)$$

Using this in the summation expression, we obtain

$$\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\partial^2 R(d\mathbf{y}, dt, \mathbf{u})}{\partial(dy_i)\,\partial(dy_j)}\,dy_i\,dy_j = \sum_{i=1}^{m}\sum_{j=1}^{m} dy_i\,r_i r_j\,dy_j$$
$$- \sum_{i=1}^{m}\sum_{j=1}^{m} dy_i\,r_i E[r_j]\,dy_j - \sum_{i=1}^{m}\sum_{j=1}^{m} dy_i\,E[r_i] r_j\,dy_j$$
$$+ 2\sum_{i=1}^{m}\sum_{j=1}^{m} dy_i\,E[r_i]\,E[r_j]\,dy_j - \sum_{i=1}^{m}\sum_{j=1}^{m} dy_i\,E[r_i r_j]\,dy_j \tag{3.49}$$

The first four sums are readily identifiable from their definitions as matrix dot products. The last sum is

$$\sum_{i=1}^{m}\sum_{j=1}^{m} dy_i\,E\left[\sum_{k=1}^{m}\sum_{l=1}^{m} r_{k,i}^{-1}\,h_i(\mathbf{x}, t) h_l(\mathbf{x}, t) r_{j,l}^{-1}\right] dy_j$$
$$= d\mathbf{y}^T\mathbf{R}^{-1}E[\mathbf{h}(\mathbf{x}, t)\mathbf{h}^T(\mathbf{x}, t)]\mathbf{R}^{-1}\,d\mathbf{y}$$
$$= E[\mathbf{h}^T(\mathbf{x}, t)\mathbf{R}^{-1}d\mathbf{y}\;d\mathbf{y}^T\mathbf{R}^{-1}\mathbf{h}(\mathbf{x}, t)] \tag{3.50}$$

which follows upon taking the transpose of the scalar quantity. Thus, we obtain for the entire sum

$$\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\partial^2 R(d\mathbf{y}, dt, \mathbf{u})}{\partial(dy_i)\,\partial(dy_j)}\,dy_i\,dy_j$$
$$= \mathbf{h}^T(\mathbf{u}, t)\mathbf{R}^{-1}d\mathbf{y}\;d\mathbf{y}^T\mathbf{R}^{-1}\mathbf{h}(\mathbf{u}, t)$$
$$- \mathbf{h}^T(\mathbf{u}, t)\mathbf{R}^{-1}d\mathbf{y}\;d\mathbf{y}^T\mathbf{R}^{-1}E[\mathbf{h}(\mathbf{x}, t)]$$
$$- E[\mathbf{h}^T(\mathbf{x}, t)]\mathbf{R}^{-1}d\mathbf{y}\;d\mathbf{y}^T\mathbf{R}^{-1}\mathbf{h}(\mathbf{u}, t)$$
$$+ E[\mathbf{h}^T(\mathbf{x}, t)]\mathbf{R}^{-1}d\mathbf{y}\;d\mathbf{y}^T\mathbf{R}^{-1}E[\mathbf{h}(\mathbf{x}, t)]$$
$$- E[\mathbf{h}^T(\mathbf{x}, t)\mathbf{R}^{-1}d\mathbf{y}\;d\mathbf{y}^T\mathbf{R}^{-1}\mathbf{h}(\mathbf{x}, t)] \tag{3.51}$$

But recall that

$$d\mathbf{y}\,d\mathbf{y}^T = \mathbf{R}\,dt; \qquad \text{w.p.1} \tag{3.52}$$

Using this in the above and then using that in the expansion for $R(d\mathbf{y},dt,\mathbf{u})$, we obtain

$$R(d\mathbf{y}, dt, \mathbf{u}) = 1 - \frac{1}{2}\mathbf{h}^T(\mathbf{u}, t)\mathbf{R}^{-1}\mathbf{h}(\mathbf{u}, t)\,dt$$
$$+ \frac{1}{2}E[\mathbf{h}^T(\mathbf{x}, t)\mathbf{R}^{-1}\mathbf{h}(\mathbf{x}, t)]$$
$$+ d\mathbf{y}^T\mathbf{R}^{-1}[\mathbf{h}(\mathbf{u}, t) - E[\mathbf{h}(\mathbf{x}, t)]]$$
$$+ \frac{1}{2}\mathbf{h}^T(\mathbf{u}, t)\mathbf{R}^{-1}\mathbf{h}(\mathbf{u}, t)\,dt$$
$$- \mathbf{h}^T(\mathbf{u}, t)\mathbf{R}^{-1}E[\mathbf{h}(\mathbf{x}, t)]\,dt$$
$$+ E[\mathbf{h}^T(\mathbf{x}, t)]\mathbf{R}^{-1}E[\mathbf{h}(\mathbf{x}, t)]\,dt$$
$$- \frac{1}{2}E[\mathbf{h}^T(\mathbf{x}, t)\mathbf{R}^{-1}\mathbf{h}(\mathbf{x}, t)]\,dt$$
$$+ o(dt) \tag{3.53}$$

Upon canceling like terms and rearranging, we obtain

$$R(d\mathbf{y}, dt, \mathbf{u}) = 1 + [d\mathbf{y} - E[\mathbf{h}(\mathbf{x}, t)]\, dt]^T \mathbf{R}^{-1} [\mathbf{h}(\mathbf{u}, t) - E[\mathbf{h}(\mathbf{x}, t)]] + o(dt) \tag{3.54}$$

Using this in (3.29), the lemma is proved. ∎

Having evaluated $q(d\mathbf{y}, dt, \mathbf{u})$ for the additive Gaussian case, we can now combine it with the preceding theorem to present the following corollary.

COROLLARY 3.1. (Kushner-Stratonovich equation). Let $\mathbf{x}(t)$ be an $(n \times 1)$-vector Markov process generated by

$$d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}, t)\, dt + d\mathbf{n}(t) \tag{3.55}$$

and let the $(m \times 1)$-vector measurement process be given by

$$d\mathbf{y} = \mathbf{h}(\mathbf{x}, t)\, dt + d\mathbf{w} \tag{3.56}$$

Then $p = p_{\mathbf{x}}(\mathbf{u}, t \,|\, O_{t_0,t})$ is given by the following partial-differential integral equation;

$$\frac{\partial p}{\partial t} = L^+ p + \left[\frac{d\mathbf{y}}{dt} - E[\mathbf{h}(\mathbf{x}, t)]\right]^T \mathbf{R}^{-1}(t) \left[\mathbf{h}(\mathbf{u}, t) - E[\mathbf{h}(\mathbf{x}, t)]\right] p \tag{3.57}$$

where the L^+ operator is

$$L^+ = -\sum_{i=1}^{n} \frac{\partial(f_i(\,))}{\partial u_i} + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2(Q_{ij}(\,))}{\partial u_i\, \partial u_j} + \sum_{i=1}^{n} \lambda_i(t)[-(\,) + p_{a_i^2}(\,)] \tag{3.58}$$

and

$$E[\mathbf{h}(\mathbf{x}, t)] = \int \mathbf{h}(\mathbf{u}, t) p_{\mathbf{x}}(\mathbf{u}, t \,|\, O_{t_0,t})\, d\mathbf{u} \tag{3.59}$$

This theorem provides the basic building block to all linear and nonlinear estimation schemes. We shall devote a great deal of Chapter 6 to methods that attempt to solve this nonlinear partial-differential integral equation. Except for a small number of special, albeit important, cases solutions are relatively unknown and virtually incapable of being known. Arguments on the existence and uniqueness of solutions, on the other hand, are known, and the reader is referred to Duncan [1] for the results. The following examples provide us with several uses for this approach.

Example. Let us assume that we want to estimate the random process given by

$$\dot{x}_1(t) = -ax_1(t) + w_1(t) \tag{3.60}$$

This process is measured through a device that intermittently fails to give an output. This process can be modeled as a Poisson process with two discrete amplitudes (+ 1 and 0). To obtain this, let $n_p(t)$ be a Poisson process and define

Figure 5.4 Model of a process sampled through a loose relay.

$$\dot{x}_2 = \frac{dn_p(t)}{dt} \tag{3.61}$$

Now $n_p(t)$ is a generalized Poisson process, here limited to jumps in amplitude of $+1$ and -1. The derivative of such a process is a set of impulses that are ± 1 at the arrival times of the process. $x_2(t)$ is merely the process $n_p(t)$. We shall assume that $x_2(t)$ has an initial condition of 0. These processes are graphically shown in Figure 5.4.

We observe $z(t)$, which is

$$z(t) = g(x_2(t))x_1(t) + w_2(t) \tag{3.62}$$

over the time interval $(0, t)$, and the nonlinearity is defined as

$$g(x) = \begin{cases} 1; & x \geqq 0 \\ 0; & x < 0 \end{cases} \tag{3.63}$$

Both $w_1(t)$ and $w_2(t)$ are white noise processes. $w_1(t)$ $w_2(t)$, and $n_p(t)$ are all independent. This model may represent the measurement of the processes $x_1(t)$ through a loose contact represented by $x_2(t)$. In state form $\mathbf{x}$ is given by the equations

$$\dot{\mathbf{x}} = \begin{bmatrix} -a & 0 \\ 0 & 0 \end{bmatrix} \mathbf{x} + \begin{bmatrix} w_1(t) \\ \dot{n}_p(t) \end{bmatrix} \tag{3.64}$$

where

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{3.65}$$

The measurments, therefore, are equal to

$$z(t) = h(\mathbf{x}, t) + w_2(t) \tag{3.66}$$

where h is the product form shown. The propagation equation for the conditional density $p = p_x(u_1, u_2, t \mid O_t)$ is

$$\frac{\partial p}{\partial t} = (z - \overline{x_1 g(x_2)})(u_1 g(u_2) - \overline{x_1 g(x_2)})p + \frac{\partial}{\partial u_1}(au_1 p) + \frac{1}{2}\frac{\partial^2}{\partial u_1^2} p$$
$$+ \lambda(t)\int \delta(u_1 - \xi_1)\left[\frac{1}{2}\delta(u_2 - \xi_2 + 1) + \frac{1}{2}\delta(u_2 - \xi_2 - 1)\right]$$
$$p(\xi_1, \xi_2, t \mid O_t)d\xi_1\, d\xi_2 - \lambda(t)p \tag{3.67}$$

where O_t is $O_{0,t}$, the minimum sub σ-field generated by the observation.

Since

$$\mathbf{Q} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \tag{3.68}$$

$$S = 0 \tag{3.69}$$

$$R = I \tag{3.70}$$

and

$$\lambda = \lambda(t)$$

is a time-invariant rate and where

$$\overline{x_1 x_2} = \int p_x(u_1, u_2; t \,|O_t) u_1 u_2 \, du_1 \, du_2 \tag{3.71}$$

Note that we have used impulses for the density function of the amplitude of the generalized Poisson portion. Now, simplifying this equation, we obtain

$$\begin{aligned} \frac{\partial p}{\partial t} &= \frac{\partial}{\partial u_1}(a u_1 p) + \frac{1}{2}\frac{\partial^2}{\partial u_1^2} \\ &+ \frac{\lambda}{2}\left(p_x(u_1, u_2 - 1; t \,|O_t) - p_x(u_1, u_2 + 1; t \,|O_t)\right) \\ &+ \left(z - \overline{x_1 g(x_2)}\right)\left(u_1 g(x_2) - \overline{x_1 g(x_2)}\right)p \end{aligned} \tag{3.72}$$

This equation is almost impossible to solve. To get $\dot{\hat{x}}_1$ we would multiply by u_1 and integrate over u_1 and u_2. This would follow the procedure outlined in Chapter 6. Under steady-state conditions, $\partial p/\partial t = 0$, and a possible solution may exist.

Example. This example comes from the field of aerospace instrumentation (McGarty [1]). We want to design a device that will obtain an estimate of a light intensity scanning across its surface. For example, we may have a rotating cylinder within which is contained optics and a photo detector. The device is scanning at a constant rate. The light source is a star that is seen through the turbulent upper atmosphere of the earth (see Figure 5.5). The light arrives at the detector in the form of photons whose average arrival rate is proportional to the intensity of light observed. The photons act as impulses exciting the photodiode. $\lambda(t)$ is the arrival rate of the photons. The photodiode can be modeled as a first-order dynamical system (a resistor and capacitor circuit). The output current of the photodiode is given by $x_1(t)$ and is

$$\dot{x}_1(t) = -a_1 x_1(t) + \frac{dn_p(t)}{dt} \tag{3.73}$$

Where $n_p(t)$ is a generalized Poisson process of rate $\lambda(t)$. The photodiode, because of impurities, does not equally weight the incoming photons. This may be modeled by considering that the amplitudes of the photon impulses are random with a density $p(\alpha_1)$. A good approximation is to assume

$$p(\alpha_1) = \beta e^{-\beta \alpha_1}; \qquad \alpha_1 \geqq 0 \tag{3.74}$$

Clearly, $x_1(t)$ is always positive as one would expect.

Figure 5.5 Star-scanning system with nonlinear measurement

The turbulence can be modeled as a square of a Gaussian process. Let

$$\dot{x}_2(t) = -a_2 x_2 + n_g(t) \tag{3.75}$$

Then the output of the system can be given by

$$z(t) = x_2^2(t)x_1(t) + w(t) \tag{3.76}$$

where we have accounted for the turbulence by multiplying the photodiode output by the turbulence squared. We assume that all the noises are independent. Now we assume

$$\mathbf{Q} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \tag{3.77}$$

$$S = 0 \tag{3.78}$$

$$R = 1 \tag{3.79}$$

and $p(u_1)$ is unspecified. Then the conditional density propagation equations become

$$\frac{\partial p}{\partial t} = (z - u_2^2 u_1)(u_2^2 u_1 - \hat{x}_2^2 \hat{x}_1)p + \frac{\partial(a_1 u_1 p)}{\partial \hat{x}_1} + \frac{\partial(a_2 u_2 p)}{\partial \hat{x}_2} + \frac{\partial^2}{\partial \hat{x}_2^2} p + \lambda(t) \int \delta(u_2 - \xi_2) p(u_1 - \xi_1) p \, d\xi_1 \, d\xi_2 - \lambda(t) p \tag{3.80}$$

It should again be obvious that any further progress would be futile. It is again necessary to make approximations in order to obtain a tractable solution.

Having discussed the additive Gaussian case, we proceed with the unit-jump Poisson case. The following lemma evaluates $q(d\mathbf{N}, dt, \mathbf{u})$ for that case.

LEMMA. 3.2. Let $\mathbf{x}(t)$ be an $(n \times 1)$-vector Markov process generated by

$$d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}, t)\, dt + d\mathbf{n}(t) \tag{3.81}$$

and let $d\mathbf{N}(t)$ be an $(m \times 1)$-vector Poisson step process with an $m \times 1$ rate parameter $\boldsymbol{\lambda}(\mathbf{x}(t), t)$. Let $\lambda_i(\mathbf{x}(t), t)$ be the ith component of $\boldsymbol{\lambda}(\mathbf{x}(t), t)$. Let $p_{\mathbf{x}}(\mathbf{u}, t \,|\, O_{t_0, t+dt})$ be the conditional probability density function of the process $\mathbf{x}(t)$ given $O_{t_0, t+dt}$, the minimum σ-field generated by $\mathbf{N}(s)$, $s \in [t_0, t + dt)$. Then to order dt

$$p_{\mathbf{x}}(\mathbf{u}, t \,|\, O_{t_0, t+dt}) = p_{\mathbf{x}}(\mathbf{u}, t \,|\, O_{t_0, t})\,[1 + dq(d\mathbf{N}, dt, \mathbf{u})] \tag{3.82}$$

where

$$dq(d\mathbf{N}, dt, \mathbf{u}) = \sum_{i=1}^{m} [\lambda_i(\mathbf{u}, t) - E[\lambda_i(\mathbf{x}, t)]]\,[E[\lambda_i(\mathbf{x}, t)]]^{-1} [dN_i(t) - E[\lambda_i(\mathbf{x}, t)]\, dt] \tag{3.83}$$

and where

$$E[\lambda_i(\mathbf{x}, t)] = \int \lambda_i(\mathbf{u}, t) p_{\mathbf{x}}(\mathbf{u}, t \,|\, O_{t_0, t})\, d\mathbf{u} \tag{3.84}$$

Proof. As with Lemma 3. we shall begin the proof with Bayes's rule. First we again assume that $O_{t_0, t+dt} = O_{t_0, t} \bigcup d\mathbf{N}(t)$. For the Poisson step process, this equivalence is easier to show than in the Gaussian case, and we refer the interested reader to J. R. Clark for the proof.

Thus,

$$p_{\mathbf{x}}(\mathbf{u}, t \,|\, O_{t_0, t+dt}) = p_{\mathbf{x}}(\mathbf{u}, t \,|\, O_{t_0, t}, d\mathbf{N}(t)) \tag{3.85}$$

Now $d\mathbf{N}(t)$ can take on only the values 0 or γ_i, where

$$\gamma_i = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \begin{matrix} 1 \\ \vdots \\ i \\ \vdots \\ m \end{matrix} \tag{3.86}$$

That is, the probability that $dN_i(t)$ *and* $dN_j(t)$ are both 1, with all others being 0, is $o(dt)$. Thus, $d\mathbf{N}(t)$ is limited to the $m \times 1$ values, $\mathbf{0}$, $\gamma_1, \cdots, \gamma_m$. Since $d\mathbf{N}(t)$ is a discrete random process, we write for Bayes's rule

$$p_{\mathbf{x}}(\mathbf{u}, t \,|O_{t_0 t}, d\mathbf{N}) = \frac{P[d\mathbf{N}|O_{t_0 t}, \mathbf{x}(t) = \mathbf{u}]}{P[d\mathbf{N}|O_{t_0 t}]} p_{\mathbf{x}}(\mathbf{u}, t|O_{t_0 t}) \tag{3.87}$$

where $P[d\mathbf{N}|O_{t_0 t}, \mathbf{x}(t) = \mathbf{u}]$ is the probability that $d\mathbf{N}$ takes on its prescribed value given $O_{t_0 t}$, and $\mathbf{x}(t)$. Again $d\mathbf{N}$ conditioned on both $O_{t_0 t}$, and $\mathbf{x}(t)$ depends only on $\mathbf{x}(t)$. Furthermore, the denominator may be written as

$$\begin{aligned} P[d\mathbf{N}|O_{t_0 t}] &= \int P[d\mathbf{N}|O_{t_0 t}, \mathbf{x}(t) = \mathbf{v}] p_{\mathbf{x}}(\mathbf{v}, t \,|O_{t_0 t}) \, d\mathbf{v} \\ &= E[\, P[d\mathbf{N}|\mathbf{x}(t)]\,] \end{aligned} \tag{3.88}$$

where as before the expectation operator $E[\ ]$ is over the conditional density. Now, for an arbitrary $\gamma_i \neq \mathbf{0}$,

$$P[d\mathbf{N} = \gamma_i|\mathbf{x}(t) = \mathbf{u}] = \lambda_i(\mathbf{u}, t)\, dt \prod_{j \neq i}^{m} (1 - \lambda_j(\mathbf{u}, t)\, dt) = \lambda_i(\mathbf{u}, t)\, dt \tag{3.89}$$

For the case $d\mathbf{N}(t) = \mathbf{0}$

$$P[d\mathbf{N} = \mathbf{0}|\mathbf{x}(t) = \mathbf{u}] = \prod_{i=1}^{m} (1 - \lambda_i(\mathbf{u}, t)\, dt) = 1 - \sum_{i=1}^{m} \lambda_i(\mathbf{u}, t)\, dt \tag{3.90}$$

Now this can be formalized if we introduce the vector γ, where

$$\gamma = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix} \tag{3.91}$$

Then

$$\begin{aligned} P[d\mathbf{N}|\mathbf{x}(t) = \mathbf{u}] &= \lambda^T(\mathbf{u}, t) d\mathbf{N}\, dt \\ &\quad + (1 - \lambda^T(\mathbf{u}, t)\gamma\, dt)(1 - d\mathbf{N}^T\gamma) \end{aligned} \tag{3.92}$$

Clearly, upon substitution this satisfies all possible γ_i and $\mathbf{0}$ cases for $d\mathbf{N}$. Thus, using this in the Bayes formula, we obtain

$$p_{\mathbf{x}}(\mathbf{u}, t \,|O_{t_0, t+dt})$$
$$= \frac{\boldsymbol{\lambda}^T(\mathbf{u}, t)d\mathbf{N}\, dt + (1 - \boldsymbol{\lambda}^T(\mathbf{u}, t)\boldsymbol{\gamma}\, dt)(1 - d\mathbf{N}^T\boldsymbol{\gamma})}{[E[\boldsymbol{\lambda}^T(\mathbf{x}, t)]d\mathbf{N}\, dt + (1 - E[\boldsymbol{\lambda}^T(\mathbf{x}, t)]\boldsymbol{\gamma}\, dt)(1 - d\mathbf{N}^T\boldsymbol{\gamma})]} p_{\mathbf{x}}(\mathbf{u}, t \,|O_{t_0, t}) \tag{3.93}$$

Now the expression to the left of $p_{\mathbf{x}}(\mathbf{u}, t \,|O_{t_0,t})$ is to be expanded to $o(dt)$ terms. To do this, we again take advantage of the fact that $d\mathbf{N}$ can be only $\mathbf{0}$ or $\boldsymbol{\gamma}_i$. Thus,

$$\frac{\boldsymbol{\lambda}^T(\mathbf{u}, t)d\mathbf{N}\, dt + (1 - \boldsymbol{\lambda}^T(\mathbf{u}, t)\boldsymbol{\gamma}\, dt)(1 - d\mathbf{N}^T\boldsymbol{\gamma})}{E[\boldsymbol{\lambda}^T(\mathbf{x}, t)]d\mathbf{N}\, dt + (1 - E[\boldsymbol{\lambda}^T(\mathbf{x}, t)]\boldsymbol{\gamma}\, dt)(1 - d\mathbf{N}^T\boldsymbol{\gamma})}$$
$$= \begin{cases} \dfrac{1 - \boldsymbol{\lambda}^T(\mathbf{u}, t)\boldsymbol{\gamma}\, dt}{1 - E[\boldsymbol{\lambda}^T(\mathbf{x}, t)]\boldsymbol{\gamma}\, dt} & d\mathbf{N} = \mathbf{0} \\ \dfrac{\boldsymbol{\lambda}^T(\mathbf{u}, t)\boldsymbol{\gamma}_i}{E[\boldsymbol{\lambda}^T(\mathbf{x}, t)]\boldsymbol{\gamma}_i} & d\mathbf{N} = \boldsymbol{\gamma}_i \end{cases}$$
$$= \frac{1 - \boldsymbol{\lambda}^T(\mathbf{u}, t)\boldsymbol{\gamma}\, dt}{1 - E[\boldsymbol{\lambda}^T(\mathbf{x}, t)]\boldsymbol{\gamma}\, dt}(1 - d\mathbf{N}^T\boldsymbol{\gamma})$$
$$+ \sum_{i=1}^{m} \frac{\boldsymbol{\lambda}^T(\mathbf{u}, t)\boldsymbol{\gamma}_i}{E[\boldsymbol{\lambda}^T(\mathbf{x}, t)]\boldsymbol{\gamma}_i} d\mathbf{N}^T\boldsymbol{\gamma}_i \tag{3.94}$$

The equivalence of the above expressions should be clear upon substitution. Now, expanding the first expression on the right in the final equality, we obtain for (3.94)

$$= 1 + \left[\sum_{i=1}^{m} (E[\lambda_i(\mathbf{x}, t)] - \lambda_i(\mathbf{u}, t))\right](1 - d\mathbf{N}^T\boldsymbol{\gamma})$$
$$- d\mathbf{N}^T\boldsymbol{\gamma} + \sum_{i=1}^{m} \frac{\lambda_i^T(\mathbf{u}, t)}{E[\lambda_i(\mathbf{x}, t)]} dN_i \tag{3.95}$$

Upon rearranging, we obtain

$$= 1 + \sum_{i=1}^{m} [\lambda_i(\mathbf{u}, t) - E[\lambda_i(\mathbf{x}, t)]][E[\lambda_i(\mathbf{u}, t)]]^{-1}$$
$$[dN_i(t) - E[\lambda_i(\mathbf{u}, t)]\, dt] \tag{3.96}$$

which proves the lemma. ∎

We can now use this lemma and the theorem to present Snyder's equation for the propagation of the conditional probability density.

COROLLARY. 3.2. (Snyder's equation). Let $\mathbf{x}(t)$ be an $(n \times 1)$-vector Markov process generated by

$$d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}, t)\, dt + d\mathbf{n}(t) \tag{3.97}$$

and let the $(m \times 1)$-vector measurement process $d\mathbf{N}(t)$ be a unit jump Poisson process with an $(m \times 1)$-vector rate parameter $\boldsymbol{\lambda}^*(t)$—to distinguish it

from the rate parameter $\lambda(t)$ of the state equation. Then $p = p_{\mathbf{x}}(\mathbf{u}, t \,|O_{t_0 t})$ is given by the following partial differential integral equation

$$\frac{\partial p}{\partial t} = L^{+} p + \sum_{i=1}^{m} [\lambda_i^*(\mathbf{u}, t) - E[\lambda_i^*(\mathbf{x}, t)]][E[\lambda_i^*(\mathbf{x}, t)]]^{-1} \left[\frac{dN_i(t)}{dt} - E[\lambda_i^*(\mathbf{x}, t)]\right] \tag{3.98}$$

where L^{+} is the characteristic operator and

$$E[\lambda_i^*(\mathbf{x}, t)] = \int \lambda_i^*(\mathbf{u}, t) p_{\mathbf{x}}(\mathbf{u}, t \,|O_{t_0 t}) \, d\mathbf{u} \tag{3.99}$$

The preceding two corollaries present the propagation equations that must be solved in order to evaluate the conditional density of the state, given the information. The following two examples obtain numerical solutions to these equations for two specific problems. The first example is for a nonlinear system with linear additive Gaussian measurements and is from Culver [1]. The second example is a linear system with the measurements being unit-jump Poisson processes. Two forms of arrival rates are considered. This example was first presented by Snyder [4]. In both examples we evaluate $p_{\mathbf{x}}(\mathbf{u}, t \,|O_{t_0 t})$ as a function of u for different t.

Example. A dynamical system is given by the state equation

$$\frac{dx}{dt} = -0.7x(t) \tag{3.100}$$

where $x(0)$ is a Gaussian random variable with mean zero and variance 0.16. The measurement process is

$$z(t) = x(t) + \dot{w}(t) \tag{3.101}$$

where $\dot{w}(t)$ is white Gaussian noise with spectral height 0.05. A block diagram is shown in Figure 5.6. The propagation equation for the system is

Figure 5.6 Linear system–linear measurement.

$$\frac{\partial p}{\partial t} = \frac{\partial(0.7up)}{\partial u} + [z(t) - \hat{x}(t)]20[u - \hat{x}(t)]p \tag{3.102}$$

where $\hat{x}(t)$ is the $E[x(t)|\mathcal{Y}_{t_0,t}]$ and where $p = p_x(u, t|\mathcal{Y}_{0,t})$, (i.e., $t_0 = 0$). Using Monte Carlo techniques, Culver [1] has obtained p for this case. The results are shown in Figure 5.7. Note that initially at $t = 0$, the density is the a priori density associated with $x(t)$ at $t = 0$. Then, as time passes, two things occur. The mean shifts toward zero as a result of the fact that the system trajectory is taking it there. Second, the variance decreases as a result of two factors. The first factor is that the system is known to decay to zero w.p.1, so eventually the variance must go to zero. The second fact is that the measurements are adding information that makes the variance even smaller. Thus, the curve

Figure 5.7 Conditional PDF for linear system with linear measurement (from Culver [1]).

Figure 5.8 Nonlinear system--linear measurement.

Figure 5.9 Conditional PDF for nonlinear system with linear measurement (from Culver[1]).

shows a translation of the mean and a sharpening of the peak. Finally, it should be noted that the system is linear so that $p_x(u, t|O_{0,t})$ is Gaussian, which is what appears in the figure. In the next chapter we shall investigate this more fully.

We can now augment the system to include nonlinear dynamics. The system is given by

$$\dot{x} = -1.2x + 0.3x^2 \tag{3.103}$$

where $x(0)$ is again Gaussian with mean 2 and variance 0.16. The same measurement scheme is used in this system as with the linear systems (see Figure 5.8). For this case $p_x(u, t|O_{0,t})$ is shown in Figure 5.9. Note that as before at $t = 0$ the a priori Gaussian statistics are present. But now, in contrast to the linear case, as time progresses, the conditional density function becomes multimodal until at $t = 2.1$ there appear three distinct peaks to the

Figure 5.10 Poisson measurement; $\lambda(xt) = 100 \exp(-xt)$; linear state (from Synder [4]).

density. This fact will become extremely important in our discussion of approximate filtering strategies.

Example. In many cases of estimation the process does not evolve in time but is constant. This type of estimation is called *parameter estimation* and can be represented by the system equation

$$\frac{dx}{dt} = 0 \tag{3.104}$$

where now there are neither dynamics nor disturbances to the system. All that is specified is $x(0) = x_0$. Thus, $x(t) = x_0$, a random variable, for all t. Such a model implies that $f(x,t) = 0$ and that both dn_p and dn_g have zero arrival rate and covariance, respectively.

In an example considered by Snyder [4], the system was given as in (3.104),

Figure 5.11 Poisson measurement; $\lambda(x,t) = x$; linear state (from Synder [4]).

but the measurement was given by a Poisson counting process with rate parameter $\lambda(x,t)$. Snyder's problem arose in the biomedical field, but a similar problem arises in the meteorological field (see McGarty [3]). We shall consider two forms of $\lambda(x,t)$ where both λ and x are scalar. The first is for

$$\lambda(x, t) = 100 \exp(-xt)u_{-1}(t) \tag{3.105}$$

where $u_{-1}(t)$ is the unit-step function. For this case and for the assumption that x is initially distributed uniformly on [0.5,1.5] the resulting a posteriori density is shown in Figure 5.10. Here we note again that at $t = 0$ the a priori statistics are used. As time increases, a unimodal distribution begins to appear.

A second and simpler example is for the same state equation but for

$$\lambda(x, t) = x \tag{3.106}$$

The simulated results for this case are shown in Figure 5.11, where now $x(0)$ was initially assumed to be uniform over [1,2]. Again the density is unimodal.

These two examples are indicative of what is to be found for most dynamical systems. In general, the variance will decrease on the a posteriori density because of the information provided by the measurements despite the presence of system noise. This variance will generally be less than that obtained by using the Fokker-Planck or the Feller-Kolmogorov equations. It should also be clear that the Kushner-Stratonovich equation reduces to one or other of the equations as $\mathbf{R}$, the measurement noise covariance, is increased to infinity. In that case, the measurements become useless and are thus disregarded, and the a priori trajectory is followed. The next section will present an alternate approach to this analysis that is more rigorous and provides a different insight into the structure of the a posteriori statistics.

5.4. THE REPRESENTATION THEOREM

There are methods of obtaining the conditional density function of the state, given the observations, other than that of the propagation equation developed in the last section. Two of these methods are discussed in this section and both were initially proposed by Bucy. They are appropriately called integral techniques because the resulting structure of the conditional densities are expressed in integral form. The first theorem presents a recursive technique for evaluating the conditional density at a given time, given the density at some previous time. From this representation we develop expressions for the conditional mean and conditional variance. This theorem is for a discretized system that we spend the first part of this section discussing.

The second and third theorems are the discrete and continuous versions of Bucy's representation theorem. These theorems present the inherent

relationship between the conditional density given data and the unconditioned density evaluated by the Fokker-Planck equation.

The arguments used will be heuristic, and the interested reader is referred to Bucy and Joseph for a more complete measure theoretic discussion. The proofs also rely a great deal on the properties of martingales as discussed briefly in Chapters 3 and 4 and extensively surveyed by Doob.

The first step in the analysis is to discretize the continuous-time system. This will provide us with the simplification necessary to perform the conditioning on the observations first on a finite number and then obtain the limiting behaviors for the continuous case. We shall deal only in Gaussian disturbances to both the state and measurement systems.

Let $\mathbf{x}(t)$ be an $n \times 1$ state vector generated by

$$d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}(t), t)\, dt + d\mathbf{n}(t) \tag{4.1}$$

where $\mathbf{n}(t)$ is a Wiener process with

$$E[d\mathbf{n}(t)\, d\mathbf{n}^T(t)] = \mathbf{Q}(t)\, dt \tag{4.2}$$

(Note that we are suppressing the Gaussian subscript.) Now integrate $d\mathbf{x}(t)$ from $t = kT$ to $t = (k + 1)\, T$, where T is some small time interval. Then we obtain

$$\mathbf{x}((k + 1)T) - \mathbf{x}(kT) = \int_{kT}^{(k+1)T} \mathbf{f}(\mathbf{x}(t), t)\, dt + \mathbf{n}((k + 1)T) - \mathbf{n}(kT) \tag{4.3}$$

Assume that T is sufficiently small such that

$$\int_{Tk}^{(k+1)T} \mathbf{f}(\mathbf{x}(t), t)\, dt \cong \mathbf{f}(\mathbf{x}(kT), kT)T \tag{4.4}$$

Then define $\mathbf{f}(\mathbf{x}(k))$ as

$$\mathbf{f}(\mathbf{x}(k)) \triangleq \mathbf{x}(kT) + \mathbf{f}(\mathbf{x}(k), k)T \tag{4.5}$$

and $\mathbf{n}(k)$ as

$$\mathbf{n}(k) \triangleq \mathbf{n}((k + 1)T) - \mathbf{n}(kT) \tag{4.6}$$

Then clearly $\mathbf{n}(k)$ is zero mean Gaussian with covariance

$$E[\mathbf{n}(k)\mathbf{n}^T(k)] \triangleq \mathbf{Q}(k) = \mathbf{Q}(kT)T \tag{4.7}$$

assuming $\mathbf{Q}(kT) \approx \mathbf{Q}((k + 1)T)$. Thus, the discrete system model becomes

$$\mathbf{x}(k + 1) = \mathbf{f}(\mathbf{x}(k)) + \mathbf{n}(k) \tag{4.8}$$

Similarly, for the measurement model

$$d\mathbf{y}(t) = \mathbf{h}(\mathbf{x}(t), t)\, dt + d\mathbf{w}(t) \tag{4.9}$$

Discretizing this in a similar fashion, we obtain

$$\mathbf{y}(k+1)T) - \mathbf{y}(kT) = \mathbf{h}(\mathbf{x}(kT), kT)T + \mathbf{w}((k+1)T) - \mathbf{w}(kT) \tag{4.10}$$

Now define $\mathbf{z}(k)$ as

$$\mathbf{z}(k) = \frac{\mathbf{y}((k+1)T) - \mathbf{y}(kT)}{T} \tag{4.11}$$

and $w(k)$ as

$$\mathbf{w}(k) = \frac{\mathbf{w}((k+1)T) - \mathbf{w}(kT)}{T} \tag{4.12}$$

Then the discrete measurement equation is

$$\mathbf{z}(k) = \mathbf{h}(\mathbf{x}(k)) + \mathbf{w}(k) \tag{4.13}$$

where $\mathbf{w}(k)$ is a zero mean Gaussian random variable with covariance

$$E[\mathbf{w}(k)\mathbf{w}^T(k)] \triangleq \mathbf{R}(k) = \frac{\mathbf{R}(kT)}{T} \tag{4.14}$$

The most important fact to note about this discretization is the distinct difference in the covariances in the measurement noise and system noise. The difference lies in the fact that as $T \to 0$, $\mathbf{w}(k)$ becomes white noise, whereas $\mathbf{n}(k)$ goes to 0 with probability 1 because of continuity of the Wiener process. Thus, going in the reverse direction with this model requires some care in the taking of limits.

We can now prove the first theorem of this section, which provides us with a recursive scheme for calculating the conditional densities of the discrete process.

THEOREM 4.1

Let a dynamical system have a discrete version given by

$$\mathbf{x}(k+1) = \mathbf{f}(\mathbf{x}(k)) + \mathbf{n}(k) \tag{4.15}$$

and a discrete measurement given by

$$\mathbf{z}(k+1) = \mathbf{h}(\mathbf{x}(k+1)) + \mathbf{w}(k+1) \tag{4.16}$$

where k represents kT, T being a sampling interval. Let

$$\mathbf{z}(k+1) = \mathbf{y}(k+1) - \mathbf{y}(k) \tag{4.17}$$

with $\mathbf{y}(0) = \mathbf{0}$. Let the measurement be defined as

$$\mathcal{Y}_k = \{\mathbf{y}(k); k = 0, \cdots, kT = t\} \tag{4.18}$$

Then,

$$J_k(\mathbf{u}) = \frac{\left[\int \exp\left[\mathbf{z}^T(k)\mathbf{R}^{-1}\mathbf{h}(\mathbf{v}) - \frac{1}{2}\mathbf{h}^T(\mathbf{v})\mathbf{R}^{-1}\mathbf{h}(\mathbf{v})\right.\right.}{\left.\left.\exp\left(-\frac{1}{2}\{[\mathbf{u}-\mathbf{f}(\mathbf{v})]^T\mathbf{Q}^{-1}[\mathbf{u}-\mathbf{f}(\mathbf{v})]\}\right)J_{k-1}(\mathbf{v})\,d\mathbf{v}\right]}{(2\pi)^{n/2}|\mathbf{Q}|^{1/2}\int \exp[\mathbf{z}^T(\mathbf{k})\mathbf{R}^{-1}\mathbf{h}(\mathbf{v}) - \frac{1}{2}\mathbf{h}^T(\mathbf{v})\mathbf{R}^{-1}\mathbf{h}(\mathbf{u})]J_{k-1}(\mathbf{v})\,d\mathbf{v}} \tag{4.19}$$

where

$$J_k(\mathbf{u}) = p_{\mathbf{x}}(\mathbf{u}, k|\mathbf{z}(1)\cdots\mathbf{z}(k)) = p_{\mathbf{x}}(\mathbf{u}, k|O_k) \tag{4.20}$$

and

$$\mathbf{Q} = E[\mathbf{n}(k)\mathbf{n}^T(k)] \tag{4.21}$$

$$\mathbf{R} = E[\mathbf{w}(k)\mathbf{w}^T(k)] \tag{4.22}$$

Proof. By definition

$$p_{\mathbf{x}}(\mathbf{u}, k|O_k) = \frac{p_{\mathbf{x}}(\mathbf{u}, k; \mathbf{y}(k)|O_{k-1})}{p_{\mathbf{y}}(\mathbf{y}(k)|O_{k-1})} \tag{4.23}$$

This follows from the Bayes theorem approach used in the last section. Now we can write the numerator as

$$p_{\mathbf{x}}(\mathbf{u}, k; \mathbf{y}(k)|O_{k-1}) = \int p_{\mathbf{x}}(\mathbf{u}, k; \mathbf{v}, k-1; \mathbf{y}(k)|O_{k-1})\,d\mathbf{v} \tag{4.24}$$

where $p_{\mathbf{x}}(\mathbf{u}, k; \mathbf{v}, k-1; \mathbf{y}(k)|O_{k-1})$ is the joint probability density of $\mathbf{x}$ at times k and $k-1$. Writing this in terms of conditional probabilities yields

$$\begin{aligned} p_{\mathbf{x}}(\mathbf{u}, k; \mathbf{y}(k)|O_{k-1}) = \int & p_{\mathbf{y}}(\mathbf{y}(k)|\mathbf{x}(k) = \mathbf{u}, \mathbf{x}(k-1) = \mathbf{v}, O_{k-1}) \\ & p_{\mathbf{x}}(\mathbf{u}, k|\mathbf{x}(k-1) = \mathbf{v}, O_{k-1}) \\ & p_{\mathbf{x}}(\mathbf{v}, k-1|O_{k-1})\,d\mathbf{v} \end{aligned} \tag{4.25}$$

The first term on the right in the integral is the conditional density of $\mathbf{y}(k)$, given $\mathbf{x}(k-1)$ and O_{k-1}. But recall that

$$\mathbf{y}(k) = \mathbf{y}(k-1) + \mathbf{h}(\mathbf{x}(k-1)) + \mathbf{w}(k-1) \tag{4.26}$$

Thus, $\mathbf{y}(k)$, given $\mathbf{x}(k-1)$ and $\mathbf{y}(k-1)$, is Gaussian, so that

$$\begin{aligned} & p_{\mathbf{y}}(\mathbf{y}(k)|\mathbf{x}(k) = \mathbf{u}, \mathbf{x}(k-1) = \mathbf{v}, O_{k-1}) \\ & = \frac{1}{(2\pi)^{m/2}|\mathbf{R}|^{1/2}} \exp\{-\frac{1}{2}[\mathbf{y}(k) - \mathbf{y}(k-1) - \mathbf{h}(\mathbf{v})]^T\mathbf{R}^{-1} \\ & \qquad \cdot [\mathbf{y}(k) - \mathbf{y}(k-1) - \mathbf{h}(\mathbf{v})]\} \end{aligned} \tag{4.27}$$

The second density in the integral is that of $\mathbf{x}(k)$, given $\mathbf{x}(k-1)$. But recall that

$$\mathbf{x}(k) = \mathbf{f}(\mathbf{x}(k-1)) + \mathbf{n}(k-1) \tag{4.28}$$

Thus,

$$p_{\mathbf{x}}(\mathbf{u}, k|\mathbf{x}(k-1) = \mathbf{v}, O_{k-1})$$
$$= \frac{1}{(2\pi)^{n/2}|\mathbf{Q}|^{1/2}} \exp\{-\frac{1}{2}[\mathbf{u} - \mathbf{f}(\mathbf{v})]^T\mathbf{Q}^{-1}[\mathbf{u} - \mathbf{f}(\mathbf{v})]\} \tag{4.29}$$

Finally, the third term is merely $J_{k-1}(\mathbf{v})$. The denominator $p_{\mathbf{y}}(\mathbf{y}(k)|O_{k-1})$ can also be written as

$$p_{\mathbf{y}}(\mathbf{y}(k)|O_{k-1}))$$
$$= \int p_{\mathbf{y}}(\mathbf{y}(k)|O_{k-1}, \mathbf{x}(k-1) = \mathbf{v}) p_{\mathbf{x}}(\mathbf{v}, k-1|O_{k-1})d\mathbf{v} \tag{4.30}$$

Clearly, the second term in the integral is merely $J_{k-1}(\mathbf{v})$. The first term we have already obtained in (4.29). Substitution of these values into the Bayes formula leads to the desired result. ∎

From this theorem we note that $J_k(\mathbf{u})$ has the property that

$$\int J_k(\mathbf{u})\, d\mathbf{u} = 1 \tag{4.31}$$

This follows directly from integration. However,

$$\int \mathbf{u} J_k(\mathbf{u})\, d\mathbf{u} = \hat{\mathbf{x}}(k) \tag{4.32}$$

which is the conditional mean. Thus, the theorem provides a method whereby the estimate can be obtained by integration. To start this process, it is necessary to have $J_0(\mathbf{u})$ or the a priori statistics of the process. The performance can also be obtained by evaluating the covariance matrix

$$\mathbf{P}(k) = \int J_k(\mathbf{u})(\mathbf{u} - \hat{\mathbf{x}}(k))(\mathbf{u} - \hat{\mathbf{x}}(k))^T\, d\mathbf{u} \tag{4.33}$$

The conditional mean can be obtained as

$$\hat{\mathbf{x}}(k) = \frac{\int M(\mathbf{z}(k), \mathbf{v})d\mathbf{v} \int \mathbf{u} \exp\{-\frac{1}{2}[\mathbf{u} - \mathbf{f}(\mathbf{v})]^T\mathbf{Q}^{-1}[\mathbf{u} - \mathbf{f}(\mathbf{v})]\}\, d\mathbf{u}}{(2\pi)^{n/2}|\mathbf{Q}|^{1/2} \int M(\mathbf{z}(k), \mathbf{v})\, d\mathbf{v}} \tag{4.34}$$

where

$$M(\mathbf{z}(k), \mathbf{v}) = \exp[\mathbf{z}^T(k)\mathbf{R}^{-1}\mathbf{h}(\mathbf{v}) - \tfrac{1}{2}\mathbf{h}^T(\mathbf{v})\mathbf{R}^{-1}\mathbf{h}(\mathbf{v})] J_{k-1}(\mathbf{v}) \tag{4.35}$$

Performing the integration in (4.34) yields for $\hat{\mathbf{x}}(k)$:

$$\hat{\mathbf{x}}(k) = \frac{\int M(\mathbf{z}(k), \mathbf{v})\mathbf{f}(\mathbf{v})\, d\mathbf{v}}{\int M(\mathbf{z}(k), \mathbf{v})\, d\mathbf{v}} \tag{4.36}$$

Similarly,

$$\int \mathbf{u}\mathbf{u}^T J_k(\mathbf{v})d\mathbf{v} = \frac{\int [\mathbf{Q} + \mathbf{f}(\mathbf{v})\mathbf{f}^T(\mathbf{v})] M(\mathbf{z}(k), \mathbf{v})\, d\mathbf{v}}{\int M(\mathbf{z}(k), \mathbf{v})\, d\mathbf{v}} \tag{4.37}$$

Thus, we can write for the covariance

$$\mathbf{P}(k) = \frac{\int [\mathbf{Q} + \mathbf{f}(\mathbf{v})\mathbf{f}^T(\mathbf{v}) - 2\mathbf{f}(\mathbf{v})\hat{\mathbf{x}}(k) + \hat{\mathbf{x}}(k)\hat{\mathbf{x}}(k)] M(\mathbf{z}(k), \mathbf{v})\, d\mathbf{v}}{\int M(\mathbf{z}(k), \mathbf{v})\, d\mathbf{v}} \tag{4.38}$$

which follows after some manipulation. Now $\mathbf{P}(k)$ can be further simplified by noting that $\mathbf{Q}$ is independent of $\mathbf{v}$ and the terms within the parentheses can be simplified to yield

$$\mathbf{P}(k) = \mathbf{Q}(k) + \frac{\int (\mathbf{f}(\mathbf{v}) - \hat{\mathbf{x}}(k))\,(\mathbf{f}(\mathbf{v}) - \hat{\mathbf{x}}(k))^T M(\mathbf{z}(k), \mathbf{v})\, d\mathbf{v}}{\int M(\mathbf{z}(k), \mathbf{v})\, d\mathbf{v}} \tag{4.39}$$

Thus, $\mathbf{P}(k)$ equals the variance associated with the system noise plus a positive factor dependent upon the measurement. This method of estimation may be implemented to yield the estimate and the performance.

To show how one computes the estimate and covariance, consider Figure 5.12. Here is shown schematically what is being performed. Note that in order to "know" $J_k(\mathbf{x})$, we must have its value for all $\mathbf{x}$. Yet if we are doing it digitally, knowledge of it at some $(2N + 1)$ points will suffice. This is what we have shown. Let l_i be the point

$$l_i = \{x_{1i}, x_{2i}, \cdots, x_{ni}\} \tag{4.40}$$

where x_{ji} is a value of a state in state space.

Let us consider the five specific operations in Figure 5.12:

1. M generator: This device takes the input conditional probability density function and uses the input signal $\mathbf{z}(k)$ to generate the M functions. It does so for each time interval and over the range of the l_i points selected.

2. Normalizer: This merely takes the integral of the M functions and is used as a normalization constant throughout.

3. Probability generator: This uses the M functions and generates the next J function. It generates $(2N + 1)^2$ products and yields $(2N + 1)$ new J functions.

4. State estimator: This operates on the M functions and produces the optimum state estimate.

5. Covariance generator: This yields the performance of the estimate. This method is quite different than that using the Kalman-Bucy routine. The method does not require a calculation of $\mathbf{P}(k)$ in order to evaluate $\hat{\mathbf{x}}(k)$.

Figure 5.12 Block diagram of Bayes Law estimator. ^ The terms $b(\ell/\ell_i)$ eq
evaluated at the ℓ_i th Point.

$$\exp\left(-\frac{1}{2}\left\{\underset{\sim}{y}-\underset{\sim}{f}(\underset{\sim}{\nu})\right\}^{T}\underset{\approx}{Q}^{-1}\left\{\underset{\sim}{y}-\underset{\sim}{f}(\underset{\sim}{\nu})\right\}\right)\Big/(2\pi)^{n/2}\,|Q|^{1/2}$$

The coupling between $\hat{\mathbf{x}}(k)$ and $\mathbf{P}(k)$ is only one-directional as compared to Snyder's approach.

The filter works in the following fashion:

1. An initial value of $J(0)$ is fed into the device.
2. A measurement $\mathbf{z}(0)$ is made, and the M function is obtained.
3. The estimate $\hat{\mathbf{x}}(1)$ is evaluated.
4. The performance of $\hat{\mathbf{x}}(1)$ is evaluated yielding $\mathbf{P}(1)$.
5. The M functions are fed back and $J(1)$ is obtained.
6. The process continues recursively.

There are several advantages that this approach provides. First, it is an exact method. No Taylor-series approximations are used (see Section 6.1 for a discussion of Taylor-series approximations). Second, all inverses are done on a priori known functions. Thus, they may be calculated before hand. Third, the performance measure is exact; that is, $\mathbf{P}(k)$ is the real covariance function. It tells how well the system is functioning.

The greatest disadvantage of the method is that a great deal of storage may be necessary, depending on how spread the conditional densities become. This is the governing criterion of when to use this method.

We are now ready to develop Bucy's representation theorem. To quote from Bucy [1], "···further progress in the important area of nonlinear filtering depends on a deeper understanding of the representation theorem." This has been seen to be true in the efforts that have been made to produce rigorous mathematical proofs of the filtering equations. The theorem has been used by Kallianpur and Striebel [1, 2] in their proof and provides the building block for other approaches. The conditions for this technique are not the most general. More general results are available using the innovations approach as shown in Fujisaki, Kallianpur, and Kunita. However, the representation-theorem approach is a useful alternative for the added insight it provides. We shall thus present this theorem and then show what it implies and how at present it is not suitable for calculation. The reader is to be forewarned that the proof will *not* be as rigorous as one would desire it to be. Such rigor was felt to be beyond the scope of this book, and it was also felt that it was secondary to the implications evident from the result.

The following theorem is the discrete-state version of the representation theorem.

THEOREM 4.2

Let the observations be given by the set $O_{0,n}$, where

$$O_{0,n} = \{\mathbf{z}(i); i = 0, \cdots, n\} \tag{4.41}$$

and $\mathbf{z}(i)$ is given by

$$\mathbf{z}(i) = \mathbf{h}(\mathbf{x}(i)) + \mathbf{w}(i) \tag{4.42}$$

Let $p_{\mathbf{x}}(\mathbf{u}_n|O_{0,n})$ be the probability density of $\mathbf{x}(n)$, given the set $O_{0,n}$. Then,

$$p_{\mathbf{x}}(\mathbf{u}_n|O_{0,n}) = \frac{\int\int e^H p_{\mathbf{x}}(\mathbf{u}_0, \cdots, \mathbf{u}_{n-1}) d\mathbf{u}_0 \cdots d\mathbf{u}_{n-1} p_{\mathbf{x}}(\mathbf{u}_n)}{\int\int e^H p_{\mathbf{x}}(\mathbf{u}_0, \cdots, \mathbf{u}_n) d\mathbf{u}_0 \cdots d\mathbf{u}_n} \tag{4.43}$$

where

$$H = \sum_{i=0}^{n} [\mathbf{z}^T(i)\mathbf{R}^{-1}(i)\mathbf{h}(\mathbf{u}_i) - \tfrac{1}{2}\mathbf{h}^T(\mathbf{u}_i)\mathbf{R}^{-1}(i)\mathbf{h}(\mathbf{u}_i)] \tag{4.44}$$

and $p_{\mathbf{x}}(\mathbf{u}_n)$ is the probability density of the variable $\mathbf{x}(n)$.

Proof. Using Bayes's law, we obtain

$$p_{\mathbf{x}}(\mathbf{u}_n|O_{0,n}) = \frac{p_{\mathbf{z}}(O_{0,n}|\mathbf{x}(n) = \mathbf{u}_n)}{p_{\mathbf{z}}(O_{0,n})} p_{\mathbf{x}}(\mathbf{u}_n) \tag{4.45}$$

But $p_{\mathbf{z}}(O_{0,n}|\mathbf{x}(n) = \mathbf{u}_n)$ can be written as

$$\begin{aligned} &p_{\mathbf{z}}(O_{0,n}|\mathbf{x}(n) = \mathbf{u}_n) \\ &= \int p_{\mathbf{z}}(O_{0,n}|\mathbf{x}(n) = \mathbf{u}_n, \cdots, \mathbf{x}(0) = \mathbf{u}_0) p_{\mathbf{x}}(\mathbf{u}_0, \cdots, \mathbf{u}_{n-1}) d\mathbf{u}_0 \cdots d\mathbf{u}_{n-1} \end{aligned} \tag{4.46}$$

We can further factor the conditional density of the measurements as

$$\begin{aligned} &p_{\mathbf{z}}(O_{0,n}|\mathbf{x}(n) = \mathbf{u}_n, \cdots, \mathbf{x}(0) = \mathbf{u}_0) \\ &= p_{\mathbf{z}}(\mathbf{z}(n), \cdots, \mathbf{z}(0)|\mathbf{x}(n) = \mathbf{u}_n, \cdots, \mathbf{x}(0) = \mathbf{u}_0) \\ &= p_{\mathbf{z}}(\mathbf{z}(0)|\mathbf{x}(0) = \mathbf{u}_0, \cdots, \mathbf{x}(n) = \mathbf{u}_n) \\ &\quad \cdot p_{\mathbf{z}}(\mathbf{z}(1)|\mathbf{x}(0) = \mathbf{u}_0, \cdots, \mathbf{x}(n) = \mathbf{u}_n; \mathbf{z}(0)) \\ &\quad \cdot p_{\mathbf{z}}(\mathbf{z}(n)|\mathbf{x}(0) = \mathbf{u}_0, \cdots, \mathbf{x}(n) = \mathbf{u}_n; \mathbf{z}(0), \cdots, \mathbf{z}(n-1)) \end{aligned} \tag{4.47}$$

But for each of these densities the conditioning makes them Gaussian, since

$$\begin{aligned} p_{\mathbf{z}}(\mathbf{z}(i)|\mathbf{x}(0) = \mathbf{u}_0, \cdots, \mathbf{x}(n) = \mathbf{u}_n; \mathbf{z}(0), \cdots, \mathbf{z}(i-1)) \\ = p_{\mathbf{z}}(\mathbf{z}(i)|\mathbf{x}(i) = \mathbf{u}_i) \end{aligned} \tag{4.48}$$

and

$$\mathbf{z}(i) = \mathbf{h}(\mathbf{x}(i)) + \mathbf{w}(i) \tag{4.49}$$

Thus,

$$\begin{aligned} &p_{\mathbf{z}}(\mathbf{z}(i)|\mathbf{x}(i) = \mathbf{u}_i) \\ &= \frac{1}{(2\pi)^{m/2}|\mathbf{R}(i)|^{1/2}} \exp\left[-\left\{\frac{1}{2}[\mathbf{z}(i) - \mathbf{h}(\mathbf{u}_i)]^T\mathbf{R}^{-1}(i)[\mathbf{z}(i) - \mathbf{h}(\mathbf{u}_i)]\right\}\right] \end{aligned} \tag{4.50}$$

Thus, for each term in (4.48), we can now write

$$\begin{aligned} &p_{\mathbf{z}}(O_{0,n}|\mathbf{x}(n) = \mathbf{u}_n, \cdots, \mathbf{x}(0) = \mathbf{u}_0) \\ &= C \exp\left[-\left[\frac{1}{2}\sum_{i=0}^{n}(\mathbf{z}(i) - \mathbf{h}(\mathbf{u}_i))^T\mathbf{R}^{-1}(i)(\mathbf{z}(i) - \mathbf{h}(\mathbf{u}_i))\right]\right] \end{aligned} \tag{4.51}$$

where the quantity H is easily identified. Furthermore, we note that

$$p_z(O_{0,n}) = \int p_z(O_{0,n} | \mathbf{x}(0) = \mathbf{u}_0, \cdots, \mathbf{x}(n) = \mathbf{u}_n) p_{\mathbf{x}}(\mathbf{u}_0, \cdots, \mathbf{u}_n) d\mathbf{u}_0 \cdots d\mathbf{u}_n \quad (4.52)$$

and using (4.51) in (4.52) and then substituting (4.45), we prove the theorem. ∎

The following example shows how the previous theorem can be applied and how the conditional density function can be evaluated. This evaluation is from the work of Bucy and Senne.

Example. Consider an aircraft that is being tracked by an air-traffic control network. The position of the aircraft is given at a constant altitude z_2 moving in the x, y-coordinate space. The state vector is given by

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} \quad (4.53)$$

Now both the system and measurements are discrete-time. The state vector is assumed to be linear with the form

$$\mathbf{x}(k+1) = \boldsymbol{\Phi}(k+1, k)\,\mathbf{x}(k) + \mathbf{n}(k) \quad (4.54)$$

where

Figure 5.13 Geometric position of sensor and aircraft.

$$\Phi(k+1,k)=\begin{bmatrix}1.0 & 0\\ 0 & 1.0\end{bmatrix} \tag{4.55}$$

with the initial condition having a given probability density function. The process $\mathbf{n}(k)$ is a Gaussian independent increment processs of zero mean with covariance Q, where

$$\mathbf{Q}=\begin{bmatrix}0.050 & 0.025\\ 0.025 & 0.050\end{bmatrix} \tag{4.56}$$

The aircraft is moving in the x, y-plane as shown in Figure 5.13, and a sensor is assumed to be rotating about the origin at a unit distance measuring the relative bearing of the craft. This form of measurement is necessary to insure that the states are observable. The sensor is assumed to be in the same plane as the aircraft. The angle θ_k is the bearing of the sensor at time kT and the angle φ_k is the bearnig of the aircraft relative to the sensor at time kT.

The output of the sensor is the angle φ_k for each kT. This is given in terms of the sensor and aircraft coordinates as

$$z(k)=h(\mathbf{x}(k),k)+w(k) \tag{4.57}$$

where

$$h(\mathbf{x}(k),k)=\tan^{-1}\left[\frac{x_2(k)-\sin\theta_k}{x_1(k)-\cos\theta_k}\right] \tag{4.58}$$

and $w(k)$ is a zero mean white Gaussian sequence with covariance R ($R=0.01$).

Now, in this example the initial density is assumed to have four distinct modes, as shown in Figure 5.14 (a). As time change, the conditional density changes into a single mode initially the incorrect one and finally the correct mode. From these densities the conditional mean and variance can be calculated. The initial caputre of the incorrect mode is a common feature of nonlinear estimators and should be recognized in any simulation. The report by Bucy, Hecht, and Senne carefully discusses this issue in detail.

We now want to consider the continuous version of the representation theorem. To do so, we shall first express the previous result in a slightly different form.

COROLLARY 4.1. For the discrete-state observation process the conditional density state, given the observations, is given by

$$p_{\mathbf{x}}(\mathbf{u}_n|O_{0,n})=\frac{E[e^H|\mathbf{x}(n)=\mathbf{u}_n]}{E[e^H]}\,p_{\mathbf{x}}(\mathbf{u}_n) \tag{4.59}$$

where H is

$$H=\sum_{i=0}^{n}\left[\mathbf{z}^T(i)\mathbf{R}^{-1}(i)\mathbf{h}(\mathbf{u}_i)-\tfrac{1}{2}\,\mathbf{h}^T(\mathbf{u}_i)\mathbf{R}^{-1}(i)\mathbf{h}(\mathbf{u}_i)\right] \tag{4.60}$$

Figure 5.14 (*a*) Initial condition of probability density; four initial modes in x_2 coordinate. (*b*) Estimated density at $k = 6$ time units; peaking at wrong state. (*c*) Estimated density at $k = 19$; further peaking at incorrect x_2 value. (*d*) Estimated density at $k = 50$; final locking of density on correct value of state.

and $E[\ ]$ represents the expectation over all $\mathbf{x}(i)$ and $E[\ \ |\mathbf{x}(n) = \mathbf{u}_n]$ is the expectation over all given $\mathbf{x}(n) = \mathbf{u}_n$.

This corollary presents the result in a more compact form by representing the integrals over the state probability densities by the expectation operator. Now let $nT = t$, some arbitrary time. Let t_0 be some initial time and divide the interval $[t_0,t]$ into n subintervals. Now as we let n go to infinity and T to to 0 such that $nT = t - t_0$, we define

$$H_t = \lim_{\substack{n\to\infty \\ T\to 0 \\ nT\to t-t_0}} H = \lim_{\substack{n\to\infty \\ T\to 0 \\ nt\to t-t_0}} \left[\sum_{i=0}^{n} \mathbf{z}^T(i)\mathbf{R}^{-1}(i)\mathbf{h}(u_i) - \frac{1}{2}\mathbf{h}^T(u_i)\mathbf{R}^{-1}(i)\mathbf{h}(u_i) \right] \tag{4.61}$$

Now recall that

$$\sum_{i=0}^{n} \mathbf{z}^T(i)\mathbf{R}^{-1}(i)\mathbf{h}(u_i)$$
$$= \sum_{i=0}^{n} \frac{[\mathbf{y}((i+1)T) - \mathbf{y}(iT)]^T}{T} \mathbf{R}^{-1}(iT)^T\mathbf{h}(x(iT)) \tag{4.62}$$

In the limit we obtain

$$\lim_{\substack{n\to\infty \\ T\to 0 \\ nT\to t-t_0}} \sum_{i=1}^{n} \mathbf{z}^T(i)\mathbf{R}^{-1}(i)\mathbf{h}(u_i) = \int_{t_0}^{t} \mathbf{h}^T(\mathbf{x}(\xi))\mathbf{R}^{-1}(\xi)\, d\mathbf{y}(\xi) \tag{4.63}$$

where the integral is interpreted in the Ito sense. Similarly,

$$\lim_{\substack{n\to 0 \\ T\to 0 \\ nT\to t-t_0}} \sum_{i=0}^{n} \mathbf{h}^T(\mathbf{u}_i)\mathbf{R}^{-1}(i)\mathbf{h}(\mathbf{u}_i) = \int_{t_0}^{t} \mathbf{h}^T(\mathbf{x}(\xi))\mathbf{R}^{-1}(\xi)\mathbf{h}(\mathbf{x}(\xi))\, d\xi \tag{4.64}$$

Thus, if we consider the expectation operators in the preceding corollary to be extended suitably as $n \to \infty$, $T \to 0$, we obtain the following continuous-time version of the representation theorem.

THEOREM 4.3

Let $\mathbf{x}(t)$ be an $(n \times 1)$-vector Markov process given by

$$d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}(t), t)\, dt + d\mathbf{n}(t) \tag{4.65}$$

and $\mathbf{y}(t)$ an $(m \times 1)$-vector Markov process given by

$$d\mathbf{y}(t) = \mathbf{h}(\mathbf{x}(t), t)\, dt + d\mathbf{w}(t) \tag{4.66}$$

where

$$E[d\mathbf{n}\, d\mathbf{n}^T] = \mathbf{Q}(t)\, dt \tag{4.67}$$

$$E[d\mathbf{w}\, d\mathbf{w}^T] = \mathbf{R}(t)\, dt \tag{4.68}$$

then

$$p_{\mathbf{x}}(\mathbf{u}, t \,|\, O_{t_0,t}) = \frac{E[e^{H_t} \,|\, \mathbf{x}(t) = \mathbf{u}]}{E[e^{H_t}]}\, p_{\mathbf{x}}(\mathbf{u}, t) \tag{4.69}$$

where

$$H_t = \int_{t_0}^{t} \mathbf{h}^T(\mathbf{x}(\xi))\mathbf{R}^{-1}(\xi)d\mathbf{y}(\xi) - \tfrac{1}{2}\int_{t_0}^{t} \mathbf{h}^T(\mathbf{x}(\xi))\mathbf{R}^{-1}(\xi)\mathbf{h}(\mathbf{x}(\xi))\, d\xi \tag{4.70}$$

and $E[\ ]$ is the expectation operator on $\mathbf{x}(s)$ and $O_{t_0,t}$ is the minimum σ-field generated by the process $\mathbf{y}(t)$ on $[t_0, t]$.

Proof. See Kallianpur and Striebel [2, 3].

The usefulness of this theorem is in obtaining the progapation of $p_{\mathbf{x}}(\mathbf{u}, t \,|\, O_{t_0,t})$ directly from it. This has been heuristically done, first in Bucy and then by Jaswinski. Problem 5.19 outlines the proof. In general, this theorem provides a useful mathematical tool for exact proofs but does not lend itself to easy implementation. In the next chapter we shall use the propagation equation developed in the last section to develop the propagation equations for the optimal estimates.

5.5 CONCLUSIONS

The quantity sought after in most filtering problems is the conditional expectation of the process $\mathbf{x}(t)$, given a record of some measurement process. Previous approaches—those by Masani and Wiener; Balakrishnan; and Dolph and Woodbury, for example—concentrated on obtaining this conditional expectation using the orthogonality properties directly. However, many of their solution techniques were quite limited, because they considered too wide a class of stochastic processes. However, because of work of Kush-

ner and Stratonovich, interest in the filtering of Markov processes led to the analysis of the conditional density function. It is this equation and its ramifications that have played the central role in this chapter.

In the first section we concentrated on the efficacy of the Markov model to describe a large enough class of processes to be useful as a theory. Two special classes of models are of particular usefulness. The first is that which represents the dynamics of a physical system such, as an electrical circuit or the trajectory of a satellite, that is perturbed by some random external force. To a first-order approximation this force can be modeled by a white noise process, either Gaussian or Poisson or both. The resulting system motion becomes a Markov process. The second class of models considers the problem of finding a system that is driven by white noise and has required second-order properties. For example, if we have a stationary Markov process with a known spectrum, how do we obtain a state-space description of this process? This is also called the spectrum-factorization problem. Now the results of this analysis can be used to model nonwhite noise disturbances that affect dynamical systems, thus expanding the repertoire of possible models. Thus, the general model given by

$$d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}(t), t)\, dt + d\mathbf{n}(t) \tag{5.1}$$

where $d\mathbf{n}(t)$ is the sum of Wiener and Poisson processes, is a robust enough model to satisfy the constraints of many problems.

The second issue of model development concerns the method of measurements. Two specific types of measurements are discussed. The first is the additive Gaussian measurement described by

$$d\mathbf{y}(t) = \mathbf{h}(\mathbf{x}(t), t)\, dt + d\mathbf{w}(t) \tag{5.2}$$

where $\mathbf{w}(t)$ is a Wiener process, and the second is a Poisson measurement process $d\mathbf{N}(t)$ whose arrival rate $\boldsymbol{\lambda}(\mathbf{x}(t), t)$ depends upon the state of the system. These equations are to be interpreted as the Ito type equations, as developed in Chapter 3. However, a formalism can be developed if in the case of linear time-invariant systems we interpret $d\mathbf{n}_g(t)/dt$ as a white noise process. This interpretation has been used extensively in more elementary treatments of filtering.

The second section developed the propagation equations for the transition densities of the Markov process model that we have developed. The approach taken in this section was to use the theorem of Bartlett-Moyal, and from it we obtained the operator equation

$$\frac{\partial p}{\partial t} = L^{+} p \tag{5.3}$$

The operator L^{+} is called the forward operator (thus the plus sign on the L). Our approach assumed that probability density function existed and was

unique. The conditions given by Duncan were assumed to hold. Other approaches to the development of the propagation equations is via semigroup theory as described by Dynkin or Wong [ ]. The evaluation of the operator L^+ was obtained for the system driven by a Wiener process or by a Poisson process. The resulting equations are called the Fokker-Planck and Feller-Kolmolgorov equations, respectively. The approach used for these cases can be extended to more arbitrary independent increment driving functions.

The final topic discussed in this section was that of the generalized Fokker-Planck equation for non-Markov processes. This equation represents an example of how the propagation equation for a generalized transition density can be developed. However, the use of this equation is quite limited and to date has not been used extensively. Yet the structure thus developed is necessary whenever the driving function is not decomposable into processes generated by independent increment processes (i.e., Markov process systems).

The third section of this chapter forms the kernel of the entire book—namely, the development of the propagation for the conditional density of the state given the sub-sigma algebra generated by measurements. The result given in Theorem 3.1 shows the relationship between L^+, the forward operator, and a function $q(d\mathbf{y}, dt, \mathbf{u})$, which depends upon the measurements. Thus, given any measurement, we must find this function and the propagation follows directly. This added function may be considered as a forcing function to the propagation equation with the L^+ operator. Thus, the measurement acts as a perturbation on the transition density function.

In this section we obtained the $q(d\mathbf{y}, dt, \mathbf{u})$ function for the Gaussian measurement using the technique developed by Kushner and the result for Poisson measurement following Snyder. Other techniques for the development of these equations can be obtained by other methods, particularly those using the representation theorem. The solution to these propagation equations are highly complex except for the linear case where the result is known to be Gaussian.

The fourth section developed the representation theorem proposed by Bucy [1]. It is essentially a function-space representation of Bayes's theorem and, as such, had already been known in the statistical literature (see Cameron and Martin). As we have shown, the heuristic derivation is quite straightforward, especially in the discrete-time case; however, the difficulties arise in obtaining a rigorous proof under sufficiently general conditions. These have been provided by Kallianpur and Striebel [3]. There is, however, a fundamental limitation on derivations using the representation theorem in that they cannot be used whenever the signal component is allowed to depend on the past observations, as in feedback communications and control systems as well as in colored noise systems solved by reduction to a white noise problem.

In the problems, we used this theorem to develop the propagation equation for the conditional-density via the Ito differential rule. Similar results were also presented by Mortensen [1], who later presented a quite readable presentation of its usefulness (see Mortensen [2]). In the presentation developed in the section, we first considered the discrete-time system and obtained an integral iterative technique for the evaluation of the conditional density. Following this, a discrete-time version of the representation theorem was developed and the continuous-time results presented. An outline of the proof of the continuous-time results was presented in the problems using the martingale convergence results.

The results on function space and dynamic systems were first applied by Darling and Siegert [1, 2] and Siegert [1, 3] and provide a basis for the results given in the representation theorem. Furthermore, the results using function-space techniques have been applied by Evans [2] for the evaluation of performance bounds of estimators.

The approach used here to obtain the conditional density, and thus the conditional expectation, is as we have said nonunique. Other approaches. such as the innovations approach (see Kailath and Frost) or the reproducing kernel Hilbert-space approach (see Kailath [7]) also provide alternatives to the conditional-density approach. In the next section we shall pursue the conditional-density approach to obtain the conditional expectation.

5.6 PROBLEMS

5.1. Let the state equation be given by

$$\dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t)$$

where $\mathbf{u}(t)$ is an $n \times 1$ Gaussian white noise process.

(a) Let $\mathbf{x}(t)$ be given for $t_1 < t$; find $\mathbf{x}(t_2)$ for $t_2 > t_1$ using the Ito interpretation.

(b) Let $\{t_i\}$ be a finite ordered set such that $t_i < t_{i+1}$. Find $\mathbf{x}(t_j)$ in terms of $\mathbf{x}(t_k)$ for $t_k < t_j$.

(c) Let $p_{\mathbf{x}}(\mathbf{u}, t_j|\mathbf{x}(t_1) \cdots \mathbf{x}(t_{j-1}))$ be the conditional density of $\mathbf{x}(t_j)$, given $\mathbf{x}(t_1) \cdots \mathbf{x}(t_{j-1})$. Show that $\mathbf{x}(t)$ is a Markov process.

(d) Find $p_{\mathbf{x}}(\mathbf{u}, t_j|\mathbf{x}(t_{j-1}))$.

5.2. Consider the problem in example on page ~~00~~ 184. Assume that

$$E[\mathbf{u}(t)] = \mathbf{m}(t)$$

(a) Rederive the result accounting for this change.

(b) Evaluate an expression for $E[\mathbf{x}(t)]$ assuming that $E[\mathbf{x}(0)] = \mathbf{x}_0$.

5.3. Let $\mathbf{x}(t)$ be given by

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 0 & 1 \\ -3 & -4 \end{bmatrix} \mathbf{x}(t) + \mathbf{u}(t)$$

where the covariance of $\mathbf{u}(t)$ is $\delta(t - s)\mathbf{I}$.

(a) Let $y(t)$ be

$$y(t) = [1 \quad 2]\mathbf{x}(t)$$

Find $K_y(t)$, the covariance of the process $y(t)$.

(b) Find $S_y(f)$, the power spectrum of $y(t)$.

5.4. Let $\mathbf{x}(t)$ be given by

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 0 & 1 \\ a & b \end{bmatrix} \mathbf{x}(t) + \mathbf{u}(t)$$

and

$$y(t) = [1 \quad 0]\mathbf{x}(t)$$

(a) Find $K_y(t)$ in terms of a and b.

(b) Find values of a,b to make $K_y(t)$ to be exponential and damped sinusoid.

(c) Evaluate $S_y(f)$ for those cases in part (b).

(d) A message $r(t)$ is composed of a signal $s(t)$ and white Gaussian noise $n(t)$

$$r(t) = n(t) + s(t)$$

The noise has spectrum

$$S_y(f) = \frac{N_0}{2}$$

while $s(t)$ has a spectrum

$$S_s(f) = \frac{10}{f^4 + 4}$$

Find a and b to obtain a state variable realization for this process.

5.5. Let $\mathbf{x}(t)$ be an $(n \times 1)$-vector Markov process given by

$$d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}, t)\, dt + \boldsymbol{\sigma}(\mathbf{x}, t)\, d\mathbf{n}_g(t) + \boldsymbol{\gamma}(\mathbf{x}, t)\, d\mathbf{n}_p(t)$$

where $\mathbf{n}_g(t)$ and $\mathbf{n}_p(t)$ are as defined in the text and are $(q \times 1)$- and $(p \times 1)$-vector processes, respectively. Assume that

$$\boldsymbol{\sigma}^T(\mathbf{x}, t)\boldsymbol{\sigma}(\mathbf{x}, t) = \Sigma(\mathbf{x}, t)$$

and

$$\boldsymbol{\gamma}^T(\mathbf{x}, t)\boldsymbol{\gamma}(\mathbf{x}, t) = \Gamma(\mathbf{x}, t)$$

are positive definite and both Σ_{ij} and Γ_{ij} satisfy the Lipschitz conditions of Definition 2.2. Obtain the propagation equation for $p_{\mathbf{x}}(\mathbf{u}, t \,|\, \mathbf{x}(s))$, $s < t$. 5.6. Show that (2.47) is the solution to (2.45) with the given initial conditions. 5.7. Consider the problem discussed in example on page 000. In state form it is

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 0 & 1 \\ 0 & -\dfrac{a}{m} \end{bmatrix} \mathbf{x}(t) + \mathbf{u}(t)$$

where

$$E[\mathbf{u}(t)\mathbf{u}(s)] = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \delta(t - s)$$

with $\mathbf{x}(t_0) = \mathbf{x}_0$.

(a) Write the complete Fokker-Planck equation for this system.

(b) Note that since the system is linear, both $x_1(t)$ and $x_2(t)$ are linear. Evaluate

i. $E[x_1(t)|x_{10}]$

ii. $E[x_2(t)|x_{20}]$

iii. $K_{x_1}(t, s)$

iv. $K_{x_2}(t, s)$

v. $E[x_1(t)x_2(s)|x_{10}, x_{20}]$

(c) With the result of part (b), find

$$p_{x_1x_2}(u_1, u_2; t\,|x_{10}, x_{20}; t_0)$$

(d) Show that the result of part (c) is a solution to part (a).

5.8. Solve Fokker-Planck equation for the following cases:

(a) Diffusion process: $\dot{x}(t) = \dot{u}(t)$; $E[u^2(t)] = t$

(b) Orenstein-Uhlenbeck process:

$$\dot{x}(t) = -ax(t) + \dot{u}(t); E[u^2(t)] = \sigma^2 t\,; x(t_0) = x_0$$

and $a > 0$.

(c) For both of the above processes, note that $x(t)$ is Gaussian. Evaluate

$$E[x(t)|x_0]$$

and

$$K_x(t, s)$$

and obtain $p_x(u, t\,|x_0)$. Show that it satisfies the corresponding Fokker-Planck equation.

5.9. Let $x(t)$ be a scalar-valued Markov process given by

$$dx(t) = f(x, t)\,dt + \sigma(x)\,dn_p(t) \quad (x(0) = 0)$$

where $n_p(t)$ is a simple step Poisson process with rate λ.

(a) Find the Feller-Kolmogorov equation.

(b) Let $f(x, t) = 0$ and $\sigma(x) = -x/|x|$. Sketch a possible sample path for $x(t)$. (This is the random telegraph wave.)

(c) Find $p_x(u, t\,|x(0) = 0)$. Note that it will be impulsive, so simplify.

5.10. Let $x(t)$ be a scalar process given by

$$\frac{dx(t)}{dt} = \frac{dn_p(t)}{dt} \qquad (x(0) = 0 \qquad \text{w. p. 1})$$

where $n_p(t)$ is a generalized Poisson process of rate λ and n_p can assume only jump in unit steps of $+1$.

(a) Show that the Feller-Kolmogorov equations can be written as

$$\frac{dP_n}{dt} = \lambda P_{n-1} - \lambda P_n$$

where $P_n = P[x(t) = n]$.

(b) Show that

$$P_n(t) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}$$

(c) Show that the Feller-Kolmogorov equation can be written as

$$\frac{\partial p(x, t)}{\partial t} = \lambda p(x - 1, t) - \lambda p(x, t)$$

where $p(x, t)$ is the envelope of a set of impulse at all the positive integers.

(d) Expand $p(x - 1, t)$ about $x = 1$ and show that $p(x, t)$ satisfies

$$\frac{\partial p}{\partial t} = -\frac{\partial p}{\partial x} + \frac{1}{2}\frac{\partial^2 p}{\partial x^2}$$

and show that the solution is

$$p(x, t) = \frac{1}{(2\pi\lambda t)^{1/2}} \exp\left(-\frac{(x - \lambda t)^2}{2\lambda t}\right)$$

(e) Compare this continuous density to the Poisson probability for λ large.

5.11. (Viterbi) The phase-lock loop is a device that is used to estimate the phase of a signal of the form

$$z(t) = A\cos(\omega_0 t + \theta(t)) + w(t)$$

where $w(t)$ is white Gaussian noise and $\theta(t)$ is the phase. A feedback loop is constructed, and the error signal

$$e(t) = \theta(t) - \hat{\theta}(t)$$

can be shown to satisfy the following equation:

$$\dot{e}(t) + c_1 \sin(e(t)) = n(t) \qquad (*)$$

where $n(t)$ is white Gaussian noise with spectral height $N_0/2$. Let $p_e(E, 0) = \delta(E - \theta_0)$.

(a) Write the Fokker-Planck equation for (*).

(b) Let $t \gg 0$ and assume $(\partial/\partial t)\, p_e(u, t) \approx 0$; find the steady state value of $p_e(u, t)$.

5.12. Let

$$\dot{x} = -ax \qquad (x(0) \text{ Gaussian } (\sigma_0, m_0))$$

$$z = x + \dot{w} \qquad (E[w(t)w^T(s)] = R(t)\min(t, s))$$

(a) Let $P = E[(x - \bar{x})^2]$. What is P?

(b) Let $\sum = E[(x - \bar{x})^2]$, where $\bar{x} = E[x(t)]$; solve from Fokker-Planck equation.

(c) Show that $\sum = \lim_{R \to \infty} P$.

(d) Sketch $P(t)$ and $\sum(t)$ versus t for $R = 10\ m_0^2,\ m_0^2,\ m_0^2/10$. Comment on the result.

5.13. Consider the linear time-varying system

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{w}(t)$$

where $\mathbf{A}(t)$ is an $n \times n$ time-varying matrix and $\mathbf{B}(t)$ is an $n \times q$ time-varying matrix. Let $\mathbf{w}(t)$ be a zero mean white noise process with

$$E[\mathbf{w}(t)\mathbf{w}^T(s)] = \mathbf{Q}(t)\delta(t - s)$$

where $\mathbf{Q}(t)$ is a $q \times q$ positive definite matrix. The measurement is

$$\mathbf{z}(t) = \mathbf{C}(t)\mathbf{x}(t) + \mathbf{u}(t)$$

where $\mathbf{C}(t)$ is $m \times n$ and $\mathbf{u}(t)$ is $m \times 1$ white noise with

$$E[\mathbf{u}(t)\mathbf{u}^T(s)] = \mathbf{R}(t)\delta(t - s)$$

Furthermore, assume that

$$E[\mathbf{w}(t)\mathbf{u}^T(s)] = \mathbf{S}(t)\delta(t - s).$$

Obtain the Kushner-Stratonovich equation for this equation.

5.14. Let $\mathbf{x}(t)$ be given by the vector equation

$$d\mathbf{x}(t) = \mathbf{A}(t)\mathbf{x}(t)\, dt + d\mathbf{n}_p(t)$$

where $\mathbf{n}_p(t)$ is a generalized Poisson process. Let $\mathbf{a}_i$ be the jumps heights on the ith jump. $\mathbf{a}_i$ is an $n \times 1$ vector. Let

$$p_{\mathbf{a}_i}(\mathbf{u}) = \frac{1}{(2\pi)^{n/2}|\Lambda_{a_i}|^{1/2}} \exp\left(-\mathbf{u}^T \Lambda_{a_i}^{-1}\mathbf{u}\right)$$

where Λ_{a_i} is the covariance matrix of $\mathbf{a}_i$. Write the Kushner-Stratonovich equation for this case when

$$d\mathbf{y}(t) = \mathbf{C}(t)\mathbf{x}(t)\, dt + \mathbf{dw}(t)$$

5.15. Let $x(t)$ be given by

$$dx(t) = A(t)x(t)\,dt + dn_p(t)$$

where $x(t)$ is scalar and $A(t)$ is a time-varying scalar. $n_p(t)$ is a scalar Poisson process with jump heights ± 1 occurring with equal probability. Let

$$dy(t) = x(t)\,dt + dw(t)$$

where $E[w^2(t)] = t$.

(a) Write the Kusher-Stratonovich-equation for this system.

(b) Let $p_y(v, t)$ be the unconditional probability density of the measurement process. Find $p_y(v, t)$.

5.16. Let $x(t)$ be a scalar given by

$$dx(t) = Ax(t)\,dt \qquad (A<0)$$

where $x(t_0)$ is a Gaussian random variable. Let $z(t)$ be a scalar Poisson counting process with

$$\lambda(x, t) = \exp[-\lambda x(t)]$$

(a) Write Snyder's equation for this case.

(b) Let $t - t_0 \to 0$, find $p_x(u, t\,|O_{t_0,t})$.

(c) Let A be as given and let $P[N(t) = k]$ be the unconditioned probability that $N(t)$ equals k (note $N(0) = 0$). Find $P[N(t) = k]$.

(d) Now let $A = 0$ in part (c). Find $P[N(t) = k]$.

5.17. Let $\mathbf{x}(t)$ be an $(n \times 1)$-vector Markov process given by

$$d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}, t)\,dt + \boldsymbol{\sigma}(\mathbf{x}, t)\,d\mathbf{u}(t)$$

and $\mathbf{y}(t)$ an $(m \times 1)$-vector Markov process satisfying

$$d\mathbf{y}(t) = \mathbf{h}(\mathbf{x}, t)\,dt + \boldsymbol{\gamma}(\mathbf{x}, t)\,d\mathbf{w}(t)$$

where both $\boldsymbol{\sigma}(\mathbf{x}, t)$ and $\boldsymbol{\gamma}(\mathbf{x}, t)$ satisfy the SSSR. Write the Kushner-Stratonovich equation for the case where both $\mathbf{u}(t)$ and $\mathbf{w}(t)$ are Wiener processes. Repeat for the case where

$$\mathbf{u}(t) = \mathbf{u}_p(t) + \mathbf{u}_g(t)$$

5.18. Let $\mathbf{x}(t)$ be an $(n \times 1)$-vector Markov process given by

$$d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}, t)\,dt + \boldsymbol{\sigma}(\mathbf{x}, t)\,d\mathbf{u}(t)$$

where $\mathbf{u}(t)$ is a $q \times 1$ Wiener process. Let $\mathbf{y}(t) = \mathbf{N}(t)$ an $m \times 1$ Poisson counting process. Evaluate Snyder's equation for this case.

5.19. The representation theorem can be written as

$$p_{\mathbf{x}}(\mathbf{u}, t\,|O_{t_0,t}) = \frac{E[e^{\xi(t)}|\mathbf{x}(t) = \mathbf{u}]p_{\mathbf{x}}(\mathbf{u}, t\,|\mathbf{x}(t_0))}{E[e^{\xi(t)}]}$$

where

$$\xi(t) = \int_{t_0}^{t} \mathbf{h}^T(\mathbf{x}(\tau), \tau)\mathbf{R}^{-1}(\tau)\, d\mathbf{y}(\tau) - \tfrac{1}{2}\int_{t_0}^{t} \mathbf{h}^T(\mathbf{x}(\tau), \tau)\mathbf{R}^{-1}(\tau)\mathbf{h}(\mathbf{x}(\tau), \tau)\, d\tau$$

(a) From the definition of $\xi(t)$ show that

$$d\xi(t) = \tfrac{1}{2}\,\mathbf{h}^T(\mathbf{x}(t), t)\mathbf{R}^{-1}(t)\mathbf{h}(\mathbf{x}(t), t) + \mathbf{h}^T(\mathbf{x}(t), t)\mathbf{R}^{-1}(t)d\mathbf{w}(t)$$

(b) Let

$$Q = E[e^{\xi(t)}|\mathbf{x}(t) = \mathbf{u}]p_{\mathbf{x}}(\mathbf{u}, t\,|\mathbf{x}(t_0))$$

and

$$P = E[e^{\xi(t)}] = \int Q(\mathbf{u})\, d\mathbf{u}.$$

Using Ito's lemma show that

$$dQ = L^{+}Q\, dt + Q\mathbf{h}^T(\mathbf{x}(t), t)\mathbf{R}^{-1}(t)\, d\mathbf{y}(t)$$

where L^{+} is the forward Fokker-Planck operator. Also show that

$$dP = PE[\mathbf{h}^T(\mathbf{x}(t), t)]\mathbf{R}^{-1}(t)\, d\mathbf{y}(t)$$

(c) Use the results of part (b) to obtain the Kushner-Stratonovich equation.

5.20. Consider the system

$$d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}(t), t)\, dt + d\mathbf{n}(t)$$

where $\mathbf{n}(t)$ is an $(n \times 1)$-vector Wiener process and the measurement is

$$d\mathbf{y}(t) = \mathbf{h}(\mathbf{x}(t), t) + d\mathbf{w}(t)$$

Let

$$E[d\mathbf{n}(t)\, d\mathbf{n}^T(t)] = \mathbf{Q}(t)\, dt$$
$$E[d\mathbf{n}(t)\, d\mathbf{w}^T(t)] = \mathbf{S}(t)\, dt$$
$$E[d\mathbf{w}(t)\, d\mathbf{w}^T(t)] = \mathbf{R}(t)\, dt$$

Obtain the representation theorem for this case.

5.21. Let $\mathbf{x}(t)$ be given by

$$d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}(t), t)\, dt + d\mathbf{n}(t)$$

and let $\mathbf{y}(t)$ be $\mathbf{N}(t)$, an $(m \times 1)$-vector Poisson counting process with $(m \times 1)$ vector rate $\boldsymbol{\lambda}(\mathbf{x},t)$. Develop a representation theorem for this case similar to that for Gaussian measurements.

5.22. Using the martingale convergence theorem of Section 4.2 of chapter 4, prove the continuous-time version of the representation theorem.

CHAPTER 6

ESTIMATION EQUATIONS

The main objective of estimation is to obtain a set of equations that will allow the measurements to be transformed into good estimates of the quantities sought. In Chapter 4 we found that the conditional mean gave an estimate that minimized the mean square error. We also found that under certain restrictions on the probability densities, it minimized other cost criteria also. For this reason we spent considerable effort in Chapter 5 obtaining an equation for the conditional probability density of the state, given measurements over some interval. Unfortunately, the resulting equation was a nonlinear partial-differential integral equation that presented little chance of solution except under certain explicit constraints on both the system and the measurement.

In this chapter we shall develop equations for the conditional mean, or MMSE estimate of the state, by making assumptions concerning the nature of the nonlinearities. Except for the case of a linear system with linear Gaussian measurements, our solutions will be approximations. Yet in many cases these approximations will be valid estimates, albeit less precise than the exact solutions. In Section 6.1 we use both the Kushner-Stratonovich equation and Snyder's equation to obtain the approximate estimates of the states for Gaussian and Poisson measurements, respectively. Both techniques assume that the nonlinearities can be expanded in multidimensional Taylor series. The resulting equations for the estimate and covariance matrix are thus obtained. Several examples are given to show how these techniques can be implemented and to demonstrate the effects of including different terms in the approximations to the nonlinearities.

A discussion of the exact nature of the linear-system linear Gaussian measurement equations for the estimate and covariance is presented. The resulting equations in this case are called the *Kalman-Bucy equations* and were first presented in 1961. The initial derivation of these equations was from a different approach, one more consistent with deterministic optimal control. The reason for this is that the linear optimal control problem with

a quadratic cost criterion is the dual of the linear MMSE problem. This was first noted by Kalman in 1959. The importance of the Kalman-Bucy equations should not be understated, and since their inception a great deal of literature has grown up about the linear problem. For those interested in a more complete discussion of this problem, the books by Meditch [2]; Sage and Melsa; and Jaswinski [2] are recommended, as well as the original paper by Kalman, republished in 1963 (see Kalman [3]).

We also discuss a linearization technique of the estimator for the case of Poisson measurements. Unfortunately, there does not seen to be a general class of problems with exact solutions. Evans [2] in 1971 found that a class of problems considered by Wonham [1] were subject to an exact analysis by the fact that the state space was discrete. Whether this restriction is equivalent to linear-system linear measurement in the Gaussian case is still an unanswered question.

In Section 6.2 we present an analysis of the problem of estimating the state of a continuous-time nonlinear system based upon discrete-time linear measurements. This analysis is presented to reinforce the ideas initiated in 6.1 concerning expansions of nonlinearities. In this section, we expand the state nonlinearity about the optimal estimate but the residual of the expansion is unspecified. By establishing a cost criterion one can obtain this residual recursively. This section is based upon the work of Athans, Wishner, and Bertolini. Similar analyses have been done for continuous systems and discrete measurements by Jaswinski [1, 2] and Culver [1, 2].

Section 6.3 presents a technique for the estimation of the state of a discrete-time system based on discrete-time measurements. It begins with the conditional density, but instead of determining conditional means, which we have found is difficult, it finds those states that maximize the a posteriori probability density. For this reason the estimates are called *maximum a posteriori* (MAP) estimates. In the case of linear systems the MAP and MMSE estimates are identical.

The problem one obtains upon maximizing the a posteriori density problem can be solved by techniques of dynamic programming. Yet these procedures can be quite time-consuming in computation time, and approximate procedures are sought. The resulting approximations again assume expansion of the nonlinearity. Upon doing this, one finds that the optimization problem can be solved in a recursive fashion yielding the discrete-time version of the results in Section 6.1. This result was originally obtained by Cox in 1964.

Issues of stability and divergence of the resulting filters are discussed in Section 6.4. We first present some of the general divergence issues, indicate how they arise, and qualitatively discuss their effects. We then choose a specific issue of incomplete model determination and quantitatively evaluate

the effect. The resulting analysis raises issues of deterministic stability with respect to the discrete-time linear filter. The complete analysis of these issues is performed in Appendix C, but the results are presented in this section. The issues raised and discussed in this section are essential in the proper implementation of the filter.

Section 6.5 discusses extensions of the results developed. It briefly discusses the use of quasimoment functions first introduced by Kuznetsov, Stratonovich, and Tikhanov in 1960 and used by Fisher in 1968 for the nonlinear estimation problem. Integration of the stochastic equations is also discussed. We also mention alternate approaches that produce the same results.

This chapter contains the main results of this book in terms of estimator structure. It can be read independently of the remaining text if one will accept the required results.

6.1 CONTINUOUS-TIME LINEARIZED ESTIMATION EQUATIONS

The previous chapter developed the machinery necessary to obtain the MMSE estimates of the state of a nonlinear system. This section presents varying techniques employed to obtain estimates. The material discussed in this section derives from Snyder [1–5] and from Bass, Norum, and Schwartz [1, 2]. We present a continuous-time approach that starts from the Kushner-Stratonovich equation and Snyder's equation of Chapter 5. In so doing, we obtain estimation equations coupled with performance equations. The basic approach here and elsewhere is a linearization of the nonlinearities. The disadvantages of such a scheme will be discussed at the end of the chapter. Its advantages often outweigh these problems, so that its consideration is essential.

What we shall do in this section is, first, to take the Kushner-Stratonovich equations and obtain the conditional mean. This is the estimate that was sought. We then expand the nonlinearities in a multidimensional Taylor-series expansion and proceed to integrate the resulting expressions. Such a procedure introduces higher-order moments. Equations for these higher moments are required. We truncate the expansion by obtaining only the second central moment. This truncation assumes that linear approximations are accurate in the Taylor-series expansions.

A similar analysis is performed for the case where the measurements are jump Poisson processes. We use Snyder's equation of Chapter 5 as the fundamental equation for both the evaluation of the conditional mean and variance. Again the nonlinearities are expanded in Taylor series and only the first three terms in the series are retained. The system is also assumed to be disturbed only by a Wiener process.

We conclude the section by the discussion of several special cases. In general, the approach used in this chapter is still too complex, although Synder [1] used the special case of linear state equations and nonlinear measurements to analyze several important communications systems. An attempt to use these results for both nonlinear systems and states is a formidable task. For this reason we introduce the extended Kalman filter, which is a first-order attempt to handle the problem of nonlinearities. The extended Kalman filter, which is exact for the case of linear systems and linear measurements, is quite frequently a useful tool in nonlinear problems. Also, as we shall see in the following three sections, a discrete-state version is desirable for computational reasons. For the reader interested in applications and practical techniques, suitable references are provided.

The system equation to be studied is given by the $(n \times 1)$-vector stochastic differential equation;

$$d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}(t), t)\, dt + d\mathbf{n}_g(t) \tag{1.1}$$

Where $d\mathbf{n}_g$ is an $(n \times 1)$-vector Wiener process with covariance

$$E\left[d\mathbf{n}_g\, d\mathbf{n}_g^T\right] = \mathbf{Q}(t)\, dt \tag{1.2}$$

We shall assume that $\mathbf{f}(\mathbf{x}(t), t)$ is sufficiently smooth in $\mathbf{x}(t)$ that a Taylor-series expansion of the nonlinearity exists about any point $\mathbf{x}^*(t)$ for all t of interest. This expansion can be written

$$\mathbf{f}(\mathbf{x}, t) = \mathbf{f}(\mathbf{x}^*, t) + \mathbf{A}(\mathbf{x}^*, t)(\mathbf{x} - \mathbf{x}^*) + \tfrac{1}{2}\sum_{i=1}^{n} \gamma_i(\mathbf{x} - \mathbf{x}^*)^T \mathbf{B}(\mathbf{x} - \mathbf{x}^*) + \cdots \tag{1.3}$$

where

$$\mathbf{A}(\mathbf{x}^*, t) = \left[\begin{array}{ccc} \dfrac{\partial f_1}{\partial x_1} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \dfrac{\partial f_n}{\partial x_1} & \cdots & \dfrac{\partial f_n}{\partial x_i} \end{array}\right]_{\mathbf{x} = \mathbf{x}^*} \tag{1.4}$$

where f_i and x_i are the ith components of the vectors $\mathbf{f}(\mathbf{x}, t)$ and $\mathbf{x}$, respectively, and

$$\gamma_i = \left[\begin{array}{c} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{array}\right] \begin{array}{l} 1 \\ \\ \\ i \\ \\ \\ n \end{array} \tag{1.5}$$

is a vector with 1 in the ith position and 0 elsewhere and

$$\mathbf{B}_i(\mathbf{x}^*, t) = \begin{bmatrix} \dfrac{\partial^2 f_i}{\partial x_1 \partial x_1} & \dfrac{\partial^2 f_i}{\partial x_1 \partial x_n} \\ \vdots & \vdots \\ \dfrac{\partial^2 f_i}{\partial x_n \partial x_1} & \dfrac{\partial^2 f_i}{\partial x_n \partial x_n} \end{bmatrix}_{\mathbf{x} = \mathbf{x}^*} \tag{1.6}$$

is an $n \times n$ matrix of the second partial derivatives of the nonlinearity. Higher-order terms in the Taylor-series expansion follow directly from the difinitions. For our purposes these three terms are sufficient for an analysis of the filtering problem.

For the case of Gaussian measurements, the measurement equation is

$$d\mathbf{y} = \mathbf{h}(\mathbf{x}, t)\, dt + d\mathbf{w}(t) \tag{1.7}$$

where $\mathbf{h}(\mathbf{x}, t)$ is an $m \times 1$ nonlinear vector function of the state. This function is also assumed to be sufficiently smooth, so that a Taylor series expansion can be performed. Thus, for an arbitrary $\mathbf{x}^*(t)$, we can write

$$\mathbf{h}(\mathbf{x}, t) = \mathbf{h}(\mathbf{x}^*, t) + \mathbf{C}(\mathbf{x}^*, t)(\mathbf{x} - \mathbf{x}^*) + \tfrac{1}{2} \sum_{i=1}^{m} \boldsymbol{\gamma}_i (\mathbf{x} - \mathbf{x}^*)^T \mathbf{F}_i(\mathbf{x}^*, t)(\mathbf{x} - \mathbf{x}^*) \tag{1.8}$$

where

$$\mathbf{C}(\mathbf{x}^*, t) = \begin{bmatrix} \dfrac{\partial h_1}{\partial x_1} & \dfrac{\partial h_1}{\partial x_n} \\ \vdots & \vdots \\ \dfrac{\partial h_m}{\partial x_1} & \dfrac{\partial h_m}{\partial x_n} \end{bmatrix}_{\mathbf{x} = \mathbf{x}^*} \tag{1.9}$$

is an $m \times n$ matrix of the partial derivatives of the measurement nonlinearity. The $m \times 1$ vector $\boldsymbol{\gamma}_i$ is a vector as before with a 1 in the ith position and 0 elsewhere. The dimension of $\boldsymbol{\gamma}_i$ will be understood from the context in which it is used. The matrix $\mathbf{F}_i(\mathbf{x}^*, t)$ is given by

$$\mathbf{F}_i(\mathbf{x}^*, t) = \begin{bmatrix} \dfrac{\partial^2 h_i}{\partial x_1 \partial x_1} & \dfrac{\partial^2 h_i}{\partial x_1 \partial x_n} \\ \vdots & \vdots \\ \dfrac{\partial^2 h_i}{\partial x_n \partial x_1} & \dfrac{\partial^2 h_i}{\partial x_n \partial x_n} \end{bmatrix}_{\mathbf{x} = \mathbf{x}^*} \tag{1.10}$$

Similarly, for the case of Poisson measurements, the vector arrival rate $\boldsymbol{\lambda}(\mathbf{x}, t)$ is also assumed to have a multidimensional Taylor series expansion given by;

$$\boldsymbol{\lambda}(x, t) = \boldsymbol{\lambda}(\mathbf{x}^*, t) + \mathbf{D}(\mathbf{x}^*, t)(\mathbf{x} - \mathbf{x}^*) + \frac{1}{2} \sum_{i=1}^{m} \boldsymbol{\gamma}_i (\mathbf{x} - \mathbf{x}^*)^T \mathbf{E}_i(\mathbf{x}^*, t)(\mathbf{x} - \mathbf{x}^*) \tag{1.11}$$

where

$$\mathbf{D}(\mathbf{x}^*, t) = \left[\begin{array}{ccc} \dfrac{\partial \lambda_1}{\partial x_1} & & \dfrac{\partial \lambda_1}{\partial x_n} \\ \vdots & & \vdots \\ \dfrac{\partial \lambda_m}{\partial x_1} & & \dfrac{\partial \lambda_m}{\partial x_n} \end{array}\right]_{\mathbf{x} = \mathbf{x}^*} \tag{1.12}$$

and

$$\mathbf{E}_i(\mathbf{x}^*, t) = \left[\begin{array}{ccc} \dfrac{\partial^2 \lambda_i}{\partial x_1 \, \partial x_1} & & \dfrac{\partial^2 \lambda_i}{\partial x_1 \, \partial x_n} \\ \vdots & & \vdots \\ \dfrac{\partial^2 \lambda_i}{\partial x_n \, \partial x_1} & & \dfrac{\partial^2 \lambda_i}{\partial x_n \, \partial x_n} \end{array}\right]_{\mathbf{x} = \mathbf{x}^*} \tag{1.13}$$

and γ_i is an $m \times 1$ vector as already defined.

We make one further set of assumptions, concerning the central moments of the process and the nature of the conditional probability density of the process $\mathbf{x}(t)$. Let us first define

$$P_{i_1 i_2 i_3} = E\left[(x_{i_1} - \hat{x}_{i_1})(x_{i_2} - \hat{x}_{i_2})(x_{i_3} - \hat{x}_{i_3}) \mid O_{t_0 t}\right] \tag{1.14}$$

to be the third conditional central moment where $\hat{x}_i$ is the expected value of the ith component of $\mathbf{x}(t)$, given $O_{t_0 t}$, namely,

$$\hat{x}_i = E\left[x_i(t) \mid O_{t_0 t}\right] \tag{1.15}$$

Let the fourth central moment be

$$P_{i_1 i_2 i_3 i_4} = E\left[(x_{i_1} - \hat{x}_{i_1})(x_{i_2} - \hat{x}_{i_2})(x_{i_3} - \hat{x}_{i_3})(x_{i_4} - \hat{x}_{i_4}) \middle| O_{t_0 t}\right] \tag{1.16}$$

The first set of assumptions are that

$$P_{i_1 i_2 i_3} = 0; \qquad \forall i_1, i_2, i_3 \tag{1.17}$$

and that the fourth central moment factors as

$$P_{i_1 i_2 i_3 i_4} = P_{i_1 i_2} P_{i_3 i_4} + P_{i_1 i_3} P_{i_2 i_4} + P_{i_1 i_4} P_{i_2 i_3} \tag{1.18}$$

These assumptions are based upon the assumption that the conditional density is almost Gaussian, for if it were, all odd central moments would be zero and the fourth central moment would factor as above (see Papoulis).

The next assumption concerns the nature of the higher moments of the process. Let $P_{i_1 \cdots i_l}$ be also the lth central moment. Then we assume that

$$P_{i_1 \cdots i_n} = 0; \qquad \forall n \geqq 5 \tag{1.19}$$

that is, $P_{i_1 i_2 i_3 i_4 i_5} = 0$, and so on. Clearly, if the Gaussian assumption holds, then this would be true for all odd n. The argument for even n is based upon the fact that by linearizing we are already assuming that $\mathbf{x} \approx \hat{\mathbf{x}}$, so that $\|\mathbf{P}(t)\|$ is small. For the Gaussian case all even moments relate to the second moment in a product fashion. Thus, if $P_{i_1 i_2}$ is small, then $P_{i_1 i_2} P_{i_3 i_4}$, $P_{i_1 i_3} P_{i_2 i_4}$, $P_{i_1 i_4} P_{i_2 i_3}$ will be much smaller and thus can be considered negligible.

The last assumption concerns the behavior of the probability density function $p_{\mathbf{x}}(\mathbf{u}, t \mid O_{t_0,t}) = p$. It is not based upon any linearization argument, but it is used in an integration by parts intercahnge in the proof of the lemmas that follow. We assume that p is bounded and that

$$p\Big|_{-\infty}^{\infty} = \frac{\partial p}{\partial u_i}\Big|_{-\infty}^{\infty} = u_i \frac{\partial p}{\partial u_i}\Big|_{-\infty}^{\infty} = f_i p\Big|_{-\infty}^{\infty} = u_i f_i p\Big|_{-\infty}^{\infty} = 0 \tag{1.20}$$

for all values of i. We shall comment on these approximations after we observe the consequence of having made them. Furthermore, for simplicity the dependence on $\mathbf{x}^*$ in the coefficients $\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D},\mathbf{E}_i$, and $\mathbf{F}_i$ will be suppressed.

The main objective of this section is to obtain the propagation equation for the conditional mean $\mathbf{x}(t)$. This is now done in the following two theorems for the Gaussian measurement case and Poisson measurement, respectively.

THEOREM 1.1

Let $\mathbf{x}(t)$ be an $(n \times 1)$-vector Markov process given by

$$d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}, t)\, dt + d\mathbf{n}(t) \tag{1.21}$$

where $d\mathbf{n}$ is an $(n \times 1)$-vector Wiener process. Let the measurement $d\mathbf{y}$ be given by

$$d\mathbf{y}(t) = \mathbf{h}(\mathbf{x}, t)\, dt + d\mathbf{w}(t) \tag{1.22}$$

Then $\mathbf{x}^*(t)$, the linearized approximation of the conditional mean, is given by

$$\begin{aligned}\frac{d\mathbf{x}^*(t)}{dt} &= \mathbf{f}(\mathbf{x}^*, t) + \frac{1}{2}\sum_{i=1}^{n} \gamma_i \operatorname{tr}[\mathbf{P}(t)\mathbf{B}_i(t)] \\ &\quad + \mathbf{P}(t)\mathbf{C}(t)\mathbf{R}^{-1}(t)\Big[\mathbf{z}(t) - \mathbf{h}(\mathbf{x}^*, t) - \frac{1}{2}\sum_{i=1}^{m} \gamma_i \operatorname{tr}[\mathbf{P}(t)\mathbf{F}_i(t)]\Big]\end{aligned} \tag{1.23}$$

(where "tr" means "trace") with $\mathbf{x}^*(t_0)$ given by the a priori estimate and where $\mathbf{P}^*(t)$ is the $n \times n$ linearized covariance matrix given by

$$\begin{aligned}\frac{d\mathbf{P}^*}{dt} &= \mathbf{P}^*(t)\mathbf{A}^T(t) + \mathbf{A}(t)\mathbf{P}^*(t) - \mathbf{P}^*(t)\mathbf{C}^T(t)\mathbf{R}^{-1}(t)\mathbf{C}(t)\mathbf{P}^*(t) \\ &\quad + \mathbf{Q}(t) + \sum_{i=1}^{m} \mathbf{P}^*(t)\mathbf{F}_i(t)\mathbf{P}^*(t)\gamma_i^T\mathbf{R}^{-1}(t) \\ &\quad \cdot \Big[\mathbf{z}(t) - \mathbf{h}(\mathbf{x}^*, t) - \frac{1}{2}\sum_{i=1}^{m} \gamma_i \operatorname{tr}[\mathbf{P}^*(t)\mathbf{F}_i(t)]\Big]\end{aligned} \tag{1.24}$$

with $\mathbf{P}^*(t_0)$ given by the a priori covariance matrix. This is called the *second-order linearized estimate*.

Before proving this theorem we shall first prove two lemmas that are independent of any approximation results and may be used independently. We shall use these lemmas to show the correctness of the continuous-time linear filter. We should also note that $\mathbf{x}^*(t)$ will always denote the linearized estimate, obtained by approximating the nonlinearities. The quantity $\hat{\mathbf{x}}(t)$

will be the actual conditional mean and MMSE estimate of the state. Likewise, $\mathbf{P}^*(t)$ will be the covariance that is obtained when the nonlinearities are linearized, whereas $\mathbf{P}(t)$ is the actual covariance. We shall show under what conditions $\mathbf{P}(t) = \mathbf{P}^*(t)$ and $\hat{\mathbf{x}}(t) = \mathbf{x}^*(t)$ at the end of this section.

LEMMA 1.1. Let $\mathbf{x}(t)$ be given by the $(n \times 1)$-vector Markov process driven by an $(n \times 1)$-vector Wiener process $\mathbf{n}(t)$ given by

$$d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}, t)\, dt + d\mathbf{n}(t) \tag{1.25}$$

and the measurement equation is an $(m \times 1)$-vector Markov process

$$d\mathbf{y}(t) = \mathbf{h}(\mathbf{x}, t)\, dt + d\mathbf{w}(t) \tag{1.26}$$

Then the MMSE estimate is generated by

$$\frac{d\hat{\mathbf{x}}(t)}{dt} = E[\mathbf{f}(\mathbf{x}, t)] + E[(\mathbf{x}(t) - \hat{\mathbf{x}}(t))\mathbf{h}^T(\mathbf{x}, t)] \cdot \mathbf{R}^{-1}(t)[\mathbf{z}(t) - E[\mathbf{h}(\mathbf{x}, t)]] \tag{1.27}$$

where $E[\ ]$ represents the expectation with respect to $p_{\mathbf{x}}(\mathbf{u}, t|O_{t_0,t})$, $\mathbf{z}(t) = d\mathbf{y}/dt$ and $\mathbf{R}(t)$ is the covariance

$$E[d\mathbf{n}(t)d\mathbf{n}^T(t)] = \mathbf{R}(t)dt \tag{1.28}$$

Proof. From the Kushner-Stratonovich equation we know that

$$\frac{\partial p}{\partial t} = -\sum_{i=1}^{n} \frac{\partial(f_i p)}{\partial u_i} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} Q_{ij}\frac{\partial^2 p}{\partial u_i \partial u_j} + p[\mathbf{h}(\mathbf{u}, t) - E[\mathbf{h}(\mathbf{x}, t)]]^T\mathbf{R}^{-1}(t)[\mathbf{z}(t) - E[\mathbf{h}(\mathbf{x}, t)]] \tag{1.29}$$

where $p = p_{\mathbf{x}}(\mathbf{u}, t|O_{t_0,t})$, the conditional density. Now multiplying the Kushner-Stratonovich equation by the $n \times 1$ vector $\mathbf{u}$ and integrating, we obtain

$$\int \mathbf{u}\frac{\partial p}{\partial t}\, d\mathbf{u} = \frac{\partial}{\partial t}\int \mathbf{u}p\, d\mathbf{u} = \frac{\partial \hat{\mathbf{x}}(t)}{\partial t} \tag{1.30}$$

Likewise, we can integrate by parts to yield

$$\int u_i \frac{\partial(f_i p)}{\partial u_i}\, d\mathbf{u} = \int f_i p\Big|_{-\infty}^{\infty} du_i - \int f_i p\, d\mathbf{u} \tag{1.31}$$

where $\int$ means integration over all except the ith $\mathbf{u}$, and then using our assumptions about the conditional density yields

$$= -\int f_i p\, d\mathbf{u} \tag{1.31}$$

Furthermore, the integrals for $j \neq i$ can be evaluated

$$\int u_j \frac{\partial(f_i p)}{\partial u_i}\, d\mathbf{u} = \int u_j(f_i p)\Big|_{-\infty}^{\infty} du_i = 0 \tag{1.32}$$

Thus, the first term on the right yields

$$\int \mathbf{u}\left(-\sum_{i=1}^{n}\frac{\partial(f_i(\mathbf{u},t)p)}{\partial u_i}\right)d\mathbf{u} = \int \mathbf{f}(\mathbf{u},t)p\,d\mathbf{u} \tag{1.33}$$

which is $E[\mathbf{f}(\mathbf{x},t)\mid O_{t_0,t}]$ by definition. By integrating the second term by parts, we easily show that

$$\int \mathbf{u}Q_{ij}\frac{\partial^2 p}{\partial u_i\,\partial u_j}\,d\mathbf{u} = \mathbf{0} \tag{1.34}$$

The third term becomes

$$\begin{aligned}&\int \mathbf{u}p[\mathbf{h}(\mathbf{u},t) - E[\mathbf{h}(\mathbf{x},t)]]^T\mathbf{R}^{-1}(t)[\mathbf{z}(t) - E[\mathbf{h}(\mathbf{x},t)]]\,d\mathbf{u}\\ &= E[\mathbf{x}(t)[\mathbf{h}(\mathbf{x},t) - E[\mathbf{h}(\mathbf{x},t)]]^T]\mathbf{R}^{-1}(t)[\mathbf{z}(t) - E[\mathbf{h}(\mathbf{x},t)]]\end{aligned} \tag{1.35}$$

But we note that

$$\begin{aligned}&E[\mathbf{x}(t)[\mathbf{h}(\mathbf{x},t) - E[\mathbf{h}(\mathbf{x},t)]]^T]\\ &= E[\mathbf{x}(t)[\mathbf{h}(\mathbf{x},t)]^T] - E[\mathbf{x}(t)]E[\mathbf{h}^T(\mathbf{x},t)]\\ &= E[[\mathbf{x}(t) - E[\mathbf{x}(t)]]\mathbf{h}^T(\mathbf{x},t)]\end{aligned} \tag{1.36}$$

Therefore, using these results in the Kushner-Stratonovich integration yields the desired result. ∎

LEMMA 1.2. Let $\mathbf{x}(t)$,$\mathbf{y}(t)$, and $\hat{\mathbf{x}}(t)$ be given as in the provious lemma. Let

$$\mathbf{P}(t) = E[(\mathbf{x}(t) - \hat{\mathbf{x}}(t))(\mathbf{x}(t) - \hat{\mathbf{x}}(t))^T \mid O_{t_0,t}] \tag{1.37}$$

Then $\mathbf{P}(t)$ satisfies the equation

$$\begin{aligned}\frac{d\mathbf{P}}{dt} &= E[(\mathbf{x} - \hat{\mathbf{x}})\mathbf{f}^T(\mathbf{x})] + E[\mathbf{f}(\mathbf{x})(\mathbf{x} - \hat{\mathbf{x}})^T] + \mathbf{Q}(t)\\ &\quad + E[(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^T(\mathbf{h}(\mathbf{x},t) - E[\mathbf{h}(\mathbf{x},t)])^T]\mathbf{R}^{-1}(t)\\ &\quad [\mathbf{z}(t) - E[\mathbf{h}(\mathbf{x},t)]]\end{aligned} \tag{1.38}$$

where $\mathbf{z}(t)$ is $d\mathbf{y}(t)/dt$ and the operator $E[\ ]$ is the expectation with respect to the density $p_{\mathbf{x}}(\mathbf{u},t\mid O_{t_0,t})$.

Note that the covariance $\mathbf{P}(t)$ is the conditional covariance and, as such, depends on the measurements. We see this directly by the appearance of $\mathbf{z}(t)$ in the propagation equation. Only in the linear case does this dependence on the measurements disappear.

Proof. By definition $d\mathbf{P}(t)$ is given by

$$\begin{aligned}d\mathbf{P}(t) &= \mathbf{P}(t+dt) - \mathbf{P}(t)\\ &= \int(\mathbf{u} - \hat{\mathbf{x}}(t+dt))(\mathbf{u} - \hat{\mathbf{x}}(t+dt))^T p_{\mathbf{x}}(\mathbf{u},t+dt \mid O_{t_0,t+dt})\,d\mathbf{u}\\ &\quad - \int(\mathbf{u} - \hat{\mathbf{x}}(t))(\mathbf{u} - \hat{\mathbf{x}}(t))^T p_{\mathbf{x}}(\mathbf{u},t\mid O_{t_0,t})\,d\mathbf{u}\end{aligned} \tag{1.39}$$

Also, we have

$$d\tilde{\mathbf{x}}(t) = \tilde{\mathbf{x}}(t + dt) - \tilde{\mathbf{x}}(t) \tag{1.40}$$

Now we can write $P(t + dt)$as

$$\mathbf{P}(t + dt) = \int (\mathbf{u} - \tilde{\mathbf{x}}(t) - d\tilde{\mathbf{x}})(\mathbf{u} - \hat{\mathbf{x}}(t) - d\hat{\mathbf{x}})^T p_{\mathbf{x}}(\mathbf{u}, t + dt \mid O_{t_0,t+dt})\, d\mathbf{u}$$

$$= \int (\mathbf{u} - \hat{\mathbf{x}}(t))(\mathbf{u} - \hat{\mathbf{x}}(t))^T p_{\mathbf{x}}(\mathbf{u}, t + dt \mid O_{t_0,t+dt})\, d\mathbf{u} + d\tilde{\mathbf{x}}\, d\hat{\mathbf{x}}^T \tag{1.41}$$

$$- \int (\mathbf{u} - \hat{\mathbf{x}}(t))\, d\hat{\mathbf{x}}^T p_{\mathbf{x}}(\mathbf{u}, t + dt \mid O_{t_0,t+dt})\, d\mathbf{u}$$

$$- \int d\tilde{\mathbf{x}}(\mathbf{u} - \hat{\mathbf{x}}(t))^T p_{\mathbf{x}}(\mathbf{u}, t + dt \mid O_{t_0,t+dt})\, d\mathbf{u} \tag{1.42}$$

But we can rewrite the third and fourth integrals on the right of the above expression differently by noting that

$$\int (\mathbf{u} - \hat{\mathbf{x}}(t)) d\hat{\mathbf{x}}^T p_{\mathbf{x}}(\mathbf{u}, t + dt \mid O_{t_0,t+dt})\, d\mathbf{u}$$

$$= \int (\mathbf{u} - \tilde{\mathbf{x}}(t + dt) + d\tilde{\mathbf{x}}) d\hat{\mathbf{x}}^T p_{\mathbf{x}}(\mathbf{u}, t + dt \mid O_{t_0,t+dt})\, d\mathbf{u} \tag{1.43}$$

$$= \int (\mathbf{u} - \tilde{\mathbf{x}}(t + dt)) d\hat{\mathbf{x}}^T p_{\mathbf{x}}(\mathbf{u}, t + dt \mid O_{t_0,t+dt})\, d\mathbf{u} + d\tilde{\mathbf{x}}\, d\hat{\mathbf{x}}^T$$

But since the estimate $\hat{\mathbf{x}}(t + dt)$ is unbiased, this equals $d\hat{\mathbf{x}}\, d\hat{\mathbf{x}}^T$. Thus, for $\mathbf{P}(t + dt)$ we can write

$$\mathbf{P}(t + dt) = \int (\mathbf{u} - \tilde{\mathbf{x}}(t))(\mathbf{u} - \hat{\mathbf{x}}(t))^T p_{\mathbf{x}}(\mathbf{u}, t + dt \mid O_{t_0,t+dt})\, d\mathbf{u} - d\tilde{\mathbf{x}}\, d\hat{\mathbf{x}}^T \tag{1.44}$$

The object now is to note that $p_{\mathbf{x}}(\mathbf{u}, t + dt \mid O_{t_0,t+dt})$ can be written in terms of $p_{\mathbf{x}}(\mathbf{u}, t \mid O_{t_0,t})$ by means of the Kushner-Stratonivich equation, also note that $d\hat{\mathbf{x}}\ d\hat{\mathbf{x}}^T$ follows directly from the preceding lemma. That is,

$$d\hat{\mathbf{x}} = E[\mathbf{f}(\mathbf{x}, t)]\, dt + E[(\mathbf{x} - \hat{\mathbf{x}})\mathbf{h}^T(\mathbf{x}, t)]\mathbf{R}^{-1}[d\mathbf{y} - E[\mathbf{h}(\mathbf{x}, t)]\, dt] \tag{1.45}$$

Thus, to $o(dt)$ it can easily be shown that

$$d\hat{\mathbf{x}}\, d\hat{\mathbf{x}}^T = E[(\mathbf{x} - \hat{\mathbf{x}})\mathbf{h}^T(\mathbf{x}, t)]\mathbf{R}^{-1}\, d\mathbf{y}\, d\mathbf{y}^T\, \mathbf{R}^{-1} E[\mathbf{h}(\mathbf{x}, t)(\mathbf{x} - \hat{\mathbf{x}})^T] + o(dt) \tag{1.46}$$

But also to $o(dt)$ we have

$$d\mathbf{y}\ d\mathbf{y} = \mathbf{R}\, dt; \qquad \text{w.p.1} \tag{1.47}$$

Thus,

$$d\hat{\mathbf{x}}\ d\hat{\mathbf{x}}^T = E[(\mathbf{x} - \hat{\mathbf{x}})\mathbf{h}^T(\mathbf{x}, t)]\mathbf{R}^{-1} E[\mathbf{h}(\mathbf{x}, t)(\mathbf{x} - \hat{\mathbf{x}})^T]\, dt \tag{1.48}$$

We can now return to the evaluation of the integral. Recall that from the Kushner-Stratonovich equation that

$$p_{\mathbf{x}}(\mathbf{u}, t + dt \mid O_{t_0,t+dt}) = p_{\mathbf{x}}(\mathbf{u}, t \mid O_{t_0,t}) + L^+ p\, dt + p[\mathbf{h}(\mathbf{u}, t) - E[\mathbf{h}(\mathbf{x}, t)]]^T \cdot \mathbf{R}^{-1}(t)[d\mathbf{y} - E[\mathbf{h}(\mathbf{x}, t)]\, dt] \tag{1.49}$$

Consider now the evaluation of

$$\int (u_k - \hat{x}_k)(u_l - \hat{x}_l) L^+ p \, d\mathbf{u}$$
$$= \int (u_k - \hat{x}_k)(u_l - \hat{x}_l) \left[- \sum_{i=1}^{n} \frac{\partial (f_i p)}{\partial u_i} + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} Q_{ij} \frac{\partial^2 p}{\partial u_i \, \partial u_j} \right] d\mathbf{u} \quad (1.50)$$

Integration by parts yields

$$= E[(x_k - \hat{x}_k) f_l(\mathbf{x})] + E[(x_l - \hat{x}_l) f_k(\mathbf{x})] + Q_{kl} \quad (1.51)$$

Thus, using this in the above, we find that after some matrix manipulation we have

$$\begin{aligned} \mathbf{P}(t + dt) = \mathbf{P}(t) &+ E[(\mathbf{x} - \hat{\mathbf{x}})\mathbf{f}^T(\mathbf{x}) | O_{t_0 t}] \, dt \\ &+ E[\mathbf{f}(\mathbf{x})(\mathbf{x} - \hat{\mathbf{x}})^T | O_{t_0 t}] \, dt + \mathbf{Q}(t) \, dt \\ &+ E[(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^T (\mathbf{h}(\mathbf{x}, t) - E[\mathbf{h}(\mathbf{x}, t)])^T | O_{t_0 t}] \\ &\cdot \mathbf{R}^{-1}[d\mathbf{y} - E[\mathbf{h}(\mathbf{x}, t)] \, dt] \\ &- E[(\mathbf{x} - \hat{\mathbf{x}})\mathbf{h}^T(\mathbf{x}, t)] \mathbf{R}^{-1} E[\mathbf{h}(\mathbf{x}, t)(\mathbf{x} - \hat{\mathbf{x}})^T] \, dt \end{aligned} \quad (1.52)$$

Formally dividing through by dt proves the lemma. ∎

Proof (of Theorem 1.1). To evaluate the propagation of the conditional mean we use the result of the first of the preceding lemmas and use the expansion of the nonlinearities about $\hat{\mathbf{x}}(t)$. Thus,

$$E\left[\mathbf{f}(\hat{\mathbf{x}}, t) + \sum_{i=1}^{n} \mathbf{A}(t)(\mathbf{x} - \hat{\mathbf{x}}) + \frac{1}{2} \sum_{i=1}^{n} \gamma_i (\mathbf{x} - \hat{\mathbf{x}})^T \mathbf{B}_i (\mathbf{x} - \hat{\mathbf{x}}) \right]$$
$$= \mathbf{f}(\hat{\mathbf{x}}, t) + \frac{1}{2} \sum_{i=1}^{n} \gamma_i \operatorname{tr}[\mathbf{P}(t) \mathbf{B}_i(t)] \quad (1.53)$$

and the second term on the left vanishes as a result of the unbiased nature of $\mathbf{x}$. Similarly, for $E[\mathbf{h}(\mathbf{x}, t)]$ we obtain

$$E[\mathbf{h}(\mathbf{x}, t)] \approx \mathbf{h}(\hat{\mathbf{x}}, t) + \frac{1}{2} \sum_{i=1}^{n} \gamma_i \operatorname{tr}[\mathbf{P}(t)\mathbf{F}_i(t)] \quad (1.54)$$

The quantity denoted by the trace results from

$$\operatorname{tr}[\mathbf{P}(t)\mathbf{B}_i(t)] = \sum_{j=1}^{n} \sum_{k=1}^{n} P_{jk}(t) B_{ijk}(t)$$
$$= E[(\mathbf{x} - \mathbf{x})^T \mathbf{B}_i (\mathbf{x} - \mathbf{x})] \quad (1.55)$$

The same holds for tr $[\mathbf{P}(t)\mathbf{F}_i(t)]$. The last quantity to be evaluated is

$$E[(\mathbf{x}(t) - \hat{\mathbf{x}}(t))\mathbf{h}^T(\mathbf{x}, t)] \simeq E[(\mathbf{x}(t) - \hat{\mathbf{x}}(t))\mathbf{h}^T(\hat{\mathbf{x}}, t)]$$
$$+ E[(\mathbf{x}(t) - \hat{\mathbf{x}}(t))((\mathbf{x}(t) - \hat{\mathbf{x}}(t))\mathbf{C}^T(t) + \cdots = \mathbf{P}(t)\mathbf{C}^T(t) \quad (1.56)$$

Thus, using these approximations, we obtain

$$\frac{d\mathbf{x}^*}{dt} = \mathbf{f}(\mathbf{x}^*, t) + \frac{1}{2}\sum_{i=1}^{n} \gamma_i \operatorname{tr}[\mathbf{P}^*(t)\mathbf{B}_i(t)] + \mathbf{P}^*(t)\mathbf{C}^T(t)\mathbf{R}^{-1}(t)\left[\mathbf{z}(t) - \mathbf{h}(\mathbf{x}^*, t) - \frac{1}{2}\sum_{i=1}^{m} \gamma_i \operatorname{tr}[\mathbf{R}^*(t)\mathbf{F}_i(t)]\right] \quad (1.57)$$

which proves the first part of the theorem. Note that we have used $\mathbf{x}^*(t)$ to denote the fact that this is a result of a linearization and is not exact. The second part requires using the second lemma and evaluating the propagation of the covariance. Using the expansions, we can easily show that

$$E[(\mathbf{x} - \hat{\mathbf{x}})\mathbf{f}^T(\mathbf{x})] \simeq E[(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^T\mathbf{A}^T(t)] = \mathbf{P}(t)\mathbf{A}^T(t) \quad (1.58)$$

and that

$$E[\mathbf{f}(\mathbf{x})(\mathbf{x} - \hat{\mathbf{x}})^T] = E[\mathbf{A}(t)(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^T] = \mathbf{A}(t)\mathbf{P}(t) \quad (1.59)$$

Likewise,

$$E[(\mathbf{x} - \hat{\mathbf{x}})\mathbf{h}^T(\mathbf{x}, t)] \simeq \mathbf{P}(t)\mathbf{C}^T(t) \quad (1.60)$$

and

$$E[\mathbf{h}(\mathbf{x}, t)(\mathbf{x} - \hat{\mathbf{x}})^T] \simeq \mathbf{C}(t)\mathbf{P}(t) \quad (1.61)$$

This leaves us with the evaluation of the term

$$E[(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^T(\mathbf{h}(\mathbf{x}, t) - E[\mathbf{h}(\mathbf{x}, t)])^T]$$

First, we note that to the order of our approximations we can write

$$\begin{aligned}&\mathbf{h}(\mathbf{x}, t) - E[\mathbf{h}(\mathbf{x}, t)] \\ &\simeq \mathbf{h}(\hat{\mathbf{x}}, t) + \mathbf{C}(t)(\mathbf{x} - \hat{\mathbf{x}}) + \frac{1}{2}\sum_{i=1}^{n} \gamma_i(\mathbf{x} - \hat{\mathbf{x}})^T\mathbf{F}_i(t)(\mathbf{x} - \hat{\mathbf{x}}) \\ &\quad - \mathbf{h}(\hat{\mathbf{x}}, t) - \frac{1}{2}\sum_{i=1}^{n} \gamma_i E[(\mathbf{x} - \hat{\mathbf{x}})^T\mathbf{F}_i(t)(\mathbf{x} - \hat{\mathbf{x}})] \end{aligned} \quad (1.62)$$

Consider now the k, l entry of $\mathbf{P}$, namely, P_{kl}. The contribution of this final term to the propagation of P_{kl} is given by(to the order of our approximations)

$$\begin{aligned}E\Big[&(x_k - \hat{x}_k)(x_l - \hat{x}_l)\Big[(\mathbf{x} - \hat{\mathbf{x}})^T\mathbf{C}^T(t) \\ &+ \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - \hat{x}_j)(x_j - \hat{x}_j)\frac{\partial^2\mathbf{h}^T}{\partial x_i^*\,\partial x_i^*} \\ &- E\Big[\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - \hat{x}_i)(x_j - \hat{x}_j)\frac{\partial^2\mathbf{h}^T}{\partial x_i^*\,\partial x_j^*}\Big]\Big]\Big]\end{aligned} \quad (1.63)$$

Where we have written

$$\begin{aligned}&\frac{1}{2}\sum_{i=1}^{n} \gamma_i(\mathbf{x} - \hat{\mathbf{x}})\mathbf{F}_i(t)(\mathbf{x} - \hat{\mathbf{x}}) \\ &= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - \hat{x}_i)(x_j - \hat{x}_j)\frac{\partial^2\mathbf{h}}{\partial x_i^*\,\partial x_j^*}\end{aligned} \quad (1.64)$$

and $\partial^2 \mathbf{h}/(\partial x_i^* \, \partial x_j^*)$ is the second partial of the $m \times 1$ vector $\mathbf{h}$ with respect to the respective components of $\mathbf{x}$ evaluated at $\hat{\mathbf{x}}(t)$. Now, because of the zero assumption of third central moments, we have for (1.63) the following:

$$E\left[\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[(x_k-\hat{x}_k)(x_l-\hat{x}_l)(x_j-\hat{x}_i)(x_j-\hat{x}_j)\right.\right.$$
$$\left.\left.-(x_k-\hat{x}_k)(x_l-\hat{x}_l)P_{ij}\right]\frac{\partial^2 \mathbf{h}^T}{\partial x_i^* \, \partial x_j^*}\right] \tag{1.65}$$

But by our assumption on the fourth moment, this equals

$$\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[P_{ki}P_{jl}+P_{kj}P_{il}\right]\frac{\partial^2 \mathbf{h}^T}{\partial x_i^* \, \partial x_j^*}$$
$$=\sum_{i=1}^{n}\sum_{j=1}^{n}P_{ki}P_{jl}\frac{\partial^2 \mathbf{h}^T}{\partial x_i^* \, \partial x_j^*} \tag{1.66}$$

Now we can write the partial as

$$\frac{\partial^2 \mathbf{h}^T}{\partial x_i^* \, \partial x_j^*}=\sum_{s=1}^{m}\boldsymbol{\gamma}_s^T\frac{\partial^2 h_s}{\partial x_i^* \, \partial x_j^*} \tag{1.67}$$

Where $\boldsymbol{\gamma}_s$ is as defined before and h_s is the sth component of $\mathbf{h}$. Thus,

$$\sum_{i=1}^{n}\sum_{j=1}^{n}P_{ki}P_{jl}\frac{\partial^2 \mathbf{h}^T}{\partial x_i^* \, \partial x_j^*}=\sum_{s=1}^{m}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}P_{ki}F_{s,i,j}P_{jl}\right)\boldsymbol{\gamma}_s^T \tag{1.68}$$

where

$$F_{s,i,j}=\frac{\partial^2 h_s}{\partial x_i^* \, \partial x_j^*} \tag{1.69}$$

But the term $\sum_{i=1}^{n}\sum_{j=1}^{n}P_{ki}F_{s,ij}P_{jl}$ is nothing more than the k,lth entry of the product $\mathbf{P}(t)\mathbf{F}_s(t)\mathbf{P}(t)$. Therefore, we obtain

$$E\left[(\mathbf{x}-\mathbf{x})(\mathbf{x}-\mathbf{x})^T\left[\mathbf{h}(\mathbf{x},t)-E[\mathbf{h}(\mathbf{x},t)]\right]^T\right]$$
$$=\sum_{i=1}^{m}\mathbf{P}(t)\mathbf{F}_i(t)\mathbf{P}(t)\boldsymbol{\gamma}_i^T \tag{1.70}$$

Thus, using these approximations in the propagation for the covariance, we satisfy the second part of the theorem. ∎

A more common form of the preceding theorem results if only the linear term is used in part of the expansions. This was first used by Snyder [1] in 1966. The derivation of the preceding result also rests on Snyder [1] in that the fourth central moment factorability is used. In the work of Bass, Norum, and Schwartz [1,2] in 1966, they assumed that all central moments higher than the second were zero. This clearly is an unrealistic assumption and leads to a different equation for the covariance. Snyder's assumption leads to a

more symmetric looking result and tends to perform better in certain circumstances. Yet care must be taken in both cases since the resulting $\mathbf{P}(t)$ is not necessarily positive definite. The following corollary presents what is called the *first-order linearized estimate*.

COROLLARY 1.1 Let $\mathbf{x}(t)$ be an $(n \times 1)$-vector Markov process given by

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\, dt + d\mathbf{n} \tag{1.71}$$

and let the measurement be an $(m \times 1)$-vector process

$$d\mathbf{y} = \mathbf{h}(\mathbf{x}, t)\, dt + d\mathbf{w} \tag{1.72}$$

Then $\hat{\mathbf{x}}(t) = E[\mathbf{x}|O_{t_0,t}]$ is approximately given by

$$\frac{d\mathbf{x}^*}{dt} = \mathbf{f}(\mathbf{x}^*, t) + \mathbf{P}^*(t)\mathbf{C}(\mathbf{x}^*, t)\mathbf{R}^{-1}(t)[\mathbf{z}(t) - \mathbf{h}(\mathbf{x}^*, t)] \tag{1.73}$$

where $\mathbf{P}^*(t)$ is the $n \times n$ conditional covariance matrix, and it is also approximately given by

$$\frac{d\mathbf{P}^*(t)}{dt} = \mathbf{A}(\mathbf{x}^*, t)\mathbf{P}^*(t) + \mathbf{P}^*(t)\mathbf{A}^T(\mathbf{x}^*, t) + \mathbf{Q}(t) - \sum_{i=0}^{n} \mathbf{P}^*(t)\mathbf{G}_i(t)\mathbf{P}^*(t) \tag{1.74}$$

where

$$\mathbf{G}_0(t) = \mathbf{C}^T(\mathbf{x}^*, t)\mathbf{R}^{-1}(t)\mathbf{C}(\mathbf{x}^*, t) \tag{1.75}$$

and for $i > 0$,

$$\mathbf{G}_i(t) = -\mathbf{F}_i(\mathbf{x}^*, t)\, \boldsymbol{\gamma}_i^T \mathbf{R}^{-1}(t)[\mathbf{z}(t) - \mathbf{h}(\mathbf{x}^*, t)] \tag{1.76}$$

The presence of the measurement in the covariance equation is extremely significant. Had we linearized both system and measurement first, this term would not have appeared. Its appearance alters both the filter and covariance matrix. The important fact to note is that the covariance equation is affected directly by its measurements only in the case of a nonlinear measurement. It is indirectly affected by all the measurements, because the expansion points depend upon past measurements.

We can also note that the estimate and covariance equations are coupled. Equations (1.73) and (1.74) completely describe the estimator. One should carefully note that both the equation for the estimate propagation (1.73) and the equation for covariance propagation (1.74) depend upon the data. This is what we mean by "coupled equations." Furthermore, the covariance equation depends upon the estimate equation. We shall see that for a linear state system and linear measurements, these dependencies vanish.

The initial conditions for these equations are quite simple. For $\mathbf{x}^*(t_0)$ we merely use the expected value of the system a time $t = t_0$. Such a value could

follow directly from the Fokker-Planck equation. The initial value of $\mathbf{P}^*(t)$, $\mathbf{P}^*(t_0)$, is determined by our initial uncertainty of $\mathbf{x}^*(t_0)$. For example, if $\mathbf{x}^*(t_0)$ is perfectly known, then $\mathbf{P}^*(t_0)=\mathbf{0}$, the zero matrix.

We also note that for the case of both a linear system and linear measurements

$$\mathbf{x}^*(t) \to \bar{\mathbf{x}}(t) \tag{1.77}$$

and

$$\mathbf{P}^*(t) \to \mathbf{P}(t) \tag{1.78}$$

Where $\bar{\mathbf{x}}(t)$ is the MMSE estimate. To date there has been little work relating the error due to linearization. One should be careful since $\mathbf{P}^*(t)$ is not the real covariance matrix but only one due to a linearization. $\mathbf{P}^*(t)$ as given by (1.74) is not even the covariance of the error with a linearized estimate since to get (1.74), further linearization was necessary. Thus, extreme care should be taken when evoking estimation performance from (1.74). Another point to note is that we continually expand the results about the estimate. This may not be an optimal point. This will be discussed in the following section.

We would now like to discuss three cases where simplifications lead to drastic reductions. The first is when we have linear measurements. In many practical instances this is what occurs. That is, we have access to measurements linearly dependent upon the state variables.

The second case is for a linear system with nonlinear measurements. This is quite common when one analyzes communication systems. For example, a message may be assumed to be a Gaussian random process with a specified power spectrum. Such a process can be generated by a linear system excited by white noise. This signal may then be nonlinearly modulated (FM) and received with additive noise. These cases are studied by Snyder [1–3].

The third case is the most classical. It is a linear system with linear measurements. This gives the classical continuous-time version of the Kalman-Bucy filter (Meditch [2], Van Trees [1]). We have already derived the Kalman-Bucy filter for a discrete-time system in Chapter 4. For this third case, we now have an exact optimum MMSE estimate and the performance as indicated by $\mathbf{P}(t)$ is also exact.

CASE I—*Linear measurement-nonlinear system*: Assume that the system equation is the nonlinear process already discussed. Let the measurement nonlinearity be given by

$$\mathbf{h}(\mathbf{x}, t) = \mathbf{C}(t)\mathbf{x}(t) \tag{1.79}$$

where $\mathbf{C}(t)$ is an $m \times n$ matrix. Then,

$$\mathbf{C}(\bar{\mathbf{x}}, t) = \mathbf{C}(t) \tag{1.80}$$

$$\mathbf{F}_i(\hat{\mathbf{x}}, t) = \emptyset; \qquad \forall i \tag{1.81}$$

The optimum estimate is now generated by

$$\frac{d\mathbf{x}^*(t)}{dt} = \mathbf{f}(\mathbf{x}^*, t) + \mathbf{P}^*(t)\mathbf{C}^T(t)\mathbf{R}^{-1}(t)[\mathbf{z}(t) - \mathbf{C}(t)\mathbf{x}^*(t)] \tag{1.82}$$

And the variance equation is

$$\frac{d\mathbf{P}^*(dt)}{dt} = \mathbf{A}^T(\mathbf{x}^*, t)\mathbf{P}^*(t) + \mathbf{P}^*(t)\mathbf{A}(\mathbf{x}^*, t) + \mathbf{Q}(t) - \mathbf{P}^*(t)\mathbf{C}^T(t)\mathbf{R}(t)^{-1}(t)\mathbf{C}(t)\mathbf{P}^*(t) \tag{1.83}$$

We now consider the case of linear system but with a nonlinear measurement.

CASE II—*Linear system-nonlinear measurement*: Assume that the measurements are given by

$$\mathbf{z}(t) = \mathbf{h}(\mathbf{x}, t) + \dot{\mathbf{w}}(t) \tag{1.84}$$

Let the system be given by

$$\dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t) + \dot{\mathbf{n}}(t) \tag{1.85}$$

where $\dot{\mathbf{n}}(t)$ is a white noise and $\mathbf{A}(t)$ is an $n \times n$ time-varying matrix. Then,

$$\frac{d\mathbf{x}^*(t)}{dt} = \mathbf{A}(t)\mathbf{x}^*(t) + \mathbf{P}^*(t)\mathbf{C}^T(\mathbf{x}^*, t)\mathbf{R}^{-1}[\mathbf{z}(t) - \mathbf{h}(t, \mathbf{x}^*)] \tag{1.86}$$

where $\mathbf{C}(\mathbf{x}^*, t)$ is as defined before.

The variance becomes

$$\frac{d\mathbf{P}^*(t)}{dt} = \mathbf{A}^T(t)\mathbf{P}^*(t) + \mathbf{P}^*(t)\mathbf{A}(t) + \mathbf{Q}(t) - \sum_{i=0}^{n} \mathbf{P}(t)\mathbf{G}_i(\mathbf{x}^*, t)\mathbf{P}(t) \tag{1.87}$$

CASE III *Linear system and measurement (Kalman-Bucy Filter)*: When we have both a linear system and a linear measurement, we have the solution for the problem posed by Kalman and Bucy. The model becomes

$$d\mathbf{x}(t) = \mathbf{A}(t)\mathbf{x}(t)\,dt + d\mathbf{n}(t) \tag{1.88}$$

The measurement is

$$d\mathbf{y}(t) = \mathbf{C}(t)\mathbf{x}(t)\,dt + d\mathbf{u}(t) \tag{1.89}$$

The estimate equation then is

$$\frac{d\mathbf{x}^*(t)}{dt} = \mathbf{A}(t)\mathbf{x}^*(t) + \mathbf{P}^*(t)\mathbf{C}^T(t)\mathbf{R}^{-1}(t)[\mathbf{z}(t) - \mathbf{C}(t)\mathbf{x}^*(t)] \tag{1.90}$$

and the variance equation is

$$\frac{d\mathbf{P}^*(t)}{dt} = \mathbf{A}^T(t)\mathbf{P}^*(t) + \mathbf{P}^*(t)\mathbf{A}(t) + \mathbf{Q}(t) - \mathbf{P}^*(t)\mathbf{C}^T(t)\mathbf{R}^{-1}(t)\mathbf{C}(t)\mathbf{P}^*(t) \tag{1.91}$$

and (1.91) is the Ricatti equation, whose solutions are well known. But, since no approximations were necessary in (1.90) or (1.91), $\mathbf{x}^*(t)$ is actually the optimum estimate and $\mathbf{P}^*(t)$ is actually $\mathbf{P}(t)$ the real covariance matrix. Thus,

$$\hat{\mathbf{x}}(t) = \mathbf{x}^*(t) \tag{1.92}$$

$$\mathbf{P}(t) = \mathbf{P}^*(t) \tag{1.93}$$

Note that with the Kalman-Bucy filter the covariance $\mathbf{P}(t)$ can be obtained a priori. That is, (1.91) can be solved and it can tell how well an estimate can be obtained before any measurement is made. Thus, the estimate equation is all that remains to be solved when the data is obtained.

The initial conditions are satisfied if we let $\mathbf{x}^*(t_0)$ or $\hat{\mathbf{x}}(t_0)$ be the estimate at time $t = t_0$. This value is based upon some a priori knowledge. Likewise, $\mathbf{P}(t_0)$ is the initial covariance matrix of the initial estimate.

The linearized estimation for the case of Poisson measurements can also be obtained by similar techniques. Recall that form the preceding chapter the propagation equation for the conditional density (Snyder's equation) yields

$$dp = -\sum_{i=1}^{n} \frac{\partial (fp)_i \, dt}{\partial u_i} + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} Q_{ij} \frac{\partial^2 p \, dt}{\partial u_i \, \partial u_j} + p \sum_{i=1}^{m} (\boldsymbol{\lambda}(t) - \hat{\boldsymbol{\lambda}}(t))^T \boldsymbol{\gamma}_i (\hat{\boldsymbol{\lambda}}^T(t) \boldsymbol{\gamma}_i)^{-1} (d\mathbf{N}(t) - \hat{\boldsymbol{\lambda}}(t) \, dt)^T \boldsymbol{\gamma}_i \tag{1.94}$$

where

$$p = p_{\mathbf{x}}(\mathbf{u}, t \mid O_{t_0,t}) \tag{1.95}$$

and $\hat{\boldsymbol{\lambda}}(t)$ is $E[\boldsymbol{\lambda}(t) \mid O_{t_0,t}]$.

Now given (1.94) the optimum estimate could easily be obtained. Unfortunately, such a procedure is quite impossible analytically in general, so that we must resort to some approximation techniques. Let us expand $\mathbf{f}(\mathbf{x})$ and $\boldsymbol{\lambda}(\mathbf{x})$ about their optimum points. That is, let

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(\hat{\mathbf{x}}) + \mathbf{A}(\hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}}) + \sum_{i=1}^{m} \boldsymbol{\gamma}_i (\mathbf{x} - \hat{\mathbf{x}})^T \mathbf{B}_i (\mathbf{x} - \hat{\mathbf{x}}) \tag{1.96}$$

$$\boldsymbol{\lambda}(\mathbf{x}) = \boldsymbol{\lambda}(\hat{\mathbf{x}}) + \mathbf{D}(\hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}}) + \frac{1}{2} \sum_{i=1}^{m} \boldsymbol{\gamma}_i (\mathbf{x} - \hat{\mathbf{x}})^T \mathbf{E}_i (\mathbf{x} - \hat{\mathbf{x}}) + \cdots \tag{1.97}$$

where as before

$$\mathbf{A}(\hat{\mathbf{x}}) = \left[\begin{array}{ccc} \frac{\partial f_1}{\partial x_1} & & \frac{\partial f_1}{\partial x_n} \\ & & \\ \frac{\partial f_n}{\partial x_1} & & \frac{\partial f_n}{\partial x_n} \end{array} \right]_{\mathbf{x} = \hat{\mathbf{x}}} \tag{1.98}$$

$$\mathbf{D}(\hat{\mathbf{x}}) = \begin{bmatrix} \dfrac{\partial \lambda_1}{\partial x_1} & \dfrac{\partial \lambda_1}{\partial x_n} \\ \dfrac{\partial \lambda_m}{\partial x_1} & \dfrac{\partial \lambda_m}{\partial x_n} \end{bmatrix} \Bigg|_{\mathbf{x} = \hat{\mathbf{x}}} \tag{1.99}$$

$$\mathbf{B}_i = \begin{bmatrix} \dfrac{\partial^2 f_i}{\partial x_1 \, \partial x_1} & \dfrac{\partial^2 f_i}{\partial x_n \, \partial x_1} \\ \dfrac{\partial^2 f_i}{\partial x_n \, \partial x_1} & \dfrac{\partial^2 f_i}{\partial x_n \, \partial x_n} \end{bmatrix}_{\mathbf{x} = \hat{\mathbf{x}}} \tag{1.100}$$

$$\mathbf{E}_i = \begin{bmatrix} \dfrac{\partial^2 \lambda_i}{\partial x_1 \, \partial x_1} & \dfrac{\partial^2 \lambda_i}{\partial x_1 \, \partial x_n} \\ \dfrac{\partial^2 \lambda_i}{\partial x_n \, \partial x_1} & \dfrac{\partial^2 \lambda_i}{\partial x_n \, \partial x_n} \end{bmatrix}_{\mathbf{x} = \hat{\mathbf{x}}} \tag{1.101}$$

Now $\hat{\mathbf{x}}(t)$ is generated by multiplying (1.94) by $\mathbf{u}$ and integrating ($\mathbf{u}$ is a $n \times 1$ vector) giving

$$d\hat{\mathbf{x}}(t) = E[\mathbf{f}(\mathbf{x}) \mid O_{t_0,t}]\, dt + \sum_{i=1}^{m} E[(\mathbf{x}(t) - \hat{\mathbf{x}}(t))\boldsymbol{\lambda}^T(\mathbf{x}) \mid O_{t_0,t}] \cdot \boldsymbol{\gamma}_i(\hat{\boldsymbol{\lambda}}^T\boldsymbol{\gamma}_i)^{-1} \cdot (d\mathbf{N}(t) - \hat{\boldsymbol{\lambda}}\, dt)^T\boldsymbol{\gamma}_i \tag{1.102}$$

We can now use the linearizations in this equation. Let us first define the covariance matrix $\mathbf{P}(t)$:

$$\mathbf{P}(t) = E[(\mathbf{x}(t) - \hat{\mathbf{x}}(t))(\mathbf{x}(t) - \hat{\mathbf{x}}(t))^T \mid O_{t_0,t}] \tag{1.103}$$

and let P_{ij} be the ijth component of P. Now, using (1.96) and (1.97) in (1.102), we obtain

$$d\mathbf{x}^*(t) = \mathbf{f}(\mathbf{x}^*)\, dt + \sum_{i=1}^{m} \mathbf{P}(t)\mathbf{D}(\mathbf{x}^*)\boldsymbol{\gamma}_i(\hat{\boldsymbol{\lambda}}^T\boldsymbol{\gamma}_i)^{-1}(d\mathbf{N}(t) - \hat{\boldsymbol{\lambda}}dt)^T\boldsymbol{\gamma}_i \tag{1.104}$$

Snyder [4] includes the $\mathbf{B}_i$ terms of $\mathbf{f}(\mathbf{x})$, but for our purposes, this will not be necessary. Now (1.104) gives the propagation equation for the linearized estimate.

We now seek to evaluate $\mathbf{P}(t)$. In so doing, we shall follow the technique used in the Gaussian measurement case. We want to obtain

$$\begin{aligned} d\mathbf{P}(t) &= \mathbf{P}(t + dt) - \mathbf{P}(t) \\ &= E[(\mathbf{u} - \hat{\mathbf{x}}(t + dt))(\mathbf{u} - \hat{\mathbf{x}}(t + dt))^T \mid O_{t_0,t+dt}] \\ &\quad - E[(\mathbf{u} - \hat{\mathbf{x}}(t))(\mathbf{u} - \hat{\mathbf{x}}(t))^T \mid O_{t_0,t}] \end{aligned} \tag{1.105}$$

Let

$$d\hat{\mathbf{x}} = \hat{\mathbf{x}}(t + dt) - \hat{\mathbf{x}}(t) \tag{1.106}$$

Then

$$\mathbf{P}(t+dt) = E[(\mathbf{u} - \hat{\mathbf{x}}(t) - d\hat{\mathbf{x}})(\mathbf{u} - \hat{\mathbf{x}}(t) - d\hat{\mathbf{x}})^T \mid O_{t_0,t+dt}]$$

$$= \int (\mathbf{u} - \hat{\mathbf{x}}(t))(\mathbf{u} - \hat{\mathbf{x}}(t))^T p_{\mathbf{x}}(\mathbf{u}, t+dt \mid O_{t_0,t+dt})\, d\mathbf{u} + d\hat{\mathbf{x}}\, d\mathbf{x}^T \quad (1.107)$$

$$- \int (\mathbf{u} - \hat{\mathbf{x}}(t)) d\hat{\mathbf{x}}^T p_{\mathbf{x}}(\mathbf{u}, t+dt \mid O_{t_0,t+dt})\, d\mathbf{u}$$

$$- \int d\hat{\mathbf{x}} (\mathbf{u} - \hat{\mathbf{x}}(t))^T p_{\mathbf{x}}(\mathbf{u}, t+dt \mid O_{t_0,t+dt})\, d\mathbf{u} \quad (1.108)$$

The last two integrals are both equal to $d\hat{\mathbf{x}}\ d\hat{\mathbf{x}}^T$ as was shown in Lemma 1.1. The first term can be evaluated since we know $p_{\mathbf{x}}(\mathbf{u}, t+dt \mid O_{t_0,t+dt})$ in terms of

$$p_{\mathbf{x}}(\mathbf{u}, t \mid O_{t_0,t})$$

We know that from Snyder's equation that

$$p_{\mathbf{x}}(\mathbf{u}, t+dt \mid O_{t_0,t+dt})$$
$$= p_{\mathbf{x}}(\mathbf{u}, t \mid O_{t_0,t}) + L^+[p_{\mathbf{x}}(\mathbf{u}, t \mid O_{t_0,t})]dt + p_{\mathbf{x}}(\mathbf{u}, t \mid O_{t_0,t})m \quad (1.109)$$

Here L^+ represents the forward Kolmogorov operator and m is

$$m = \sum_{i=1}^{m} (\boldsymbol{\lambda}(t) - \hat{\boldsymbol{\lambda}}(t))^T \boldsymbol{\gamma}_i (\hat{\boldsymbol{\lambda}}^T \boldsymbol{\gamma}_i)^{-1} (d\mathbf{N} - \hat{\boldsymbol{\lambda}}\, dt)^T \boldsymbol{\gamma}_i \quad (1.110)$$

It is a simple matter to show that

$$d\mathbf{P} + d\hat{\mathbf{x}}\, d\hat{\mathbf{x}}^T = \mathbf{A}^T(t)\mathbf{P}(t)\, dt + \mathbf{P}(t)\mathbf{A}(t)dt + \mathbf{Q}(t)\, dt$$
$$+ \sum_i \mathbf{P}(t)\mathbf{E}_i\mathbf{P}(t)\hat{\lambda}_i^{-1}\boldsymbol{\gamma}_i^T(d\mathbf{N} - \hat{\boldsymbol{\lambda}}\, dt) \quad (1.111)$$

This follows directly from the analysis done to prove Theorem 1.1. We must now calculate $d\hat{\mathbf{x}}\ d\hat{\mathbf{x}}^T$. To order dt,

$$d\hat{\mathbf{x}}\, d\hat{\mathbf{x}}^T = \sum_{i=1}^{m}\sum_{j=1}^{m} \mathbf{P}(t)\mathbf{D}\boldsymbol{\gamma}_i(\boldsymbol{\lambda}^{*T}\boldsymbol{\gamma}_i)^{-1}\boldsymbol{\gamma}_i^T d\mathbf{N}(t) d\mathbf{N}^T(t)\boldsymbol{\gamma}_j(\boldsymbol{\lambda}^{*T}\boldsymbol{\gamma}_j)^{-1}\boldsymbol{\gamma}_j^T\mathbf{D}^T\mathbf{P}(t) \quad (1.112)$$

where $\boldsymbol{\lambda}^*$ is the evaluation of $\hat{\boldsymbol{\lambda}}$ at $\mathbf{x}^*$.
But

$$\boldsymbol{\gamma}_i^T\, d\mathbf{N}(t)\, d\mathbf{N}^T(t)\boldsymbol{\gamma}_j = dN_i(t)\delta_{ij} \quad (1.113)$$

so that

$$d\hat{\mathbf{x}}\, d\hat{\mathbf{x}}^T = \sum_{i=1}^{m} \mathbf{P}(t)\frac{\mathbf{D}}{\lambda_i^*}\boldsymbol{\gamma}_i\boldsymbol{\gamma}_u^T\frac{\mathbf{D}^T}{\lambda_i^*}\mathbf{P}(t)\boldsymbol{\gamma}_i^T d\mathbf{N}(t) \quad (1.114)$$

But

$$\frac{\mathbf{D}}{\lambda_i^*} = \begin{bmatrix} \frac{\partial \lambda_i}{\partial x_1} \\ \vdots \\ \frac{\partial \lambda_i}{\partial x_n} \end{bmatrix} \quad (1.115)$$

Now let

$$\mathbf{H}_i = \begin{bmatrix} \dfrac{\partial^2 ln\ \lambda_i}{\partial x_1\ \partial x_1} & \dfrac{\partial^2 ln\ \lambda_i}{\partial x_1\ \partial x_n} \\ \dfrac{\partial^2 ln\ \lambda_i}{\partial x_n\ \partial x_1} & \dfrac{\partial^2 ln\ \lambda_i}{\partial x_n\ \partial x_n} \end{bmatrix} = \frac{\mathbf{D}\boldsymbol{\gamma}_i}{\lambda_i^*}\left(\frac{\mathbf{D}\boldsymbol{\gamma}_i}{\lambda_i^*}\right)^T + \frac{\mathbf{E}_i}{\lambda_i^*} \tag{1.116}$$

Thus, the covariance equation becomes

$$\frac{d\mathbf{P}^*(t)}{dt} = \mathbf{A}^T(t)\mathbf{P}^*(t) + \mathbf{P}^*(t)\mathbf{A}(t) + \mathbf{Q}(t) + \sum_{i=1}^{m} \mathbf{P}^*(t)\mathbf{H}_i\mathbf{P}(t)\boldsymbol{\gamma}_i^T \frac{d\mathbf{N}(t)}{dt} - \sum_{i=1}^{m} \mathbf{P}^*(t)\mathbf{E}_i\mathbf{P}^*(t) \tag{1.117}$$

One can immediately note that

$$\lambda_i^{*-1}\boldsymbol{\gamma}_i^T\boldsymbol{\lambda}^* = 1 \tag{1.118}$$

which led to the reduction in (1.117). This result is summarized in the following theorem.

THEOREM 1.2

Let $\mathbf{x}(t)$ be an $(n \times 1)$-vector Markov process given by

$$d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}, t)\,dt + d\mathbf{n}(t) \tag{1.119}$$

The measurement is an $m \times 1$ unit-jump Poisson process with a vector rate parameter $\boldsymbol{\lambda}(\mathbf{x}, t)$. The linearized first-order MMSE equations are given by

$$\frac{d\mathbf{x}^*}{dt} = \mathbf{f}(\mathbf{x}^*, t) + \sum_{i=1}^{m} \mathbf{P}^*(t)\mathbf{D}(\mathbf{x}^*, t)\boldsymbol{\gamma}_i(\boldsymbol{\lambda}^{*T}\boldsymbol{\gamma}_i)^{-1} \cdot \left(\frac{d\mathbf{N}(t)}{dt} - \boldsymbol{\lambda}(\mathbf{x}^*, t)\right)^T\boldsymbol{\gamma}_i \tag{1.120}$$

and $\mathbf{P}^*(t)$, the linearized covariance, is given by

$$\frac{d\mathbf{P}^*(t)}{dt} = \mathbf{A}^T(\mathbf{x}^*, t)\mathbf{P}^*(t) + \mathbf{P}^*(t)\mathbf{A}(\mathbf{x}^*, t) + \mathbf{Q}(t) + \sum_{i=1}^{m} \mathbf{P}^*(t)\mathbf{H}_i(\mathbf{x}^*, t)\mathbf{P}^*(t)\frac{d\mathbf{N}(t)}{dt} - \sum_{i=1}^{m} \mathbf{P}^*(t)\mathbf{E}_i(\mathbf{x}^*, t)\mathbf{P}^*(t) \tag{1.121}$$

where $\mathbf{A}$, $\mathbf{D}$, $\mathbf{H}_i$, and $\mathbf{E}_i$ are given in (1.98), (1.99), (1.116), and (1.101), respectively. Also, $\hat{\mathbf{x}}(t_0)$ and $\mathbf{P}^*(t_0)$ are assumed known.

Extensions of the previous theorems to include the second-order effects follow easily from the analysis of the Gaussian case. The simplifications of the Poisson measurement case are not as broad and encompasing as the Kalman-Bucy equations for the Gaussian measurement. Some of these have been dis-

cussed by Snyder [4, 5], Evans [2], and J. R. Clark. The following two examples follow from McGarty [2, 3] and are based upon a meteorological experiment to determine atmospheric structure. The first example is for a Gaussian measurement for a parameter estimation problem. The second case is for parameter estimation from Poisson measurements. The relationship between the problems is that the measurement nonlinearities are identical.

Example. In an analysis of data from a meteorological satellite experiment, it is found that the light intensity measured by a photodetector is related to the density of particles along its line of sight in an experimental relationship. For each of m wavelengths, a measurement $z_i(t)$ is given by

$$z_i(t) = h_i(\mathbf{x}, t) + \dot{w}_i(t) \tag{1.122}$$

where $\mathbf{x}$ is an $n \times 1$ vector representing the density at discrete points along the line of sight. The signal $z_i(t)$ is the voltage at the output of a photodetector tuned to the ith wavelength. The nonlinear function is

$$h_i(\mathbf{x}, t) = h_{0i} \exp[-\mathbf{g}_i^T(t)\mathbf{x}(t)] \tag{1.123}$$

$\mathbf{g}_i(t)$ is an $n \times 1$ vector that is determined by the geometry of the satellite and the scanning technique involved. The noise $\dot{\mathbf{w}}(t)$ is an $m \times 1$ white noise vector with constant spectral height R (see Figure 6. 1 for the geometry).

It is assumed that the state is a constant parameter that has a random initial value. This means that we assume that the density is constant during the scanning time. Thus, the state equation is simply

Figure 6.1 Geometry of satellite and scan.

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{0} \tag{1.124}$$

The linearized estimation equation is

$$\frac{d\mathbf{x}^*(t)}{dt} = \mathbf{P}^*(t)\mathbf{C}^T(\mathbf{x}^*(t), t)\mathbf{R}^{-1}[\mathbf{z}(t) - \mathbf{h}(\mathbf{x}^*, t)] \tag{1.125}$$

where the matrix $\mathbf{R}$ is the noise power spectral matrix associated with the measurement. We shall assume that $\mathbf{R}$ has only nonzero diagonal terms. By having zero off diagonal terms, we are assuming that the noise is uncorrelated at different wavelengths.

The $\mathbf{h}$ vector is the received signal vector without noise. It is defined as

$$\mathbf{h}(\mathbf{x}, t) = \mathbf{H}\begin{bmatrix} h_1(\mathbf{x}, t) \\ \vdots \\ h_m(\mathbf{x}, t) \end{bmatrix} \tag{1.126}$$

where $\mathbf{H}$ is an $m \times m$ diagonal matrix. The $\mathbf{H}$ matrix is a constant matrix and converts photometric units to electrical units.

The $\mathbf{C}$ matrix is a gradient matrix:

$$\mathbf{C}(\mathbf{x}(t), t) = \begin{bmatrix} \dfrac{\partial h_1}{\partial x_1} & \cdots & \dfrac{\partial h_1}{\partial x_n} \\ \vdots & & \vdots \\ \dfrac{\partial h_m}{\partial x_1} & \cdots & \dfrac{\partial h_m}{\partial x_n} \end{bmatrix} \tag{1.127}$$

Clearly, $\mathbf{C}$ is an $m \times n$ matrix. For the nonlinearity in question,

$$\frac{\partial h_i}{\partial x_j} = -g_{ij}h_i \tag{1.128}$$

where g_{ij} is the jth component of $\mathbf{g}_i$. The linearized covariance $\mathbf{P}^*(t)$ satisfies the following Riccati equation:

$$\frac{d\mathbf{P}^*(t)}{dt} = -\mathbf{P}^*(t)\mathbf{C}^T(\mathbf{x}^*, t)\mathbf{R}^{-1}\mathbf{C}(\mathbf{x}^*, t)\mathbf{P}^*(t) - \sum_{i=1}^{m} \alpha_i \mathbf{P}^*(t)\mathbf{F}_i(\mathbf{x}^*, t)\mathbf{P}^*(t) \tag{1.129}$$

where

$$\alpha_i = -\gamma_i^T \mathbf{R}^{-1}(\mathbf{z}(t) - \mathbf{h}(\mathbf{x}^*(t), t)) \tag{1.130}$$

$$\mathbf{F}_i(\mathbf{x}^*, t) = \begin{bmatrix} \dfrac{\partial^2 h_i}{\partial x_1\, \partial x_1} & \cdots & \dfrac{\partial^2 h_i}{\partial x_1\, \partial x_n} \\ \vdots & & \vdots \\ \dfrac{\partial^2 h_i}{\partial x_n\, \partial x_1} & \cdots & \dfrac{\partial^2 h_i}{\partial x_n\, \partial x_n} \end{bmatrix} \tag{1.131}$$

Also, α_i can be expressed as

$$\alpha_i = -\frac{(z_i(t) - h_i(\mathbf{x}^*(t), t))}{r_i} \tag{1.132}$$

where r_i is the ith diagonal component of $\mathbf{R}$, and $\mathbf{R}$ is written as

$$\mathbf{R} = \begin{bmatrix} r_1 & 0 & 0 \\ 0 & r_2 & \vdots \\ \vdots & & \ddots \\ 0 & \cdots & r_m \end{bmatrix} \tag{1.133}$$

γ_i is an $m \times 1$ vector with a 1 in the ith element written as

$$\gamma_i = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \; i\text{th} \tag{1.134}$$

Now, using (1.123) in (1.131), the $\mathbf{F}_i$ can be written as

$$\mathbf{F}_i(\mathbf{x}^*, t) = \begin{bmatrix} g_{i1}g_{i1}h_i \cdots g_{i1}g_{in}h_i \\ \vdots \qquad\qquad \vdots \\ g_{in}g_{i1}h_i \cdots g_{in}g_{in}h_i \end{bmatrix} \tag{1.135}$$

Let us consider the matrix $\mathbf{B}^T\mathbf{R}^{-1}\mathbf{B}$. Clearly, $\mathbf{R}^{-1}$ can be written as

$$\mathbf{R}^{-1} = \begin{bmatrix} m_1 & 0 & 0 & 0 \\ 0 & m_2 & 0 & 0 \\ 0 & 0 & m_3 & \\ & & & \ddots \\ & & & m_m \end{bmatrix} \tag{1.136}$$

where

$$m_i = \frac{1}{r_i} \tag{1.137}$$

Also

$$\mathbf{B}^T\mathbf{R}^{-1} = \begin{bmatrix} \dfrac{\partial h_1}{\partial x_1} & \cdots & \dfrac{\partial h_m}{\partial x_1} \\ \vdots & & \vdots \\ \dfrac{\partial h_1}{\partial x_n} & \cdots & \dfrac{\partial h_m}{\partial x_1} \end{bmatrix} \begin{bmatrix} m_1 & & 0 \\ & \ddots & \\ 0 & & m_m \end{bmatrix}$$

$$= \begin{bmatrix} m_1\dfrac{\partial h_1}{\partial x_1} & \cdots & m_m\dfrac{\partial h_m}{\partial x_1} \\ \vdots & & \vdots \\ m_1\dfrac{\partial h_m}{\partial x_n} & \cdots & m_m\dfrac{\partial h_m}{\partial x_n} \end{bmatrix} \tag{1.138}$$

and finally we can combine everything to read

$$\mathbf{B}^T\mathbf{R}^{-1}\mathbf{B} = \begin{bmatrix} \sum_{i=1}^{m} m_i \frac{\partial h_i}{\partial x_1}\frac{\partial h_i}{\partial x_1} & \cdots & \sum_{i=1}^{m} m_i \frac{\partial h_i}{\partial x_1}\frac{\partial h_i}{\partial x_n} \\ \vdots & & \vdots \\ \sum_{i=1}^{m} m_i \frac{\partial h_i}{\partial x_n}\frac{\partial h_i}{\partial x_1} & \cdots & \sum_{i=1}^{m} m_i \frac{\partial h_i}{\partial x_n}\frac{\partial h_i}{\partial x_n} \end{bmatrix} \tag{1.139}$$

Let us concentrate on the terms involving the $\mathbf{F}_i$ matrices:

$$\sum_{i=1}^{m} \alpha_i \mathbf{P}(t)\mathbf{F}_i\mathbf{P}(t) = \mathbf{P}(t)\left(\sum_{i=1}^{m} \alpha_i \mathbf{F}_i\right)\mathbf{P}(t) \tag{1.140}$$

But we can also write this as

$$\sum_{i=1}^{m} \alpha_i \mathbf{F}_i = \begin{bmatrix} \sum \alpha_i \frac{\partial^2 h_i}{\partial x_1 \partial x_1} & \cdots & \sum \alpha_i \frac{\partial^2 h_i}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \sum \alpha_i \frac{\partial^2 h_i}{\partial x_n \partial x_1} & \cdots & \sum \alpha_i \frac{\partial^2 h_i}{\partial x_n \partial x_n} \end{bmatrix} \tag{1.141}$$

By the functional structure of h_i we can relate the second partial derivative to the first partial derivative by

$$\frac{\partial^2 h_i}{\partial x_j \partial x_k} = \frac{1}{h_i}\frac{\partial h_i}{\partial x_j}\frac{\partial h_i}{\partial x_k} \tag{1.142}$$

This is an immediate consequence of (1.123). Now let us define a function β_i given by

$$\beta_i = \frac{\alpha_i}{h_i} \tag{1.143}$$

Then (1.141) becomes

$$\sum \alpha_i \mathbf{F}_i = \begin{bmatrix} \sum \beta_i \frac{\partial h_i}{\partial x_1}\frac{\partial h_i}{\partial x_1} & \cdots & \sum \beta_i \frac{\partial h_i}{\partial x_1}\frac{\partial h_i}{\partial x_n} \\ \vdots & & \vdots \\ \sum \beta_i \frac{\partial h_i}{\partial x_n}\frac{\partial h_i}{\partial x_1} & \cdots & \sum \beta \frac{\partial h_i}{\partial x_n}\frac{\partial h_i}{\partial x_n} \end{bmatrix} \tag{1.144}$$

Then, using (1.144) and (1.139) in (1.129), we find that

$$\frac{d\mathbf{P}^*(t)}{dt} = -\mathbf{P}^*(t)\mathbf{U}\mathbf{P}^*(t) \tag{1.145}$$

where $\mathbf{U}$ is given by

$$\mathbf{U} = \begin{bmatrix} \sum_{i=1}^{m}(m_i + \beta_i)\frac{\partial h_i}{\partial x_1}\frac{\partial h_i}{\partial x_1} & \cdots & \sum_{i=1}^{m}(m_i + \beta_i)\frac{\partial h_i}{\partial x_1}\frac{\partial h_i}{\partial x_n} \\ \vdots & & \vdots \\ \sum_{i=1}^{m}(m_i + \beta_i)\frac{\partial h_i}{\partial x_n}\frac{\partial h_i}{\partial x_1} & \cdots & \sum_{i=1}^{m}(m_i + \beta_i)\frac{\partial h_i}{\partial x_n}\frac{\partial h_i}{\partial x_n} \end{bmatrix} \tag{1.146}$$

But we can see that if we define a new matrix $\mathbf{S}^{-1}$, where

$$\mathbf{S}^{-1} = \begin{bmatrix} m_1 + \beta_1 & 0 & 0 \\ 0 & m_2 + \beta_2 & \\ & & \ddots \\ & & m_m + \beta_m \end{bmatrix} \tag{1.147}$$

then $\mathbf{U}$ becomes

$$\mathbf{U} = \mathbf{C}^T(x^*, t)\mathbf{S}^{-1}\mathbf{C}(x^*, t) \tag{1.148}$$

Thus, $\mathbf{S}$ is a new noise matrix. The elements of $\mathbf{S}$ are

$$\begin{aligned} \frac{1}{m_i + \beta_i} &= \frac{1}{m_i + (\alpha_i/h_i)} \\ &= \frac{h_i}{m_i h_i + \alpha_i} \\ &= \frac{r_i h_i}{h_i + r_i \alpha_i} \end{aligned} \tag{1.149}$$

But from (1.132)

$$S_i = \frac{1}{m_i + \beta_i} = \frac{r_i h_i}{h_i + (h_i - z_i)} \tag{1.150}$$

Now the new matrix $\mathbf{S}$ can be written in terms of the old $\mathbf{R}$ if we introduce the new matrix Δ such that

$$\mathbf{S} = \mathbf{R}\Delta = \Delta\mathbf{R} \tag{1.151}$$

Figure 6.2 Constituent densities versus altitude.

Figure 6.3 (*a*) $\Sigma_{ii}(t)/\Sigma_{ii}(0)$ for components of the state vector as a function of tangent height; nonlinear term included; signal-to-noise ratio 33/1. (*b*) $\Sigma_{ii}(t)/\Sigma_{ii}(0)$ for components of the state vector as a function of tangent height: linear term; signal-to-noise ratio 33/1. (*c*) $\Sigma_{ii}(t)/\Sigma_{ii}(0)$ for components of the state vector as a function of tangent height; nonlinear term included; signal-to-noise ratio 100/1. (*d*) $\Sigma_{ii}(t)/\Sigma_{ii}(0)$ for components of the state vector as a function of tangent height; linear term; signal-to-noise ratio 100/1.

$$\Delta = \begin{bmatrix} \dfrac{h_1}{h_1 + (h_1 - z_1)} & & 0 \\ & \ddots & \\ 0 & & \dfrac{h_m}{h_m + (h_m - z_m)} \end{bmatrix} \tag{1.152}$$

Now a great deal of insight into the nonlinear effects can be gained from this. First, Δ acts to modulate the noise matrix. If our estimates are good, that is, if

$$h_m \approx z_m$$

then Δ is close to the identity matrix and our system acts just as it did before. But if

$$h_m \neq z_m$$

and the inequality is strong, then Δ tends to change in such a fashion to drive the system faster.

A specific example of an analysis using this technique appears in McGarty [2]. The results are shown in Figure 6.2 and 6.3. In Figure 6.2 we plot the profiles that were to be estimated on a pointwise basis; that is, estimates of both ozone and neutral density were to be obtained at a fixed number of altitudes. Using the filter developed in this example, estimates were obtained and the calculated covariances (normalized to their peak value) are shown in Figure 6.3 for both a first-order Kushner–Stratonovich (K–S) filter [extended Kalman–Bucy (K–B) filter using last estimate as expansion point] and a complete second-order Kushner-Stratonovich K–S filter for two different signal-to-noise conditions.

The following example considers parameter estimation when the measurement is comprised of m simple Poisson processes. It uses the same nonlinearity of the preceding example in the analysis.

Example. This example is an extension of the previous one. Again, we seek to estimate a random parameter, so the state equation becomes

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{0} \tag{1.153}$$

But this time the light that is recevied is of such a low level that what is detected is only a single photon at a time. The arrival rate of these photons though depends on the average light intensity. This intensity at wavelength i is $I_i(t)$ and is given by

$$I_i(t) = I_{0i} \exp[-\mathbf{g}_i^T(t)\mathbf{x}(t)] \tag{1.154}$$

where I_{0i} is the source intensity at wavelength i and $\mathbf{g}_i(t)$ is an $n \times 1$ vector that depends on the geometry. Thus, $\lambda_i(x, t)$ depends upon the state as follows:

$$\lambda_i(\mathbf{x}, t) = \lambda_{i0} \exp[-\mathbf{g}_i^T(t)\mathbf{x}(t)] \tag{1.155}$$

The estimate is

$$\frac{d\mathbf{x}^*(t)}{dt} = \sum_{i=1}^{m} \mathbf{P}^*(t)\mathbf{D}\boldsymbol{\gamma}_i\lambda_i^{*-1}\boldsymbol{\gamma}_i^T\left(\frac{d\mathbf{N}}{dt} - \boldsymbol{\lambda}^*\right) \tag{1.156}$$

where the **D** matrix is given by

$$\mathbf{D} = \left[\begin{array}{cc} \frac{\partial\lambda_1}{\partial x_1} & \wedge \\ \wedge & \frac{\partial\lambda_m}{\partial x_n} \end{array}\right]_{\mathbf{x}=\mathbf{x}^*} = -\left[\begin{array}{cc} g_{11}(t)\lambda_1 & g_{1n}(t)\lambda_1 \\ g_{m1}(t)\lambda_m & g_{mn}(t)\lambda_m \end{array}\right]_{\mathbf{x}=\mathbf{x}^*} \tag{1.157}$$

and

$$\mathbf{D}\boldsymbol{\gamma}_i\lambda_i^{-1} = -\mathbf{g}_i(t) \tag{1.158}$$

Thus, we can reduce the estimate equation to the simple form

$$\frac{d\mathbf{x}^*(t)}{dt} = \mathbf{P}(t)\mathbf{R}(t)\left(\frac{d\mathbf{N}}{dt} - \boldsymbol{\lambda}^*(t)\right) \tag{1.159}$$

where $\mathbf{R}(t)$ is given by

$$\mathbf{R}(t) = -\sum_{i=1} \mathbf{g}_i\boldsymbol{\gamma}_i^T \tag{1.160}$$

or, in matrix form,

$$\mathbf{R}(t) = -\left[\begin{array}{cc} \lambda_{11}(t) & \gamma_{1m}(t) \\ g_{n1}(t) & \gamma_{nm}(t) \end{array}\right] \tag{1.161}$$

The covariance equation is simplified if we note that the matrix $\mathbf{H}_i$

$$\mathbf{H}_i = \left[\begin{array}{ccc} \frac{\partial^2 ln\lambda_i}{\partial x_1\,\partial x_1} & \cdots & \frac{\partial^2 ln\lambda_i}{\partial x_1\,\partial x_n} \\ \vdots & & \\ \frac{\partial^2 ln\lambda_i}{\partial x_n\,\partial x_1} & & \frac{\partial^2 ln\lambda i}{\partial x_n\,\partial x_n} \end{array}\right]_{\mathbf{x}=\mathbf{x}^*} = \mathbf{0} \tag{1.162}$$

since

$$ln\lambda_i = -\mathbf{g}_i^T(t)\mathbf{x}(t) \tag{1.163}$$

and the second partial derivatives are clearly all 0. The variance equation then becomes;

$$\frac{d\mathbf{P}(t)}{dt} = -\mathbf{P}(t)\mathbf{E}(t)\mathbf{P}(t) \tag{1.164}$$

where

$$\mathbf{E}(t) = \sum_{i=1}^{m} \mathbf{E}_i(t) \tag{1.165}$$

Now each $\mathbf{E}_i$ is easily evaluated also. Recall that

$$\mathbf{E}_i = \begin{bmatrix} \dfrac{\partial^2 \lambda_i}{\partial x_1 \, \partial x_1} & \dfrac{\partial^2 \lambda_i}{\partial x_1 \, \partial x_n} \\ \dfrac{\partial^2 \lambda_i}{\partial x_n \, \partial x_1} & \dfrac{\partial^2 \lambda_i}{\partial x_n \, \partial x_n} \end{bmatrix} \tag{1.166}$$

Using the λ_i we have

$$\mathbf{E}_i = \lambda_i \begin{bmatrix} g_{i1} g_{i1} & g_{i1} g_{in} \\ g_{in} g_{i1} & g_{in} g_{in} \end{bmatrix} \tag{1.167}$$

Thus, given the $\mathbf{h}$'s, $\mathbf{x}(t_0)$, and $\mathbf{P}(t_0)$, the filter is completely defined. We should also note that the covariance equation is independent of the measurement. In contrast to the Gaussian measurement case, this independence is met only for a nonlinear measurement, namely, an exponential.

These two examples represent dynamical estimation problems that have as their basis an actual physical phenomenon. Both have been used successfully in such an analysis (see McGarty [2, 3]). They also represent an analysis in which the continuous-time version was used for both the filter and the covariance.

The preceding analysis obtained linearized estimation equations and covariance equations by linearizing only after using either the Kushner-Stratonovich equation or the Snyder equation. In both cases, the measurements appeared in the covariance equation. In the examples and special cases, we saw that this dependency disappeared if the measurements were linear in the Gaussian case and exponential in the Poisson. Thus, if we linearize first in an appropriate manner, then some simplification may be obtained. Of course, the cost of this simplification is a more inaccurate filter.

We shall analyze Gaussian measurements. Now recall that the state equation is given by

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}, t) + \frac{d\mathbf{n}(t)}{dt} \tag{1.168}$$

Let us assume that $d\mathbf{n}/dt$ is a Gaussian white noise process with no Poisson part. Let

$$\mathbf{u}(t) = \frac{d\mathbf{n}(t)}{dt} \tag{1.169}$$

and formally

$$E[\mathbf{u}(t)\mathbf{u}^T(s)] = \mathbf{Q}(t)\delta(t - s) \tag{1.170}$$

where $\delta(\ )$ is the *Dirac delta function*. Let us now linearize $\mathbf{f}(\mathbf{x}, t)$ about some nominal trajectory $\mathbf{x}'(t)$ so that

$$\mathbf{f}(\mathbf{x}, t) = \mathbf{f}(\mathbf{x}', t) + \mathbf{A}(\mathbf{x}', t)(\mathbf{x} - \mathbf{x}') + \cdots \tag{1.171}$$

where $\mathbf{A}(\mathbf{x}', t)$ is defined in(1.4). Let us also assume that we can neglect the higher-order terms.

In like fashion the measurement equation can be formally written as

$$\mathbf{z}(t) = \mathbf{h}(\mathbf{x}, t) + \mathbf{v}(t) \tag{1.172}$$

where

$$\mathbf{z}(t) = \frac{d\mathbf{y}}{dt} \tag{1.173}$$

and

$$\mathbf{v}(t) = \frac{d\mathbf{w}(t)}{dt} \tag{1.174}$$

Thus, $\mathbf{v}(t)$ is an $m \times 1$ white noise process with covariance

$$E[\mathbf{v}(t)\mathbf{v}^T(s)] = \mathbf{R}(t)\delta(t - s) \tag{1.175}$$

Now expand $\mathbf{h}(\mathbf{x}, t)$about the same $\mathbf{x}'(t)$, retaining only linear terms:

$$\mathbf{h}(\mathbf{x}, t) = \mathbf{h}(\mathbf{x}', t) + \mathbf{C}(\mathbf{x}', t)(\mathbf{x} - \mathbf{x}') \tag{1.176}$$

where $\mathbf{C}(\mathbf{x}', t)$ is as in (1.9). Now define the two functions

$$\mathbf{r}(t) = \mathbf{f}(\mathbf{x}', t) - \mathbf{A}(\mathbf{x}', t)\mathbf{x}'(t) \tag{1.177}$$

and

$$\mathbf{s}(t) = \mathbf{h}(\mathbf{x}', t) - \mathbf{C}(\mathbf{x}', t)\mathbf{x}'(t) \tag{1.178}$$

Then the state equation becomes

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}(\mathbf{x}', t)\mathbf{x}(t) + \mathbf{r}(t) + \mathbf{u}(t) \tag{1.179}$$

and the measurement equation is

$$\mathbf{z}(t) = \mathbf{C}(\mathbf{x}', t)\mathbf{x}(t) + \mathbf{s}(t) + \mathbf{v}(t) \tag{1.180}$$

If the expansion $\mathbf{x}'(t)$ is deterministic, then $\mathbf{r}(t)$ and $\mathbf{s}(t)$ are known forcing functions; we shall assume they are. Now from case III for the Kalman-Bucy equations (linear system and linear measurements), we can obtain for the linearized estimate $\mathbf{x}^*(t)$

$$\frac{d\mathbf{x}^*(t)}{dt} = \mathbf{A}(\mathbf{x}', t)\mathbf{x}^*(t) + \mathbf{r}(t) + \mathbf{P}^*(t)\mathbf{C}^T(\mathbf{x}', t)\mathbf{R}^{-1}(t)$$
$$[\mathbf{z}(t) - \mathbf{s}(t) - \mathbf{C}(\mathbf{x}', t)\mathbf{x}^*(t)] \tag{1.181}$$

The linearized covariance equation is

$$\frac{d\mathbf{P}^*(t)}{dt} = \mathbf{A}(\mathbf{x}', t)\mathbf{P}^*(t) + \mathbf{P}^*(t)\mathbf{A}(\mathbf{x}', t) + \mathbf{Q}(t)$$
$$- \mathbf{P}^*(t)\mathbf{C}^T(\mathbf{x}', t)\mathbf{R}^{-1}(t)\mathbf{C}(\mathbf{x}', t)\mathbf{P}^*(t) \tag{1.182}$$

The covariance equation is now independent of the measurement, and if $\mathbf{x}'(t)$ is known a priori, then $\mathbf{P}^*(t)$ can be calculated a priori and thus the performance can be ascertained a priori. Clearly, in the linear case these equations are exact, since

$$\mathbf{f}(\mathbf{x}', t) = \mathbf{A}(t)\mathbf{x}'(t) \tag{1.183}$$

and

$$\mathbf{A}(\mathbf{x}', t)\mathbf{x}'(t) = \mathbf{A}(t)\mathbf{x}'(t) \tag{1.184}$$

so that $\mathbf{r}(t)$ is zero. Likewise, $\mathbf{s}(t)$ is also zero. For the nonlinear case, neither $\mathbf{r}(t)$ nor $\mathbf{s}(t)$ are zero and they act as driving functions to correct for the nonlinearity.

There are many ways to choose $\mathbf{x}'(t)$. The first way is to use the Fokker-Planck equation for this system and from it obtain the a priori estimate or the mean. This may not be a trivial thing to do, since it involves solving a nonlinear differential integral equation. A second method is to let

$$\mathbf{x}^*(t) = \mathbf{x}'(t) \tag{1.185}$$

The resulting equations are similar to those under the initial linearization discussed. This form or realization is called the *extended Kalman filter*.

Example. A simple first-order linear time-invariant system is given by the equation

$$\dot{x} = ax \tag{1.186}$$

with suitable initial conditions. In order that this be a stable system, $a < 0$.

Often the exact value of a is unknown, so some method must be devised to obtain it. Let us assume that the system is also driven by white Gaussian noise with covariance Q. Then we have

$$\dot{x}(t) = ax(t) + n(t) \tag{1.187}$$

Furthermore, we are able to measure $x(t)$ as

$$z(t) = x(t) + w(t) \tag{1.188}$$

where $w(t)$ is white Gaussian noise of variance R. From the measurement $z(t)$, we want to estimate a. First change this into state variable form. Let $x(t)$ equal $x_1(t)$ and a equal $x_2(t)$. Then the state equation becomes

$$\dot{x}_1(t) = -x_2(t)x_1(t) + n(t) \tag{1.189}$$

$$\dot{x}_2(t) = 0 \tag{1.190}$$

with the measurement

$$z(t) = x_1(t) + w(t) \tag{1.191}$$

With this formulation we can try to estimate $x_2(t)$ using a nonlinear filter. The estimator equations are given by

$$\dot{\hat{x}}_1(t) = -\hat{x}_2(t)\hat{x}_1(t) - P_{12}(t) + P_{11}(t)R^{-1}(z(t) - \hat{x}_1(t)) \tag{1.192}$$

$$\dot{\hat{x}}_2(t) = P_{12}(t)R^{-1}(z(t) - \hat{x}_1(t)) \tag{1.193}$$

where the covariance components are given by

$$\dot{P}_{11}(t) = -2\hat{x}_2(t)P_{11}(t) - 2\hat{x}_1(t)P_{12}(t) - R^{-1}P_{11}^2(t) + Q \tag{1.194}$$

$$\dot{P}_{12}(t) = -\hat{x}_2(t)P_{12}(t) - \hat{x}_1(t)P_{22}(t) - R^{-1}P_{11}(t)P_{12}(t) \tag{1.195}$$

$$\dot{P}_{22}(t) = -R^{-1}P_{12}^2(t) \tag{1.196}$$

with the initial conditional given.

In Figures 6.4–6.9 we present the results of this filter applied to a specific example. In this case, we assumed that

$$x_1(0) = 0.1 \tag{1.197}$$

$$x_2(0) = 0.08 \tag{1.198}$$

and

$$P_{11}(0) = 0.0; \qquad P_{12}(0) = 0.0; \qquad P_{22}(0) = 0.5 \tag{1.199}$$

with R equal to 0.002 and Q equal to 0.2 .The actual value of the time constant was 0.1. The structure of the filter can be observed by this example.

Figure 6.4 $\hat{x}_1$ and $\hat{x}_2$ versus time for Identification Problem.

Figure 6.5 x_1 and $\hat{x}_2$ versus time for Identification Problem.

The estimate of $x_1(t)$ was quite reasonable with only small errors occurring. However, $\hat{x}_2(t)$ initially converged on the actual value, but after some time, it began to diverge away. This can be attributed to the structure for $\hat{x}_2(t)$. Note that $\dot{\hat{x}}_2(t)$ depends upon the difference between the measurement and the estimate of the state. Since $P_{12}(t)$ is negative, it implies that if the state estimate is less than the measurement, it is decaying faster than the estimate. This would mean that we should decrease our estimate of $x_2(t)$. Likewise, if $x_1(t)$ is greater than $z(t)$, we are lagging behind and $\hat{x}_2(t)$ should be increased. That is, initially what occurs as long as $P_{12}(t)$ is large. However, by observing $P_{12}(t)$ in Figure 6.8, we see that it decays to zero and oscillates about that point in a random fashion. Observation of the equation for $\dot{P}_{12}(t)$ shows why this occurs. $P_{12}(t)$ is given by a decaying exponential type of equation where the time constant is $\hat{x}_2(t)$. Thus, it will tend to decay to a small number. The result of this is that $P_{12}(t)$ as $t \to \infty$ will be almost constant. Then $\hat{x}_2(t)$ is given by the equation

Figure 6.6 Measurement versus time for Identification Problem.

Figure 6.7 Covariance p_{11} versus time.

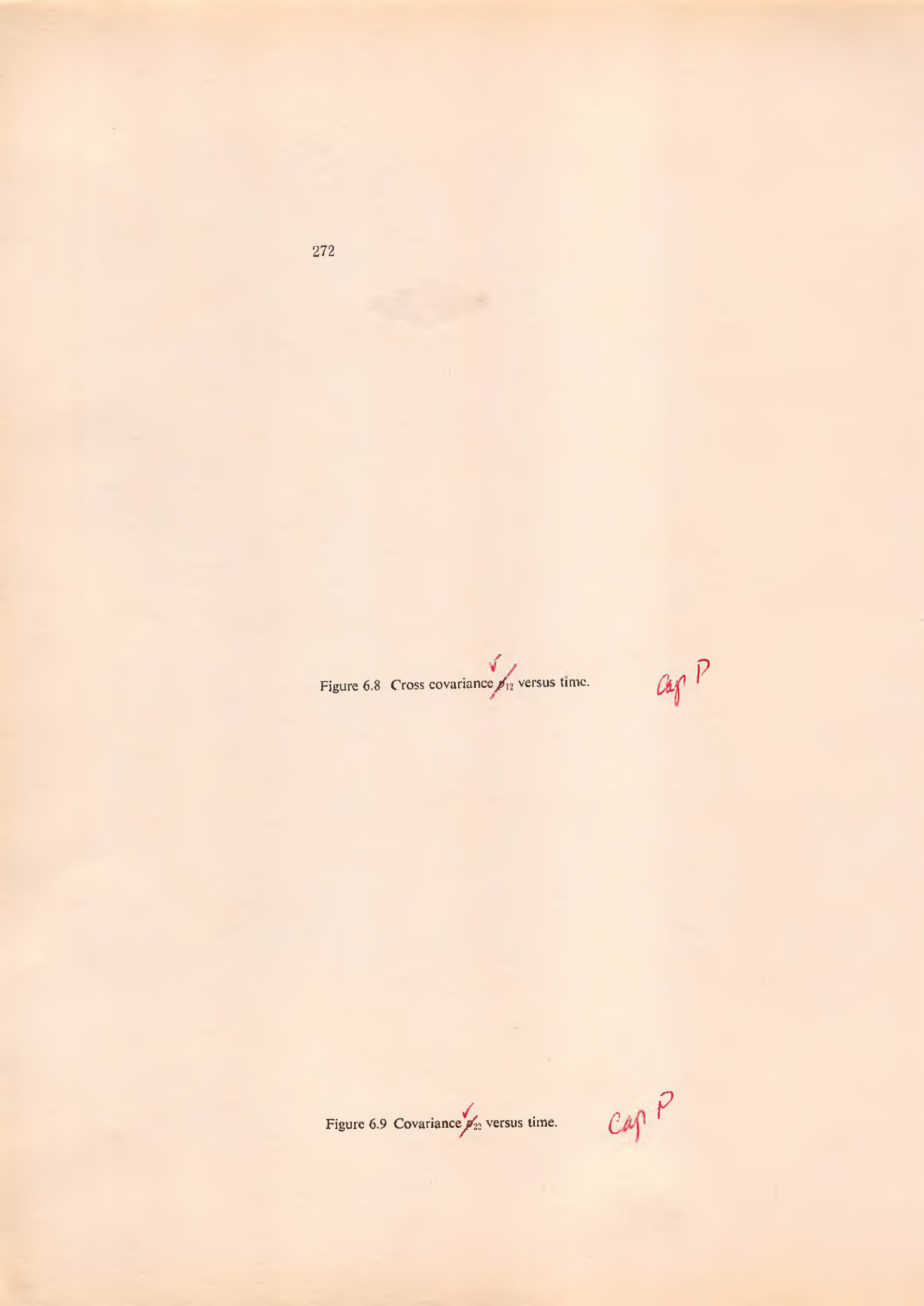

Figure 6.8 Cross covariance p_{12} versus time.

Figure 6.9 Covariance p_{22} versus time.

$$\dot{\hat{x}}_2(t) \approx P_{12}(\infty) R^{-1}(z(t) - \hat{x}_1(t)) \tag{1.200}$$

Now $z(t) - \hat{x}_1(t)$ is given by

$$z(t) - \hat{x}_1(t) = \tilde{x}_1(t) + w(t) \tag{1.201}$$

That is, it has a white noise component. so that $\hat{x}_2(t)$ becomes a Wiener process and thus useless.

This qualitative discussion represents one case where the direct application of a nonlinear filter may very well lead to serious problems; specifically, the filter diverges under certain conditions (see Section 6.4). What is even more interesting about this example is that if we had taken the spectrum of $z(t)$ and looked for the 3*db* points of that part due to the $x(t)$ process, the estimate of a would have been obtained quite directly!

Example. In the previous example, we discussed parameter or system identification via the nonlinear system

$$\dot{x}_2(t) = 0 \tag{1.202}$$

$$\dot{x}_1(t) = -x_2(t)x_1(t) + n(t) \tag{1.203}$$

with the linear measurement

$$z(t) = x_1(t) + w(t) \tag{1.204}$$

Now, if $n(t)$ were forced to 0 and $x_1(0)$ were known then we could write $x_1(t)$ directly in terms of $x_2(t)$ as

$$x_1(t) = x_1(0) \exp[-x_2(t)\, t] \tag{1.205}$$

Thus, the measurement can be written

$$z(t) = x_1(0) \exp[-x_2(t)t] + w(t) \tag{1.206}$$

This allows us to reduce the problem to a linear system

$$\dot{x}_2(t) = 0 \tag{1.207}$$

with a nonlinear measurement. But this problem has already been discussed in the example on page ~~000~~.

Example. A *phase-lock loop* is a device used to determine the phase of a signal. The signal is a sinusoid that is additively disturbed by noise. The form of the received signal is

$$z(t) = \cos[2\pi f_0 t + \phi(t)] + w(t) \tag{1.208}$$

where $\phi(t)$, the phase, is a random process. The frequency f_0 is the carrier frequency and $w(t)$ is white Gaussian noise with covariance

$$E[w(t)w(t+\tau)] = R\delta(\tau) \tag{1.209}$$

A simple example for $\phi(t)$ would be a first -orderMarkov process generated by $x(t)$, where

Figure 6.10 Sample state function for phase of phase lock loop.

$$\dot{x}(t) = -\alpha x(t) + n(t) \tag{1.210}$$

where $n(t)$ is a white Gaussian process with

$$E[n(t)n(t+\tau)] = Q\delta(\tau) \tag{1.211}$$

Then we let $\dot{\phi}(t)$ equal $\beta x(t)$, where β is the modulation index.

Now the modulation constant can be absorbed into the process $x(t)$, so that it is equivalent to let $x(t)$ represent $\dot{\phi}(t)$ directly. Thus, the estimation problem is to obtain an estimate of $x(t)$, given $z(t)$, where:

$$z(t) = \cos[2\pi f_0 t + x(t)] + w(t) \tag{1.212}$$

This is a linear-system–nonlinear-measurement problem. A sample function for this process $x(t)$ is shown in Figure 6.10 for $x(0)$ equal to zero and $Q = 0.01$. The corresponding output is in Figure 6.11 for $f_0 = 1$ and $R = 0.001$.

In Figures 6.12–6.19 we compare the results of using four different approximate nonlinear estimators. The first is the extended Kalman filter that has been linearized about the average trajectory $x(t) = 0$. This estimate is very poor. The second estimator is also an extended Kalman filter, but now the expansion point is $x^*(t)$. There is a marked increase in performance. Note, however, that the variance is almost identical but not exact.

Figure 6.11 Sample measurement of a phase lock loop.

The third filter was a first-order nonlinear filter of the form

$$\dot{x}^*(t) = -\alpha x(t) - \sin(2\pi f_0 t + x^*(t))P(t)R^{-1} [z(t) - \cos(2\pi f_0 t + x^*(t))(1 - 0.5P(t))] \quad (1.213)$$

$$\dot{P}(t) = -2\alpha P(t) + Q + P^2(t)\sin^2(2\pi f_0 t + x^*(t)) - 0.5\cos^2(2\pi f_0 t + x^*(t))P^3(t) \quad (1.214)$$

Its performance is identical to the extended Kalman filter along $x^*(t)$. The fourth filter is the second-order filter with the measurements appearing in the covariance equation. Note that in Figure 6.19, which is the covariance obtained from this approach there do appear perturbations compared to the third filter's covariance.

The extended Kalman filter of case II leads to an interesting interpretation. The estimate is given by

$$\dot{x}^*(t) = -\alpha x^*(t) - \sin[2\pi f_0 t + x^*(t)]P(t)R^{-1} \{z(t) - \cos[2\pi f_0 t + x^*(t)]\} \quad (1.215)$$

Now this can be viewed as a system that demodulates the difference between the output and the expected output and then filters it through the system that generates $x^*(t)$. If this system is a low-frequency system compared to f_0, the

Figure 6.12 Estimate of state using extended K-B filter about a nominal trajectory.

Figure 6.13 Covariance of estimate using linearized K-B filter about nominal trajectory.

Figure 6.14 State estimate for first-order K-S filter.

Figure 6.15 Covariance for first-order K-S filter.

Figure 6.16 State estimate for modified second-order K-S filter.

Figure 6.17 Covariance for modified second-order K-S filter.

Figure 6.18 State estimate for extended second-order K-S filter.

Figure 6.19 Covariance for extended second-order K-S filter.

term $\sin [2\pi f_0 t + x^*(t)] \cos [2\pi f_0 t + x^*(t)]$ will be filtered out. Furthermore, if

$$z(t) = \cos [2\pi f_0 t + x^*(t)] + w(t) \tag{1.216}$$

and if we assume $x^*(t)$ is close to the actual value, then we have

$$\sin (2\pi f_0 t + x^*(t)) z(t) \approx \hat{x}^*(t) - x(t) + w'(t) + \begin{bmatrix} \text{higher} \\ \text{frequencies} \end{bmatrix} \tag{1.217}$$

Thus, with this assumption, $\dot{x}(t)$ is given by

$$\dot{\hat{x}}(t) = -\alpha \hat{x}(t) + (x(t) - \hat{x}(t)) + w'(t) \tag{1.218}$$

where $w'(t)$ is the modulated white noise that is still white noise. Thus, the system is driven by the error. This feedback nature is why this is called the phase-lock loop. In Figure 6.20 we have schematically sketched the complete loop structure.

A comparison of the performance of these filters has been done by Schwartz and Stear for two different systems with distinctively different nonlinearities and in Mehra [2] for orbital dynamics. The results indicate that linearization about a priori nominal $\mathbf{x}'(t)$ in the preceding analysis may lead to unacceptable filter performance. Yet the other methods—Snyder's linearization; Bass, Norum, and Schwartz's linearization; and the extended Kalman filter—lead to similar results. This in general seems to be the case. Only in

Figure 6.20 Optimum demodulator for phase-modulated signal—phase-locked loop.

the case of some severe nonlinearity does one method excel the others, but in those cases, other, more complex methods may be necessary.

This section completes our discussion of the continuous linearized estimation equations. The following sections will discuss the discrete versions but will rely upon the analytical insight developed in this section.

6.2 OPTIMALLY DRIVEN FILTERING

This section will present a technique that uses a great deal of a priori knowledge of "good" filter structure to develop a method for filtering of nonlinear systems. The method is that developed by Athans, Wishner, and Bertolini. We shall present the results applicable for a linear measurement system and refer the reader to the problems for the treatment of nonlinear measurement. The extension to a nonlinear measurement is almost trivial once we have presented the results for the nonlinear system.

The underlying motivation for this presentation is to show that if we know how the filter "should" look, and realize that a slight forcing function improves performance, as we said in the extended Kalman filter technique, then a simple method of filtering can be proposed. The choice of the forcing function must be made with care. The improvements obtained by the use of this technique may often warrant its inclusion despite the increased computational complexity. Also, it gives a more general case of the extended Kalman filter and at the same time has some additional unique properties that are computationally more useful.

The system to be analyzed will be a continuous-time state equation, but the measurements will only be made at discrete instants. We then seek estimates of $\mathbf{x}(t)$ at discrete times $\{kT\}$ associated with the arrival of a measurement $\mathbf{z}(k)$. As we observed in the last section by linearizing about the estimates both the state and measurement equations had extra driving functions, which were the residuals of the linearization. What we shall do in this analysis is to let this residual be arbitrary and then, by using an appropriate cost criterion, to obtain the form they should take to minimize this criterion. Thus, the state equation will be driven by some function $\phi(t)$ between measurements $\mathbf{z}(k)$, $\mathbf{z}(k+1)$ so that the estimate $\hat{\mathbf{x}}(k+1)$ can be optimized.

Let us begin by defining the system equation and the measurement. Let

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t)) + \frac{d\mathbf{n}(t)}{dt} \tag{2.1}$$

be the system equation. Note that we assume the nonlinearity to be time-independent. The continuous measurement equation is

$$\mathbf{z}(t) = \mathbf{C}(t)\mathbf{x}(t) + \frac{d\mathbf{w}(t)}{dt} \tag{2.2}$$

where $\mathbf{C}(t)$ is an $m \times n$ matrix relating state to measurement. Again, both $\mathbf{w}(t)$ and $\mathbf{n}(t)$ are Wiener processes. Their covariances are as previously stated: namely,

$$E[\mathbf{n}(t)\mathbf{n}^T(s)] = \mathbf{Q}(t)\min(t, s) \tag{2.3}$$

and

$$E[\mathbf{w}(t)\mathbf{w}^T(s)] = \mathbf{R}(t)\min(t, s) \tag{2.4}$$

where $\mathbf{Q}(t)$ is an $n \times n$ matrix. Also, for discrete time, we shall denote $\mathbf{x}(k)$ to be $\mathbf{x}(kT)$, where T is some selected sample time.

In this method we shall again need to make use of a Taylor-series expansion of the nonlinearity about some arbitrary point $\bar{\mathbf{x}}$:

$$\mathbf{f}(\mathbf{x}) \cong \mathbf{f}(\bar{\mathbf{x}}) + \mathbf{A}(\bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}}) + \frac{1}{2}\sum_{i=1}^{n} \gamma_i(\mathbf{x} - \bar{\mathbf{x}})^T \mathbf{B}_i(\bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}}) \tag{2.5}$$

The matrix $\mathbf{A}(\bar{\mathbf{x}})$ is

$$\mathbf{A}(\bar{\mathbf{x}}) = \left[\begin{array}{ccc} \dfrac{\partial f_1}{\partial x_1} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ & & \\ \dfrac{\partial f_n}{\partial x_1} & \cdots & \dfrac{\partial f_n}{\partial x_n} \end{array}\right]_{\mathbf{x}=\bar{\mathbf{x}}} \tag{2.6}$$

and

$$\mathbf{B}_i(\bar{\mathbf{x}}) = \left[\begin{array}{ccc} \dfrac{\partial f_i}{\partial x_1\ \partial x_1} & \cdots & \dfrac{\partial f_i}{\partial x_1\ \partial x_n} \\ & & \\ \dfrac{\partial f_i}{\partial x_n\ \partial x_1} & \cdots & \dfrac{\partial f_i}{\partial x_n\ \partial x_n} \end{array}\right]_{\mathbf{x}=\bar{\mathbf{x}}} \tag{2.7}$$

and γ_i is as defined before. Both $\mathbf{A}(\bar{\mathbf{x}})$ and $\mathbf{B}_i(\bar{\mathbf{x}})$ are $n \times n$ matrices and the γ_i are $n \times 1$ vectors.

As before $\hat{\mathbf{x}}(k)$ will represent the estimate of $\mathbf{x}(k)$ and $\tilde{\mathbf{x}}(k)$ will be the error defined as

$$\tilde{\mathbf{x}}(k) = \mathbf{x}(k) - \hat{\mathbf{x}}(k) \tag{2.8}$$

We shall develop a filter by means of the following reasoning. If the system is unperturbed, then the state will be some value $\mathbf{x}^*(t)$. Furthermore, the system will follow the trajectory

$$\dot{\mathbf{x}}^*(t) = \mathbf{f}(\mathbf{x}^*(t)) \tag{2.9}$$

Unfortunately, the system is perturbed by some noise, and indeed, the trajectory differs from that obtained by solving (2.9). But if we were to apply some forcing function to (2.9), say $\phi(t)$, then possibly by choosing $\phi(t)$ in some "nice" manner, the solution to this driven equation would give us a better estimate. Let us then add this perturbation to (2.9):

Figure 6.21 System trajectories.

$$\dot{\mathbf{x}}^*(t) = \mathbf{f}(\mathbf{x}^*(t)) + \phi(t) \tag{2.10}$$

The three different trajectories of the state from t_0 to t_1 are shown in Figure 6.21, where a is the undriven trajectory described by (2.9), b is that driven by $\phi(t)$ (equation 2.10), and c is the true trajectory described by (2.1). We are also interested in the discrete-time propagation of the system. Define $\tilde{\mathbf{x}}^*(k)$ as the error at time kT between the real state and that of our estimate given by

$$\tilde{\mathbf{x}}^*(k) = \mathbf{x}(k) - \mathbf{x}^*(k) \tag{2.11}$$

Assume that at time $k-1$ we have some estimate $\hat{\mathbf{x}}(k-1|k-1)$. Again, the notation implies that the estimate is at time $k-1$, given data to time $k-1$. Now let (2.10) take the system from $\mathbf{x}^*(k-1) = \hat{\mathbf{x}}(k-1|k-1)$ to the state at time k. Then

$$\dot{\mathbf{x}}^*(t) = \mathbf{f}(\mathbf{x}^*(t)) + \phi(t) \tag{2.12}$$

where the initial condition is given by

$$\mathbf{x}^*(k-1) = \hat{\mathbf{x}}(k-1 \mid k-1) \tag{2.13}$$

and the equation is defined on the interval $t \in [(k-1)T, kT)$. At time k the

Figure 6.22 Predicted and filtered estimates.

state given by the solution of (2.12) is $\mathbf{x}^*(k)$; it will be called the *predicted estimate* of $x(t)$ at time kT. Thus, in our notation,

$$\mathbf{x}^*(k) \triangleq \hat{\mathbf{x}}(k \mid k-1) \tag{2.14}$$

which is the solution of (2.12).

Such a trajectory is shown in Figure 6.22. Now $\hat{\mathbf{x}}(k|k)$ is the estimate of $\mathbf{x}$ at time kT, given the data at time kT. Such an estimate is also shown on Figure 6.22. This is then what we seek to obtain, an estimate of $\mathbf{x}(k)$, given the new $\mathbf{z}(k)$, knowing $\hat{\mathbf{x}}(k-1|k-1)$. This procedure follows that for the usual derivation of the Kalman filter.

At time $(k-1)$, $\tilde{\mathbf{x}}^*(k-1)$ of our system is

$$\tilde{\mathbf{x}}^*(k-1) = \mathbf{x}(k-1) - \mathbf{x}^*(k-1) \tag{2.15}$$

But, by definition [see (2.13)] $\mathbf{x}^*(k-1)$ was $\hat{\mathbf{x}}(k-1 \mid k-1)$. Then (2.15) becomes

$$\tilde{\mathbf{x}}^*(k-1) = \mathbf{x}(k-1) - \hat{\mathbf{x}}(k-1|k-1) = \tilde{\mathbf{x}}(k-1) \tag{2.16}$$

which says that the approximate error is the true error at this time. Now for any time during the interval $[k-1, k)$, we have

$$\dot{\tilde{\mathbf{x}}}^*(t) = \dot{\mathbf{x}}(t) - \dot{\mathbf{x}}^*(t) = \mathbf{f}(\mathbf{x}(t)) - \mathbf{f}(\mathbf{x}^*(t)) - \boldsymbol{\phi}(t) + \boldsymbol{\eta}(t) \tag{2.17}$$

Expand the system about the predicted state, namely, $\mathbf{x}^*(t)$. This gives for (2.17) (to a second-order approximation)

$$\begin{aligned}\dot{\tilde{\mathbf{x}}}^*(t) &= \mathbf{f}(\mathbf{x}^*(t)) + \mathbf{A}(\mathbf{x}^*)(\mathbf{x}(t) - \mathbf{x}^*(t)) \\ &\quad + \sum_{i=1}^{n} \gamma_i(\mathbf{x}(t) - \mathbf{x}^*(t))^T\mathbf{B}_i(\mathbf{x}^*)(\mathbf{x}(t) - \mathbf{x}^*(t)) \\ &\quad - \mathbf{f}(\mathbf{x}^*(t)) - \boldsymbol{\phi}(t) + \mathbf{n}(t)\end{aligned} \tag{2.18}$$

Canceling and using the definition of $\mathbf{x} - \mathbf{x}^*$, we obtain

$$\dot{\tilde{\mathbf{x}}}^*(t) = \mathbf{A}(\mathbf{x}^*)\tilde{\mathbf{x}}^*(t) + \sum_{i=1}^{n} \gamma_i\tilde{\mathbf{x}}^{*T}\mathbf{F}_i(\tilde{\mathbf{x}}^*)\tilde{\mathbf{x}}^* - \boldsymbol{\phi}(t) + \mathbf{n}(t) \tag{2.19}$$

Now we demand that both $\hat{\mathbf{x}}(k-1 \mid k-1)$ and $\mathbf{x}^*(t)$; $t \in [k-1, k)$ be unbiased. This then yields

$$E[\tilde{\mathbf{x}}^*(t)] = \mathbf{0} \tag{2.20}$$

and

$$E[\dot{\tilde{\mathbf{x}}}^*(t)] = \mathbf{0} \tag{2.21}$$

Thus, taking the expectation of both sides of (2.19) will give us a means for computing $\boldsymbol{\phi}(t)$.

Before doing so, we shall take note of one fact that will help us in simplifying the result. From the previous section, we found that

$$\sum_{i=1}^{n} \gamma_i\tilde{\mathbf{x}}^{*T}\mathbf{B}_i(\mathbf{x}^*)\tilde{\mathbf{x}}^* = \sum_{i=1}^{n} \gamma_i \operatorname{tr}(\mathbf{B}_i(\mathbf{x}^*)\tilde{\mathbf{x}}^*\tilde{\mathbf{x}}^{*T}) \tag{2.22}$$

where tr means trace. Now using this notation in (2.14) and taking the expectations, we obtain

$$\mathbf{0} = \frac{1}{2}\sum_{i=1}^{n} \gamma_i \operatorname{tr}[\mathbf{B}_i(\mathbf{x}^*)E[\tilde{\mathbf{x}}^*\tilde{\mathbf{x}}^{*T}]] - \boldsymbol{\phi}(t) \tag{2.23}$$

Define the expectation as

$$E[\tilde{\mathbf{x}}^*\tilde{\mathbf{x}}^{*T}] \triangleq \mathbf{M}(t) \tag{2.24}$$

Then $\boldsymbol{\phi}(t)$ becomes

$$\boldsymbol{\phi}(t) = \frac{1}{2}\sum_{i=1}^{n} \gamma_i \operatorname{tr}[\mathbf{B}_i(\mathbf{x}^*)\mathbf{M}(t)] \tag{2.25}$$

The matrix $\mathbf{M}(t)$ $(n \times n)$ is a covariance matrix on the predicted estimate over the prescribed interval. It is now possible to obtain a differential equation that will generate a solution to $\mathbf{M}(t)$. Now

$$\dot{\mathbf{M}}(t) = E[\dot{\tilde{\mathbf{x}}}^*(t)\tilde{\mathbf{x}}^{*T}(t) + \tilde{\mathbf{x}}^*(t)\dot{\tilde{\mathbf{x}}}^{*T}(t)] \tag{2.26}$$

Using (2.19) and (2.25), we obtain

$$\begin{aligned}\dot{\mathbf{M}}(t) = {} & E[\mathbf{A}(\mathbf{x}^*)\tilde{\mathbf{x}}^*(t)\tilde{\mathbf{x}}^{*T}(t) + \frac{1}{2}(\sum_{i=1}^{n}\gamma_i \operatorname{tr}[\mathbf{B}_i(\tilde{\mathbf{x}}^*)[\tilde{\mathbf{x}}^*(t)\tilde{\mathbf{x}})^{*T}(t) - \mathbf{M}(t)]])\tilde{\mathbf{x}}^{*T}(t) \\ & + \mathbf{n}(t)\mathbf{x}^{*T}(t) \\ & + \tilde{\mathbf{x}}^*(t)\tilde{\mathbf{x}}^{*T}(t)\mathbf{A}^T(\mathbf{x}^*) + \tilde{\mathbf{x}}^*(t)\frac{1}{2}(\sum_{i=1}^{n}\gamma_i \operatorname{tr}[\mathbf{B}_i(\mathbf{x}^*)[\tilde{\mathbf{x}}^*(t)\tilde{\mathbf{x}}^{*T}(t) - \mathbf{M}(t)]]) \\ & + \mathbf{x}^*(t)\mathbf{n}^T(t)]\end{aligned} \tag{2.27}$$

If we assume that $\tilde{\mathbf{x}}^*(t)$ is Gaussian with zero mean, all the odd moments are zero. This reduces (2.27) to

$$\dot{\mathbf{M}}(t) = \mathbf{A}(\mathbf{x}^*)\mathbf{M}(t) + \mathbf{M}(t)\mathbf{A}^T(\mathbf{x}^*) + \mathbf{Q}(t) \tag{2.28}$$

with the initial condition

$$\mathbf{M}(k-1) = \mathrm{E}[\tilde{\mathbf{x}}^*(k-1)\tilde{\mathbf{x}}^{*T}(k-1)] = E[\tilde{\mathbf{x}}(k-1)\tilde{\mathbf{x}}^T(k-1)] \triangleq \mathbf{P}(k-1) \tag{2.29}$$

where $\mathbf{P}(k-1)$ is the covariance matrix of the filtered estimate at time $(k-1)$ given data to time $(k-1)$ and $\mathbf{M}(k-1)$ is the covariance matrix at time $(k-1)$. Thus, these two are equated at $(k-1)$ because of the updating of the system. We can obtain a solution to (2.28) if we let

$$\mathbf{M}(t) = \begin{bmatrix} m_{11}(t) & m_{1n}(t) \\ m_{n1}(t) & m_{nn}(t) \end{bmatrix} \tag{2.30}$$

$$\mathbf{A}(\mathbf{x}^*) = \begin{bmatrix} a_{11} & a_{1n} \\ a_{n1} & a_{nn} \end{bmatrix} \tag{2.31}$$

and assume the system to be noiseless; that is, $\mathbf{Q}(t) = \mathbf{0}$.
By defining

$$\mathbf{m}(t) = \begin{bmatrix} m_{11}(t) \\ \vdots \\ m_{1n}(t) \\ \vdots \\ m_{21}(t) \\ \vdots \\ m_{nn}(t) \end{bmatrix} \tag{2.32}$$

Then (2.28) becomes

$$\dot{\mathbf{m}}(t) = \mathbf{B}\mathbf{m}(t); \qquad \mathbf{m}(k-1) = \mathbf{p}(k-1) \tag{2.33}$$

and the $n^2 \times n^2$ matrix $\mathbf{B}$ is

$$
\mathbf{B} = \begin{array}{c} \\ n \\ n \\ \\ + \end{array}
\begin{array}{c}
\overbrace{}^{n} \\
\left[\begin{array}{cccc|ccc|c}
a_{11} & 0 & \cdots & 0 & a_{12} & 0 \cdots 0 & a_{1n} \; 0 \cdots \\
0 & a_{11} & \cdots & 0 & 0 & a_{12} \cdots 0 & \\
0 & 0 & & a_{11} & & & \\
\hline
a_{21} & 0 & \cdots & & & & \\
\hline
0 & 0 & & a_{n1} & & & 0 \quad a_{n1}
\end{array}\right] \\
\left[\begin{array}{cccc|ccc|c}
a_{11} & a_{12} & a_{13} & a_{1n} & & & \\
a_{21} & \cdots & \cdots & a_{2n} & & & \\
a_{n1} & \cdots & \cdots & a_{nn} & & 0 & \\
\hline
 & & & & & & \\
\hline
 & & & & & & \mathbf{A}
\end{array}\right]
\end{array}
\begin{array}{l} \\ \\ \\ \\ n^2 \end{array}
\tag{2.34}
$$

Thus, $\mathbf{m}(t)$ is

$$\mathbf{m}(t) = e^{\mathbf{B}(t-(k-1)T)}\mathbf{p}(k-1) \tag{2.35}$$

Now we shall *define* the filtered estimate to be

$$\hat{\mathbf{x}}(k|k) = \hat{\mathbf{x}}(k|k-1) + \mathbf{K}(k)[\mathbf{z}(k) - \mathbf{C}(k)\hat{\mathbf{x}}(k|k-1)] \tag{2.36}$$

This is identical to the discrete Kalman filter equation developed in Chapter 4 and thus its selection. We will now produce a method to obtain the gain matrix $\mathbf{K}(k)$. Let us obtain first the error of this estimate:

$$\tilde{\mathbf{x}}(k|k) = \mathbf{x}(k) - \hat{\mathbf{x}}(k|k) \tag{2.37}$$

Using (2.36) and the system equation, we obtain

$$\tilde{\mathbf{x}}(k|k) = \tilde{\mathbf{x}}^*(k) - \mathbf{K}(k)[\mathbf{z}(k) - \mathbf{C}(k)\hat{\mathbf{x}}(k|k-1)] \tag{2.38}$$

But $\mathbf{z}(k)$ is also given by

$$\mathbf{z}(k) = \mathbf{C}(k)\mathbf{x}(k) + \mathbf{v}(k) \tag{2.39}$$

Using this in (2.38) yields

$$\tilde{\mathbf{x}}(k|k) = \tilde{\mathbf{x}}^*(k) - \mathbf{K}(k)[\mathbf{C}(k)\tilde{\mathbf{x}}^*(k) + \mathbf{v}(k)] \tag{2.40}$$

Now we want to obtain the covariance on this estimate

$$E[\tilde{\mathbf{x}}(k|k)\tilde{\mathbf{x}}^T(k|k)] = \mathbf{P}(k) \tag{2.41}$$

Using (2.40) in (2.41), we obtain

$$\mathbf{P}(k) = [\mathbf{I} - \mathbf{K}(k)\mathbf{C}(k)]\mathbf{M}(k)[\mathbf{I} - \mathbf{K}(k)\mathbf{C}(k)]^T + \mathbf{K}(k)\mathbf{R}(k)\mathbf{K}^T(k) \tag{2.42}$$

We now discuss the choice of the gain matrix $\mathbf{K}(k)$. In order to do so, we must establish some cost criterion. To do so, let $\mathbf{S}(k)$ be an arbitrary symmetric positive definite matrix and consider the cost function.

$$J(k) = E[\tilde{\mathbf{x}}^T(k|k)\mathbf{S}(k)\tilde{\mathbf{x}}(k|k)] \tag{2.43}$$

Using the $\hat{\mathbf{x}}(k|k)$ as in (2.36), we desire to find matrix $\mathbf{K}(k)$ that will minimize this cost function. The $J(k)$ can be written as

$$J(k) = \text{tr}[\mathbf{S}(k)\mathbf{P}(k)] \tag{2.44}$$

where $\mathbf{P}(k)$ is the function of $\mathbf{K}(k)$. Thus, finding the stationary point of (2.44) with respect to $\mathbf{K}(k)$ can be obtained by differentiating and setting the result to 0; that is,

$$\frac{\partial J(k)}{\partial \mathbf{K}(k)} = 0 \qquad \text{(formally only)} \tag{2.45}$$

Performing the indicated differentiation yields

$$\begin{aligned} &-\mathbf{S}^T(k)\mathbf{M}^T(k)\mathbf{C}^T(k) - \mathbf{S}(k)\mathbf{M}(k)\mathbf{C}^T(k) + \mathbf{S}^T(k)\mathbf{K}(k)[\mathbf{C}(k)\mathbf{M}(k)\mathbf{C}^T(k) \\ &\qquad + \mathbf{R}(k)]^T + \mathbf{S}(k)\mathbf{K}(k)[\mathbf{C}(k)\mathbf{M}(k)\mathbf{C}^T(k) + \mathbf{R}(k)] = \mathbf{0} \end{aligned} \tag{2.46}$$

Since $\mathbf{S}^{-1}(k)$ exists by hypothesis, one can solve for $\mathbf{K}(k)$;

$$\mathbf{K}(k) = \mathbf{M}(k)\mathbf{C}^T(k)[\mathbf{C}(k)\mathbf{M}(k)\mathbf{C}^T(k) + \mathbf{R}(k)]^{-1} \tag{2.47}$$

Now substitute (2.47) into (2.42) to obtain

$$\begin{aligned} \mathbf{P}(k) = \mathbf{M}(k) &- \mathbf{M}(k)\mathbf{C}^T(k)[\mathbf{C}(k)\mathbf{M}(k)\mathbf{C}^T(k) + \mathbf{R}(k)]^{-1}\mathbf{C}(k)\mathbf{M}(k) \\ &- \mathbf{M}(k)\mathbf{C}^T(k)[\mathbf{C}(k)\mathbf{M}(k)\mathbf{C}^T(k) + \mathbf{R}(k)]^{-1T}\mathbf{C}(k)\mathbf{M}^T(k) \\ &+ \mathbf{M}(k)\mathbf{C}^T(k)[\mathbf{C}(k)\mathbf{M}(k)\mathbf{C}^T(k) + \mathbf{R}(k)]^{-1}\mathbf{M}(k) \\ &\quad [[\mathbf{C}(k)\mathbf{M}(k)\mathbf{C}^T(k) + \mathbf{R}(k)]^{-1}]^T\mathbf{C}(k)\mathbf{M}^T(k) \\ &+ \mathbf{M}(k)\mathbf{C}^T(k)[\mathbf{C}(k)\mathbf{M}(k)\mathbf{C}^T(k) + \mathbf{R}(k)]^{-1}\mathbf{R}(k) \\ &\quad [[\mathbf{C}(k)\mathbf{M}(k)\mathbf{C}^T(k) + \mathbf{R}(k)]^{-1}]^T\mathbf{C}(k)\mathbf{M}^T(k) \end{aligned} \tag{2.48}$$

Using several matrix identities for inverses and after much algebra,* one obtains for $\mathbf{P}(k)$

$$\mathbf{P}(k) = \mathbf{M}(k) - \mathbf{M}(k)\mathbf{C}^T(k)[\mathbf{C}(k)\mathbf{M}(k)\mathbf{C}^T(k) + \mathbf{R}(k)]^{-1}\mathbf{C}(k)\mathbf{M}(k) \tag{2.49}$$

This completes the filter derivation. One now seeks a path of implementation. We initially have $\hat{\mathbf{x}}(0)$ and $\mathbf{P}(0)$, and the estimation proceeds as follows:

1. Given $\hat{\mathbf{x}}(0)$, $\mathbf{P}(0)$.
2. Compute $\mathbf{M}(1)$ from $\mathbf{P}(0)$ using (2.35).
3. Compute $\phi(t)$ using $\mathbf{M}(t)$ and (2.25) with the expansion about $\hat{\mathbf{x}}(0)$.
4. Compute $\mathbf{x}^*(t)$ and $\hat{\mathbf{x}}(1|0)$ using $\phi(t)$in (2.12).
5. Compute $\mathbf{K}(1)$ from (2.47).
6. Receive $\mathbf{z}(1)$ from instruments.
7. Use (2.36) to obtain $\hat{\mathbf{x}}(1|1)$.
8. Evaluate the performance by using (2.49) to obtain $\mathbf{P}(1)$.
9. Return to step (2) and begin again for other $\mathbf{z}$.

*See Chapter 4, Section 4, for the identities used and an example of such a reduction.

The success of this method now depends upon how well the perturbation $\phi(t)$ follows the real trajectory. Notice that $\phi(t)$ is a second-order effect on this, since it depends on the second partial of the nonlinearity. Note also that the greater the uncertainty in the system, the greater the $\phi(t)$ that must be applied. Also note that $\phi(t)$ will always be of exponential form if the system is time-invariant.

Example. (from Athans, Bertolini, and Wishner). An incoming ballistic missile is to be tracked by a radar that measures range. From this measurement an estimate of the altitude and downward velocity is sought. The equations of motion of the craft are given by

$$\dot{x}_1 = -x_2(t) \tag{2.50}$$

$$\dot{x}_2 = -\frac{C_D A \rho}{2m} x_2^2(t) \tag{2.51}$$

where x_1 is the altitude above the surface of the earth and x_2 the downward velocity. C_D is a drag coefficient; A, the drag area; ρ, the density of the atmosphere; and m, the mass of the craft. The measurement is that of range, which for any time t is given by

$$r(t) = \sqrt{D^2 + [x_1(t) - H]^2} \tag{2.52}$$

where D is the distance form the radar sight to the line of fall and H is the distance of the radar from the surface of the earth. The measurements are made at a discrete set of times $\{kT\}$; thus,

$$z(k+1) = h(\mathbf{x}(k+1)) + w(k+1) \tag{2.53}$$

where

$$h(\mathbf{x}(k+1)) = [D^2 + [x_1(k+1) - H]^2]^{1/2} \tag{2.54}$$

where $\mathbf{x}(k)$ is the state vector

$$\mathbf{x}(k) = \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} \tag{2.55}$$

Using the technique developed in this section, a comparison between the performance of this technique, an extended Kalman filter, and actual performance was made. The results appear in Figure 6.23. A noticeable decrease in performance error is noted by use of this technique. This curve gives the performance of the filter in terms of the rms error. These errors are actual errors that result from a simulation. They are not calculated covariances but averaged simulation errors. Clearly, the inclusion of the driving term decreases the errors significantly. Also note that the theoretical errors are understimating the actual error performance.

This concludes the discussion of estimation for continuous systems with discrete measurements. An alternate analysis has been done by Jaswinski

Figure 6.23 Comparison of performances.

[1, 2] and by Culver [1, 2] and is outlined in the problems. Their approach differs from that discussed in this section in that it starts form an analysis of the conditional density equation for discrete measurements. Jaswinski [1] obtains a noticeable difference in the estimation equations and this is a direct of the fact that the discrete measurement has Gaussian random variables for noise vectors rather than white noise processes. This immediately eliminates the second-order effects we had to consider in Chapter 5. Culver [1] uses Jaswinski's result and completes it with the use of quasi-moment functions to obtain discrete measurement estimators that can function well for severe measurement nonlinearities.

In the next section, we consider both discret-time measurements and discrete-time systems. This extension completes the possibilities of the systems presented to the designer for possible analyis.

6.3 MAXIMUM A POSTERIORI TECHNIQUES

The preceding methods have all been a straightforward attempt to obtain a realizable approximation to the optimum filter in an MMSE sense. This section will deviate from that path and show the reader that there are other methods that may not be as theoretically exact but are computationally more efficacious. The computational aspects of the problem are therefore to be

considered first, and the performance will be a secondary matter, ascertained and evaluated only as a perturbation to the solution.

To obtain the MMSE estimate, it was first necessary to obtain a propagation equation for the conditional probability density of the state. The analysis of the continuous-time problem presented many difficulties in obtaining such a propagation equation, but with the help of the Ito calculus, one was evaluated. The analysis of discrete-time problems present less of a theoretical problem. Furthermore, for many actual applications, discrete-time models are representative of the dynamics of the situation and thus may be preferred. For these reasons we shall concentrate on this type of problem.

As before, we will center our interest around linear filters, since conceptual operation and actual implementation are much simpler. The continual emphasis on the linear is a serious problem when we encounter severe nonlinearities. Such trade-offs as time and accuracy must be made if one is to succeed.

In this section we shall review the work of Cox; Bryson and Frazier; Mowery; and Neal [1, 2]; and discuss some of their implementations and suggestions. Of these, Neal [1] is the only one to consider second-order variations of the nonlinearity. Bryson and Frazier were among the first to realize that Kalman's results could readily be obtained by using the maximum-likelihood approach. They effectively set up an optimum control problem with an appropriate cost function and followed through on the results. Cox presented a simpler, but much more extensive, coverage of the same problem but solved it for the discrete case. His paper is undoubtedly the most readable and his conclusions the most useful for actual computer implementation. Both Mowery and Neal extend Cox's results in different directions. Mowery suggests a method whereby he seeks a linear processor of a certain form and then proceeds to generate a cost criterion suitable for that method. This method may not be very rigorously pleasing, but it provides an adequate answer. Cox's results also depend upon an understanding of the techniques used in dynamic programming. We shall not introduce this material but refer the interested to the book by Bellman and Kalaba. The linearized results do not depend upon dynamic programming, so this section can be read independently of knowledge of this topic.

Recall that the discrete nonlinear system and measurement was modeled by the following set of vector equations:

$$\mathbf{x}(k+1) = \mathbf{f}(\mathbf{x}(k), k) + \mathbf{n}(k) \tag{3.1}$$

$$\mathbf{z}(k) = \mathbf{h}(\mathbf{x}(k), k) + \mathbf{w}(k) \tag{3.2}$$

where $\mathbf{n}(k)$ is an $n \times 1$ Gaussian random variable with covariance $\mathbf{Q}(k)$ and $\mathbf{w}(k)$ is an $m \times 1$ Gaussian random variable with covariance $\mathbf{R}(k)$.

Now, for MMSE estimation we obtained the estimate of the state by eval-

uating the conditional probability density function of $\mathbf{x}(n)$, given all previous data, that is, $\mathbf{z}(0), \ldots, \mathbf{z}(n)$. This in general required a knowledge of the joint probability density function of the $\mathbf{x}(k)$ conditioned on a knowledge of the $\mathbf{z}(k)$. What we shall do here is to consider only the probability density function and see if from its appearance anything can be said regarding the estimate. What we are seeking is

$$p_{\mathbf{x}}(\mathbf{x}(0),\cdots,\mathbf{x}(n)|\mathbf{z}(0),\cdots,\mathbf{z}(n)): \qquad \forall n \tag{3.3}$$

the posterior density of the states $\mathbf{x}(0),\cdots,\mathbf{x}(n)$. But this equals

$$\begin{aligned} &p_{\mathbf{x}|\mathbf{z}}(\mathbf{x}(0),\cdots,\mathbf{x}(n)|\mathbf{z}(0),\cdots,\mathbf{z}(n)) \\ &= \frac{p_{\mathbf{z}|\mathbf{x}}(\mathbf{z}(0),\cdots,\mathbf{z}(n)|\mathbf{x}(0),\cdots,\mathbf{x}(n))p_{\mathbf{x}}(\mathbf{x}(0),\cdots,\mathbf{x}(n))}{p_{\mathbf{z}}(\mathbf{z}(0),\cdots,\mathbf{z}(n))} \end{aligned} \tag{3.4}$$

Now, instead of trying to obtain this expression exactly, which we have already noted for the continuous case to be quite difficult, consider obtaining those $\mathbf{x}(k)$ that maximize this density. Clearly, if the system were linear, the $\mathbf{x}(k)$ that maximize this quantity would also be the conditional means and thus the MMSE estimates, given the data set $\mathbf{z}(k)$. However, for nonlinear systems such a density may be either multimodal—that is, have several local maxima—or the highest point may not correspond to the conditional mean. Thus, the set of $\mathbf{x}(k)$ obtained by maximizing the posterior density function may not be the MMSE estimates. Yet, for singly peaked posterior densities they may be quite close to them. Furthermore, the simplification obtained by this method warrants its inclusion.

The estimates obtained by maximizing the posterior density are called *maximum a posteriori* (MAP) *estimates*. The continuous verion of MAP estimation is presented in Van Trees [2] and in Detchmendy and Sridhar.

It should also be noted that this form of estimate relies only upon maximizing the numerator of the probability density function. As we shall find, this numerator is easily obtained for the given discrete Markov system. The maximization will not be arbitrary, beacuse we must constrain the $\mathbf{x}(k)$ obtained so that they obey the system trajectories given by (3.1). Thus, the problem of estimation is reduced to a constrained discrete-time optimization problem. The estimation problem was first considered in this context by Bryson and Frazier in 1962 for the continuous version. Their results rely heavily upon techniques developed in optimal control theory (see Athans and Falb).

Let us begin by factoring the conditional densities of the $\mathbf{z}(i)$ into the following form:

$$\begin{aligned} &p_{\mathbf{z}}(\mathbf{z}(0), \cdots, \mathbf{z}(n)|\mathbf{x}(0), \cdots, \mathbf{x}(n)) \\ &= p_{\mathbf{z}}(\mathbf{z}(n)|\mathbf{x}(0), \cdots, \mathbf{x}(n), \mathbf{z}(n-1), \cdots, \mathbf{z}(0)) \\ &\quad p_{\mathbf{z}}(\mathbf{z}(n-1)|\mathbf{x}(0), \cdots, \mathbf{x}(n), \mathbf{z}(n-2), \cdots, \mathbf{z}(0)), \cdots, \\ &\quad p_{\mathbf{z}}(\mathbf{z}(0)|\mathbf{x}(0), \cdots, \mathbf{x}(n)) \end{aligned} \tag{3.5}$$

But from (3.2) we see that each $\mathbf{z}(i)$, given $\mathbf{x}(i)$, is a Gaussian random vector. Also $\mathbf{z}(i)$, given $\mathbf{x}(i)$, depends upon no other parameter; that is, it is independent of the other $\mathbf{x}(j)$ and $\mathbf{z}(j)$. Therefore, each of the factors in (3.5) on the right-hand side are Gaussian densities. The means of these densities are easily obtained again from (3.2). The expected values are

$$E[\mathbf{z}(k)] = \mathbf{h}(\mathbf{x}(k), k) \tag{3.6}$$

Likewise, the variances of the random variables is the variance of the noise, which is $\mathbf{R}(k)$. Therefore, (3.5) becomes

$$\begin{aligned} &p_{\mathbf{z}}(\mathbf{z}(0), \cdots, \mathbf{z}(n)|\mathbf{x}(0), \cdots, \mathbf{x}(n)) \\ &= C_1 \exp\Big[-\tfrac{1}{2}\sum_{i=1}^{n}[\mathbf{z}(i) - \mathbf{h}(\mathbf{x}(i), i)]^T\mathbf{R}^{-1}(i)[\mathbf{x}(i) - \mathbf{h}(\mathbf{x}(i), i)]\Big] \end{aligned} \tag{3.7}$$

where C_1 is a normalization constant. In a similar fashion $p_{\mathbf{x}}(\mathbf{x}(0), \cdots, \mathbf{x}(n))$ can be factored as

$$\begin{aligned} p_{\mathbf{x}}(\mathbf{x}(0), \cdots, \mathbf{x}(n)) = {} & p_{\mathbf{x}}(\mathbf{x}(n)|\mathbf{x}(n-1), \cdots, \mathbf{x}(0)) \\ & \cdot p_{\mathbf{x}}(\mathbf{x}(n-1)|\mathbf{x}(n-2), \cdots, \mathbf{x}(0)) \cdots \\ & p_{\mathbf{x}}(\mathbf{x}(1)|\mathbf{x}(0)) \cdot p_{\mathbf{x}}(\mathbf{x}(0)) \end{aligned} \tag{3.8}$$

But since the process is Markov, this becomes

$$\begin{aligned} p_{\mathbf{x}}(\mathbf{x}(0), \cdots, \mathbf{x}(n)) = {} & p_{\mathbf{x}}(\mathbf{x}(n)|\mathbf{x}(n-1)) \\ & p_{\mathbf{x}}(\mathbf{x}(n-1)|\mathbf{x}(n-2)), \cdots, p_{\mathbf{x}}(\mathbf{x}(0)) \end{aligned} \tag{3.9}$$

Now, $\mathbf{x}(i)$, given $\mathbf{x}(i-1)$, is Gaussian if the model follows (3.1) and if $\mathbf{n}(k)$ is Gaussian, which it is by assumption. Thus, the expected value of $\mathbf{x}(i)$, given $\mathbf{x}(i-1)$, is

$$\mathbf{f}(\mathbf{x}(i-1), i-1) \tag{3.10}$$

and the covariance of the corresponding Gaussian form is

$$\mathbf{Q}(k) = E\big[\{\mathbf{x}(k) - E[\mathbf{x}(k)]\}\{\mathbf{x}(k) - E[\mathbf{x}(k)]\}^T\big] \tag{3.11}$$

Finally, we shall assume that $\mathbf{x}(0)$ is Gaussian also with a mean $\mathbf{m}$ and a covariance matrix $\mathbf{P}(0)$. Using these in (3.9), we obtain

$$\begin{aligned} p_{\mathbf{x}}(\mathbf{x}(0), \cdots, \mathbf{x}(n)) = C_2 \exp\Big[& -\tfrac{1}{2}\sum_{i=1}^{n}(\mathbf{x}(i) - \mathbf{f}(\mathbf{x}(i-1), i-1))^T\mathbf{R}^{-1}(i) \\ & (\mathbf{x}(i) - \mathbf{f}(\mathbf{x}(i-1), i-1)) - \tfrac{1}{2}(\mathbf{x}(0) - \mathbf{m})^T\mathbf{P}^{-1}(0)(\mathbf{x}(0) - \mathbf{m})\Big] \end{aligned} \tag{3.12}$$

where $\mathbf{m}$ is the a priori mean of $\mathbf{x}(0)$, which is assumed known. Again C_2 is normalization constant. The only factor we have left to fully complete (3.4) is the denominator that is the joint density function for the output vectors. But recall that what we are interested in is a maximization on $\mathbf{x}(k)$, that is, to find those $\mathbf{x}(k)$ that maximize the a posteriori conditional probability density function. Thus, the denominator that is solely $\mathbf{z}(k)$-dependent will act merely as a normalizing constant for any $\mathbf{x}(k)$ maximization. Therefore,

we can lump all these constants into one large constant called C and rewrite (3.4) as follows:

$$p_{\mathbf{x}|\mathbf{z}}(\mathbf{x}(0), \cdots, \mathbf{x}(n)|\mathbf{z}(0), \cdots, \mathbf{z}(n)) = C \exp\Big[-\tfrac{1}{2}\sum_{i=0}^{n-1} \|\mathbf{x}(i+1) - \mathbf{f}(\mathbf{x}(i), i)\|^2_{\mathbf{Q}^{-1}(i)} - \tfrac{1}{2}\sum_{i=0}^{n} \|\mathbf{z}(i) - \mathbf{h}(\mathbf{x}(i), i)\|^2_{\mathbf{R}^{-1}(i)} - \tfrac{1}{2}\|x(0) - m\|^2_{P^{-1}(0)}\Big] \tag{3.13}$$

Here we have used the notation

$$\|\mathbf{x}\|^2_{\mathbf{Q}} \triangleq \mathbf{x}^T\mathbf{Q}\mathbf{x} \tag{3.14}$$

Now we want to find those $\mathbf{x}$ that *maximize* this conditional probability density function. It should be obvious that those $\mathbf{x}$ that do this should also *minimize* the following function:

$$J_n = \tfrac{1}{2}\|\mathbf{x}(0) - \mathbf{m}\|^2_{\mathbf{P}^{-1}(0)} + \tfrac{1}{2}\sum_{i=0}^{n-1} \|\mathbf{x}(i+1) - \mathbf{f}(\mathbf{x}(i), i)\|^2_{\mathbf{Q}^{-1}(i)} + \tfrac{1}{2}\sum_{i=0}^{n} \|\mathbf{z}(i) - \mathbf{h}(\mathbf{x}(i), i)\|^2_{\mathbf{R}^{-1}(i)} \tag{3.15}$$

Such a minimization may be difficult because of the coupling between $\mathbf{x}(i+1)$ and $\mathbf{x}(i)$. This can be avoided if we realize that

$$\mathbf{x}(i+1) - \mathbf{f}(\mathbf{x}(i), i) = \mathbf{n}(i) \tag{3.16}$$

Then (3.15) becomes

$$J_n = \tfrac{1}{2}\|\mathbf{x}(0) - \mathbf{m}\|^2_{\mathbf{P}^{-1}(0)} + \tfrac{1}{2}\sum_{i=0}^{n-1} \|\mathbf{n}(i)\|^2_{\mathbf{Q}^{-1}(i)} + \tfrac{1}{2}\sum_{i=0}^{n} \|\mathbf{z}(i) - \mathbf{h}(\mathbf{x}(i), i)\|^2_{\mathbf{R}^{-1}(i)} \tag{3.17}$$

The maximization now occurs over all the $\mathbf{x}(i)$ and $\mathbf{n}(i)$. At the same time, we are not arbitrarily free to choose any $\mathbf{x}(i)$ and $\mathbf{n}(i)$: we must choose only those that satisfy the system trajectory. Therefore, we must append to our optimization a system constraint that ensures us that not only are we obtaining a minimization but that in obtaining it we stay on the path defined by the system constraints. Following Cox, we shall use the Lagrange multipliers $\lambda(k)$. Thus, the constrained cost function becomes I_n, defined as

$$I_n = \tfrac{1}{2}\|\mathbf{x}(0) - \mathbf{m}\|^2_{P^{-1}(0)} + \tfrac{1}{2}\sum_{i=0}^{n} \|\mathbf{z}(i) - \mathbf{h}(x(i), i)\|_{\mathbf{R}^{-1}(i)} + \sum_{i=0}^{n-1}\{\tfrac{1}{2}\|\mathbf{n}(i)\|_{\mathbf{Q}^{-1}(i)} + \lambda^T(i)[\mathbf{x}(i+1) - \mathbf{f}(\mathbf{x}(i), i) - \mathbf{n}(i)]\} \tag{3.18}$$

We shall now use a variational approach in order to find the optimum values of $\mathbf{x}(i)$ and $\mathbf{n}(i)$ to minimize the above expression. Before doing so, the reader should take note of two important facts. First, our system was driven

only stochastically; thus if a known drive or forcing function were applied to the system, one would have to take this into account only when it affects the mean of the state. This is obvious, and one would see that this forcing function would appear unaltered in the final filtering equations. The second point is that of the noise being nonzero mean. This is merely a special case of the above forcing function problem. It is handled in like manner by appending to the mean its appropriate value.

We shall use a variational approach to the solution of the optimization. Let us take the partial derivative with respect to $\mathbf{x}(0)$ first:

$$\begin{aligned}\frac{\partial I_n}{\partial \mathbf{x}(0)} &= \mathbf{P}^{-1}(0)(\mathbf{x}(0) - \mathbf{m}) \\ &\quad - \frac{\partial \mathbf{h}(\mathbf{x}(0), 0)}{\partial \mathbf{x}(0)} \mathbf{R}^{-1}(0)(\mathbf{z}(0) - \mathbf{h}(\mathbf{x}(0), 0)) \\ &\quad - \frac{\partial \mathbf{f}(\mathbf{x}(0), 0)}{\partial \mathbf{x}(0)} \boldsymbol{\lambda}(0)\end{aligned} \tag{3.19}$$

Define

$$\frac{\partial \mathbf{h}(\mathbf{x}(0), 0)}{\partial \mathbf{x}(0)} = \mathbf{C}(\mathbf{x}(0)) \tag{3.20}$$

and

$$\frac{\partial \mathbf{f}(\mathbf{x}(0), 0)}{\partial \mathbf{x}(0)} = \mathbf{A}(\mathbf{x}(0)) \tag{3.21}$$

where $\mathbf{C}(\mathbf{x}(0))$ and $\mathbf{A}(\mathbf{x}(0))$ are $m \times n$ and $n \times n$ matrices, respectively. Thus,

$$\frac{\partial \mathbf{h}(\mathbf{x}(0), 0)}{\partial \mathbf{x}(0)} = \begin{bmatrix} \dfrac{\partial h_1(\mathbf{x}(0), 0)}{\partial x_1(0)} & \cdots & \dfrac{\partial h_1(\mathbf{x}(0), 0)}{\partial x_n(0)} \\ \dfrac{\partial h_m(\mathbf{x}(0), 0)}{\partial x_1(0)} & \cdots & \dfrac{\partial h_m(\mathbf{x}(0), 0)}{\partial x_n(0)} \end{bmatrix} \tag{3.22}$$

and a similar expression results for $\mathbf{A}(\mathbf{x}(0))$. Now equating (3.19) to zero, we obtain

$$\mathbf{P}^{-1}(0)(\mathbf{x}(0) - \mathbf{m}) = \mathbf{C}(\mathbf{x}(0))\mathbf{R}^{-1}(0)(\mathbf{z}(0) - \mathbf{h}(\mathbf{x}(0), 0)) + \mathbf{A}(\mathbf{x}(0))\boldsymbol{\lambda}(0) \tag{3.23}$$

Premultiplying by $\mathbf{P}(0)$ yields

$$\mathbf{x}(0) = \mathbf{m} + \mathbf{P}(0)\mathbf{C}(\mathbf{x}(0))\mathbf{R}^{-1}(0)(\mathbf{z}(0) - \mathbf{h}(\mathbf{x}(0), 0)) + \mathbf{P}(0)\mathbf{A}(\mathbf{x}(0))\boldsymbol{\lambda}(0) \tag{3.24}$$

Use the notation $\hat{\mathbf{x}}(0|n)$ as the optimum estimate of $\mathbf{x}(0)$, given data up to n. This may not be directly evident in (3.24), but as we see what $\boldsymbol{\lambda}(0)$ is, we shall see that indeed $\hat{\mathbf{x}}(0)$ depends on all $\mathbf{z}(i)$, where $i = 0, \cdots, n$. Therefore, using this notation, (3.24) becomes

$$\begin{aligned}\hat{\mathbf{x}}(0|n) = \mathbf{m} + \mathbf{P}(0) &+ \mathbf{C}(\hat{\mathbf{x}}(0|n))\mathbf{R}^{-1}(0) \cdot (\mathbf{z}(0) - \mathbf{h}(\hat{\mathbf{x}}(0|n), 0)) \\ &+ \mathbf{P}(0)\mathbf{A}((\hat{\mathbf{x}}(0|n))\boldsymbol{\lambda}(0)\end{aligned} \tag{3.25}$$

By definition

$$\boldsymbol{\lambda}(n) = \mathbf{0} \tag{3.26}$$

Let us obtain now a stable point that will maximize I_n for an $\hat{\mathbf{x}}(k|n)$, where $k \in (1,2,\cdots,n)$. For any k

$$\frac{\partial I_n}{\partial \mathbf{x}(k)} = \frac{\partial \mathbf{h}(x(k), k)}{\partial \mathbf{x}(k)} \mathbf{R}^{-1}(k)(\mathbf{z}(k) - \mathbf{h}(\mathbf{x}(k), k)) + \boldsymbol{\lambda}(k-1) - \frac{\partial \mathbf{f}(x(k), k))}{\partial \mathbf{x}(k)} \boldsymbol{\lambda}(k) \tag{3.27}$$

Again this derivative must be equal to zero. We shall use the notation of $\hat{\mathbf{x}}(k|n)$ as the value of the state at that point. This then yields for $\boldsymbol{\lambda}(k-1)$ the relationship

$$\boldsymbol{\lambda}(k-1) = \mathbf{A}(\hat{\mathbf{x}}(k|n))\,\boldsymbol{\lambda}(k) + \mathbf{C}(\hat{\mathbf{x}}(k|n))\mathbf{R}^{-1}(k)(\mathbf{z}(k) - \mathbf{h}(\hat{\mathbf{x}}(k|n), k)) \tag{3.28}$$

This is a *backward* recursive relationship that will yield $\boldsymbol{\lambda}(i)$ for all i. Note that here $\boldsymbol{\lambda}(0)$ will contain all the information necessary from the n, $\mathbf{z}$ measurements. Thus, as we initially conjectured, $\hat{\mathbf{x}}(0|n)$ is indeed influenced by all $\mathbf{z}$ from $\mathbf{z}(0)$ to $\mathbf{z}(n)$. Now let us take the derivative of (16) with respect to the $\mathbf{n}(k)$, where $k = 0, \cdots, n$. This gives us our final set of equations to solve for a maximum.

$$\frac{\partial I_n}{\partial \mathbf{n}(k)} = \mathbf{Q}^{-1}(k)\mathbf{n}(k) - \boldsymbol{\lambda}(k) \tag{3.29}$$

Setting this term equal to zero yields

$$\mathbf{Q}^{-1}(k)\mathbf{n}(k) = \boldsymbol{\lambda}(k) \tag{3.30}$$

Then, multiplying both sides by $\mathbf{Q}(k)$, we obtain

$$\mathbf{n}(k) = \mathbf{Q}(k)\,\boldsymbol{\lambda}(k) \tag{3.31}$$

which yields a solution for the optimum $\mathbf{n}(k)$ in terms of the Lagrange multipliers. Along the optimum trajectory we must have

$$\hat{\mathbf{x}}(k+1) = \mathbf{f}(\hat{\mathbf{x}}(k), k) + \mathbf{n}(k) \tag{3.32}$$

But this must hold for the optimum $\mathbf{n}(k)$. Using (3.31) in (3.32) for $\mathbf{n}(k)$ and recalling that this is really an estimate of $\mathbf{x}(k+1)$ based upon n data points, we obtain

$$\hat{\mathbf{x}}(k+1|n) = \mathbf{f}(\hat{\mathbf{x}}(k|n), k) + \mathbf{Q}(k)\,\boldsymbol{\lambda}(k) \tag{3.33}$$

where $\mathbf{x}(k|n)$ represents the kth time estimate, given n data points. Now equations (3.33), (3.28), (3.26), and (3.25) can be used to solve recursively for the "optimal" estimate. Unfortunately, even these equations are quite difficult to manage, and they require further simplification. Cox presents a technique whereby use of dynamic programming allows one in principle to obtain a

solution. Unfortunately, such a solution may be far from practical and require extensive computer time and memory. Therefore, we are placed in the position of obtaining an approximation to the solution in a form that will be palatable to the machine-user.

We now derive a set of linearized equations that give a straightforward solution to the filtering problem. What is done is to follow Cox's approach, which is in general the approach followed by almost all other methods. That is, we shall expand the recursive state equations about the last estimate. Thus, the reader should be made aware of the fact that this expansion point may not be optimal and should be prepared to "play" with added biases on simulations to see if they result in significant improvements.

Now let us take (1) and expand it about the point termed the last approximate estimate, called $\mathbf{x}^*(k|k)$ rather than $\hat{\mathbf{x}}(k|k)$. This will yield

$$\mathbf{x}(k+1) \simeq \mathbf{f}(\mathbf{x}^*(k|k), k) + \mathbf{A}(\mathbf{x}^*(k|k))(\mathbf{x}(k) - \mathbf{x}^*(k|k)) + \mathbf{n}(k) \tag{3.34}$$

for any k. Here we have dropped the higher-order terms of the expansion. This again assumes that the nonlinearity is sufficiently smooth and that our expansion point is "close" to the real trajectory. Define a new cost criterion based upon the new linearized system of equations where for simplicity we assume that the measurements are linear in the state vector. This yields

$$I_n^* = \tfrac{1}{2}\|\mathbf{x}(0) - \mathbf{m}\|^2_{\mathbf{P}^{-1}(0)} + \sum_{i=0}^{n} \tfrac{1}{2}\|\mathbf{z}(i) - \mathbf{C}(i)\mathbf{x}(i)\|^2_{\mathbf{R}^{-1}(i)} + \sum_{i=0}^{n-1}\{\tfrac{1}{2}\|\mathbf{n}(i)\|^2_{\mathbf{Q}^{-1}(i)} + [\mathbf{x}(i+1) - \mathbf{f}(\mathbf{x}^*(i|i), i) - \mathbf{A}(\mathbf{x}^*(i|i))\cdot(\mathbf{x}(i) - \mathbf{x}^*(i|i)) - \mathbf{n}(i)]^T\boldsymbol{\lambda}(i)\} \tag{3.35}$$

In a fashion similar to the previous analysis, we can find those $\mathbf{x}(i)$, $\mathbf{n}(i)$ that minimize the expression subject to the constraints. The result of these operations is the following set of equations:

$$\mathbf{x}^*(k+1|n) = \mathbf{f}(\mathbf{x}^*(k|k), k) + \mathbf{A}(\mathbf{x}^*(k|k))(\mathbf{x}^*(k|n) - \mathbf{x}^*(k|k)) + \mathbf{Q}(k)\boldsymbol{\lambda}(k) \tag{3.36}$$

$$\boldsymbol{\lambda}(k-1) = \mathbf{A}(\mathbf{x}^*(k|k))\boldsymbol{\lambda}(k) + \mathbf{C}^T(k)\mathbf{R}^{-1}(k)\cdot(\mathbf{z}(k) - \mathbf{C}(k)\mathbf{x}^*(k|n)) \tag{3.37}$$

$$\mathbf{x}^*(0|n) = \mathbf{m} + \mathbf{P}(0)\mathbf{C}^T\mathbf{R}^{-1}(0)(\mathbf{z}(0) - \mathbf{C}(0)\mathbf{x}^*(0|n)) + \mathbf{P}(0)\mathbf{A}^T(0)\boldsymbol{\lambda}(0) \tag{3.38}$$

and

$$\boldsymbol{\lambda}(n) = \mathbf{0} \tag{3.39}$$

This is a two-point boundary-value problem, which will be solved in four steps. First, we shall define a linear system similar to the linearized system. Second, we shall augment the linear system with a *known* forcing function $\mathbf{v}(k)$. This will give us a new solution to the optimization only in that now $\mathbf{v}(k)$ must appear in (3.37). Third, we shall solve this linear problem. Fourth, we shall identify $\mathbf{v}(k)$ with the known part of our nonlinear expansion and im-

mediately obtain the nonlinear solution. This method differs from that of Cox in that it allows the reader to see that the nonlinear solution is known and further, that external known controls can be handled by this estimation procedure.

Consider the system defined by

$$\mathbf{x}(k+1) = \mathbf{A}(k)\mathbf{x}(k) + \mathbf{v}(k) + \mathbf{n}(k) \tag{3.40}$$

where $\mathbf{A}(k)$ is an $n \times n$ matrix, $\mathbf{v}(k)$ is an $n \times 1$ known forcing function, and $\mathbf{n}(k)$ is the Gaussian noise vector. We shall let

$$\mathbf{v}(k) = \mathbf{f}(\mathbf{x}^*(k|k), k) - \mathbf{A}(\mathbf{x}^*(k|k))\mathbf{x}^*(k|k) \tag{3.41}$$

and

$$\mathbf{A}(\mathbf{x}^*(k|k)) \triangleq \mathbf{A}(k) \tag{3.42}$$

Equation (3.42) states that $\mathbf{v}(k)$ is the residual "drive" between the nonlinear trajectory and the linearized trajectory. If we have a linear system, then

$$\mathbf{f}(\mathbf{x}^*(k|k), k) = \mathbf{A}(k)\mathbf{x}^*(k|k) \tag{3.43}$$

so that $\mathbf{v}(k)$ would be zero for all linear systems. For a nonlinear system it is that "extra push" that is necessary for our linear trajectory to keep up with the nonlinear one. For the present assume $\mathbf{v}(k)$ is nonzero. It could be possible that it may even include a real forcing function, and in that case, we would have to augment (3.41) with that knowledge. It should be immediately obvious then that (3.36) and (3.37) become

$$\mathbf{x}^*(k+1|n) = \mathbf{A}(k)\mathbf{x}^*(k|n) + \mathbf{v}(k) + \mathbf{Q}(k)\boldsymbol{\lambda}(k) \tag{3.44}$$

and

$$\boldsymbol{\lambda}(k-1) = \mathbf{A}(k)\boldsymbol{\lambda}(k) + \mathbf{C}^T(k)\mathbf{R}^{-1}(k)\cdot(\mathbf{z}(k) - \mathbf{C}(k)\mathbf{x}^*(k|n)) \tag{3.45}$$

Recall that for the linear case, $\mathbf{x}^*(k|n)$ is indeed $\hat{\mathbf{x}}(k|n)$. We shall return to the starred notation, recalling it to mind when necessary.

Our boundary conditions are

$$\mathbf{x}^*(0|n) = \mathbf{m} + \mathbf{P}(0)\mathbf{C}^T(0)\mathbf{R}^{-1}(0)(\mathbf{z}(0) - \mathbf{C}(0)\mathbf{x}^*(0|n)) + \mathbf{P}(0)\mathbf{A}^T(0)\boldsymbol{\lambda}(0) \tag{3.46}$$

and

$$\boldsymbol{\lambda}(n) = \mathbf{0} \tag{3.47}$$

We will use an induction proof after having established some initial trends. Let us first solve for $\mathbf{x}^*(0|0)$. Using (3.46).

$$\mathbf{x}^*(0|0) = \mathbf{m} + \mathbf{P}(0)\mathbf{C}^T(0)\mathbf{R}^{-1}(0)(\mathbf{z}(0) - \mathbf{C}(0)\,\mathbf{x}^*(0|0)) + \mathbf{P}(0)\mathbf{A}^T(0)\boldsymbol{\lambda}(0) \tag{3.48}$$

But $\boldsymbol{\lambda}(0)$ for $\mathbf{x}^*(0|0)$ is 0, since $\boldsymbol{\lambda}(0)$ is a function of the measurements and $\mathbf{x}^*(0|0)$ implies that there are no measurements—in this case $\boldsymbol{\lambda}(n) = \boldsymbol{\lambda}(0) = \mathbf{0}$ since $n = 0$,—therefore using this in (3.48), we obtain

$$(\mathbf{I} + \mathbf{P}(0)\mathbf{C}^T(0)\mathbf{R}^{-1}(0)\mathbf{C}(0))\mathbf{x}^*(0|0) = \mathbf{m} + \mathbf{P}(0)\mathbf{C}^T(0)\mathbf{R}^{-1}(0)\mathbf{z}(0) \tag{3.49}$$

Now add and subtract

$$\mathbf{P}(0)\mathbf{C}^T(0)\mathbf{R}^{-1}(0)\mathbf{C}(0)\mathbf{m} \tag{3.50}$$

yielding

$$(\mathbf{I} + \mathbf{P}(0)\mathbf{C}^T(0)\mathbf{R}^{-1}(0)\mathbf{C}(0))\mathbf{x}^*(0|0) = (\mathbf{I} + \mathbf{P}(0)\mathbf{C}^T(0)\mathbf{R}^{-1}(0)\mathbf{C}(0))\mathbf{m} + \mathbf{P}(0)\mathbf{C}^T(0)\mathbf{R}^{-1}(0)(\mathbf{z}(0) - \mathbf{C}(0)\mathbf{m}) \tag{3.51}$$

Now premultiply both sides by

$$\mathbf{M}(0) \triangleq (\mathbf{I} + \mathbf{P}(0)\mathbf{C}^T(0)\mathbf{R}^{-1}(0)\mathbf{C}(0))^{-1} \tag{3.52}$$

to yield a more familiar form;

$$\mathbf{x}^*(0|0) = \mathbf{m} + \mathbf{K}(0)\mathbf{C}^T(0)\mathbf{R}^{-1}(0)(\mathbf{z}(0) - \mathbf{C}(0)\mathbf{m}) \tag{3.53}$$

where

$$\mathbf{K}(0) = \mathbf{M}(0)\mathbf{P}(0) \tag{3.54}$$

Using (3.46), we obtain for the general case of $n \times 1$ measurements

$$\mathbf{x}^*(0|n) = \mathbf{m} + \mathbf{P}(0)\mathbf{C}^T(0)\mathbf{R}^{-1}(0)\mathbf{z}(0) - \mathbf{P}(0)\mathbf{C}^T(0)\mathbf{R}^{-1}(0)\mathbf{C}(0)\mathbf{x}^*(0|n) + \mathbf{P}(0)\mathbf{A}^T(0)\boldsymbol{\lambda}(0) \tag{3.55}$$

In this case $\boldsymbol{\lambda}(0)$ is not $\mathbf{0}$, since $n \neq 0$, but $\boldsymbol{\lambda}(n) = \mathbf{0}$.

Again add and subtract

$$\mathbf{P}(0)\mathbf{C}^T(0)\mathbf{R}^{-1}(0)\mathbf{C}(0)\mathbf{m}$$

to obtain

$$\mathbf{M}^{-1}(0)\mathbf{x}^*(0|n) = \mathbf{M}^{-1}(0)\mathbf{m} + \mathbf{P}(0)\mathbf{C}^T(0)\mathbf{R}^{-1}(0)(\mathbf{z}(0) - \mathbf{C}(0)\mathbf{m}) + \mathbf{P}(0)\mathbf{A}^T(0)\boldsymbol{\lambda}(0) \tag{3.56}$$

Multiply through by $\mathbf{M}(0)$:

$$\mathbf{x}^*(0|n) = \mathbf{m} + \mathbf{K}(0)\mathbf{C}^T(0)\mathbf{R}^{-1}(0)(\mathbf{z}(0) - \mathbf{C}(0)\mathbf{m}) + \mathbf{K}(0)\mathbf{A}^T(0)\boldsymbol{\lambda}(0) \tag{3.57}$$

But, using (3.53) for the first two terms on the right of (3.57) yields

$$\mathbf{x}^*(0|n) = \mathbf{x}^*(0|0) + \mathbf{K}(0)\mathbf{A}^T(0)\boldsymbol{\lambda}(0) \tag{3.58}$$

Recall that in (3.58) $\boldsymbol{\lambda}(0)$ depends on all the $\mathbf{z}(i)$, whereas in (3.48) $\boldsymbol{\lambda}(0)$ was $\mathbf{0}$. $\boldsymbol{\lambda}(0)$ depends on the condition of data. Now, let us do this for $\mathbf{x}^*(1|1)$. Using (3.44), we obtain

$$\mathbf{x}^*(1|1) = \mathbf{A}(0)\mathbf{x}^*(0|1) + \mathbf{v}(0) + \mathbf{Q}(0)\boldsymbol{\lambda}(0) \tag{3.59}$$

Also, using (3.45),

$$\boldsymbol{\lambda}(0) = \mathbf{A}(1)\boldsymbol{\lambda}(1) + \mathbf{C}^T(1)\mathbf{R}^{-1}(1)(\mathbf{z}(1) - \mathbf{C}(1)\mathbf{x}^*(1|1)) \tag{3.60}$$

Now $n = 1$, we have $\boldsymbol{\lambda}(1) = \mathbf{0}$. Therefore, we have $\boldsymbol{\lambda}(0)$ from (3.60) to use in (3.59). Also, from (3.58) we have $\mathbf{x}^*(0|1)$. It is

$$\mathbf{x}^*(0|1) = \mathbf{x}^*(0|0) + \mathbf{K}(0)\mathbf{A}^T(0)\boldsymbol{\lambda}(0) \tag{3.61}$$

Using (3.61) in (3.59), we obtain

$$\mathbf{x}^*(1|1) = \mathbf{A}(0)\mathbf{x}^*(0|0) + \mathbf{v}(0) + (\mathbf{A}(0)\mathbf{K}(0)\mathbf{A}^T(0) + \mathbf{Q}(0))\boldsymbol{\lambda}(0) \tag{3.62}$$

Define $\mathbf{P}(1)$ as

$$\mathbf{P}(1) = \mathbf{A}(0)\mathbf{K}(0)\mathbf{A}^T(0) + \mathbf{Q}(0) \tag{3.63}$$

Then using our knowledge of $\boldsymbol{\lambda}(0)$ from (3.60), we obtain

$$\mathbf{x}^*(1|1) = \mathbf{A}(1)\mathbf{x}^*(0|0) + \mathbf{v}(0) + \mathbf{P}(1)[\mathbf{C}^T(1)\mathbf{R}^{-1}(1)(\mathbf{z}(1) - \mathbf{C}(1)\mathbf{x}^*(1|1))] \tag{3.64}$$

Now solve (3.64) for $\mathbf{x}^*(1|1)$ by defining

$$\mathbf{M}(1) = (\mathbf{I} + \mathbf{P}(1)\mathbf{C}^T(1)\mathbf{R}^{-1}(1)\mathbf{C}(1))^{-1} \tag{3.65}$$

and

$$\mathbf{K}(1) = \mathbf{M}(1)\mathbf{P}(1) \tag{3.66}$$

Then

$$\begin{aligned}\mathbf{x}^*(1|1) = \mathbf{A}(1)\mathbf{x}^*(0|0) + \mathbf{v}(0) + \mathbf{A}(1)\mathbf{C}^T(1)\mathbf{R}^{-1}(1)\\ (\mathbf{z}(1) - \mathbf{C}(1)(\mathbf{A}(1)\mathbf{x}^*(0|0) + \mathbf{v}(0)))\end{aligned} \tag{3.67}$$

Again, in (3.67) we added and subtracted equal terms. Now let us generalize. We see that (3.63) will give us $\mathbf{P}(k+1)$ as a function of $\mathbf{P}(k)$ through $\mathbf{K}(k)$. Let us begin by obtaining $\mathbf{x}^*(1|n)$. Now, by (3.44),

$$\mathbf{x}^*(1|n) = \mathbf{A}(0)\mathbf{x}^*(0|n) + \mathbf{v}(0) + \mathbf{Q}(0)\boldsymbol{\lambda}(0) \tag{3.68}$$

But from (3.58)

$$\mathbf{x}^*(0|n) = \mathbf{x}^*(0|0) + \mathbf{K}(0)\mathbf{A}^T(0)\boldsymbol{\lambda}(0) \tag{3.69}$$

Then

$$\mathbf{x}^*(1|n) = \mathbf{A}(0)\mathbf{x}^*(0|0) + \mathbf{v}(0) + (\mathbf{A}(0)\mathbf{K}(0)\mathbf{A}^T(0) + \mathbf{Q}(0))\boldsymbol{\lambda}(0) \tag{3.70}$$

But, by (3.63), this is

$$\mathbf{x}^*(0|n) = \mathbf{A}(0)\mathbf{x}^*(0|0) + \mathbf{v}(0) + \mathbf{Q}(1)\boldsymbol{\lambda}(0) \tag{3.71}$$

Now, $\boldsymbol{\lambda}(0)$ is given by (3.45):

$$\boldsymbol{\lambda}(0) = \mathbf{A}^T(1)\boldsymbol{\lambda}(1) + \mathbf{C}^T(1)\mathbf{R}^{-1}(1)(\mathbf{z}(1) - \mathbf{C}(1)\mathbf{x}^*(1|n)) \tag{3.72}$$

Using (3.72) in (3.71), we obtain

$$\begin{aligned}\mathbf{x}^*(1|n) = \mathbf{A}(0)\mathbf{x}^*(0|0) + \mathbf{v}(0) + \mathbf{P}(1)\mathbf{A}^T(1)\boldsymbol{\lambda}(1)\\ + \mathbf{P}(1)\mathbf{C}^T(1)\mathbf{R}^{-1}(1)\mathbf{z}(1) - \mathbf{P}(1)\mathbf{C}^T(1)\mathbf{R}^{-1}(1)\mathbf{C}(1)\cdot\mathbf{x}^*(1|n)\end{aligned} \tag{3.73}$$

Factoring, we obtain

$$\begin{aligned}&(\mathbf{I} + \mathbf{P}(1)\mathbf{C}^T(1)\mathbf{R}^{-1}(1)\mathbf{C}(1))\mathbf{x}^*(1|n)\\ &= \mathbf{A}(0)\mathbf{x}^*(0|0) + \mathbf{v}(0) + \mathbf{P}(1)\mathbf{A}^T(1)\boldsymbol{\lambda}(1) + \mathbf{P}(1)\mathbf{C}^T(1)\mathbf{R}^{-1}(1)\mathbf{z}(1)\end{aligned} \tag{3.74}$$

Add and subtract the term

$$\mathbf{P}(1)\mathbf{C}^T(1)\mathbf{R}^{-1}(1)\mathbf{C}(1)(\mathbf{A}(0)\mathbf{x}^*(0|0) + \mathbf{v}(0)) \tag{3.75}$$

and recall that we have defined $\mathbf{M}(1)$ in (3.65) to give

$$\begin{aligned}\mathbf{M}^{-1}(1)\mathbf{x}^*(1|n) = {} & \mathbf{M}^{-1}(1)[\mathbf{A}(0)\mathbf{x}^*(0|0) + \mathbf{v}(0)] \\ & + \mathbf{P}(1)\mathbf{C}^T(1)\mathbf{R}^{-1}(1)(\mathbf{z}(1) - \mathbf{C}(1)(\mathbf{A}(1)\mathbf{x}^*(0|0) + \mathbf{v}(1))) \\ & + \mathbf{P}(1)\mathbf{A}^T(1)\boldsymbol{\lambda}(1)\end{aligned} \tag{3.76}$$

Then premultiplying both sides by $\mathbf{M}(1)$ yields

$$\begin{aligned}\mathbf{x}^*(1|n) = {} & \mathbf{A}(0)\mathbf{x}^*(0|0) + \mathbf{v}(0) + \mathbf{K}(1)\mathbf{C}^T(1)\mathbf{R}^{-1}(1) \\ & (\mathbf{z}(1) - \mathbf{C}(1)(\mathbf{A}(1)\mathbf{x}^*(0|0) + \mathbf{x}(0))) + \mathbf{K}(1)\mathbf{A}^T(1)\boldsymbol{\lambda}(1)\end{aligned} \tag{3.77}$$

But, using (3.67) for $\mathbf{x}^*(1|1)$, we have

$$\mathbf{x}^*(1|n) = \mathbf{x}^*(1|1) + \mathbf{K}(1)\mathbf{A}^T(1)\boldsymbol{\lambda}(1) \tag{3.78}$$

Now, comparing (3.78) to (3.58), we see that in general

$$\mathbf{x}^*(k|n) = \mathbf{x}^*(k|k) + \mathbf{K}(k)\mathbf{A}^T\boldsymbol{\lambda}(k) \tag{3.79}$$

where

$$\mathbf{P}(k) = \mathbf{A}(k-1)\mathbf{K}(k-1)\mathbf{A}^T(k-1) + \mathbf{Q}(k-1) \tag{3.80}$$

and

$$\mathbf{K}(k) = [\mathbf{I} + \mathbf{P}(k)\mathbf{C}^T(k)\mathbf{R}^{-1}(k)\mathbf{C}(k)]^{-1}\mathbf{P}(k) \tag{3.81}$$

Let us prove this by induction. Assume that it is true for $(k-1)$ and mimic the previous proof to show that it holds true for k. Starting from (3.59), we have for $k-1$

$$\mathbf{x}^*(k|k) = \mathbf{A}(k-1)\mathbf{x}^*(k-1|k) + \mathbf{v}(k-1) + \mathbf{Q}(k-1)\boldsymbol{\lambda}(k-1) \tag{3.82}$$

Also,

$$\boldsymbol{\lambda}(k-1) = \mathbf{A}(k)\boldsymbol{\lambda}(k) + \mathbf{C}^T(k)\mathbf{R}^{-1}(k)(\mathbf{z}((k) - \mathbf{C}(k)\mathbf{x}^*(k|k))) \tag{3.83}$$

Since $n = k$, $\boldsymbol{\lambda}(k) = 0$. And since we assumed (3.79), then we obtain

$$\mathbf{x}^*(k|k) = \mathbf{A}(k-1)\mathbf{x}^*(k-1|k-1) + \mathbf{K}(k-1)\mathbf{A}^T(k-1)\boldsymbol{\lambda}(k-1) \tag{3.84}$$

Using (3.84) in (3.82), we have

$$\begin{aligned}\mathbf{x}^*(k|k) = {} & \mathbf{A}(k-1)\mathbf{x}^*(k-1|k-1) + \mathbf{v}(k-1) \\ & + (\mathbf{A}(k-1)\mathbf{K}(k-1)\mathbf{A}^T(k-1) + \mathbf{Q}(k-1))\boldsymbol{\lambda}(k-1)\end{aligned} \tag{3.85}$$

But using (3.80), (3.85) becomes

$$\mathbf{x}^*(k|k) = \mathbf{A}(k-1)\mathbf{x}^*(k-1|k-1) + \mathbf{v}(k-1) + \mathbf{P}(k)\boldsymbol{\lambda}(k-1) \tag{3.86}$$

Now using the knowledge of $\boldsymbol{\lambda}(k-1)$, we have

$$\begin{aligned}\mathbf{x}^*(k|k) = {} & \mathbf{A}(k-1)\mathbf{x}^*(k-1|k-1) + \mathbf{v}(k-1) \\ & + \mathbf{P}(k)[\mathbf{C}^T(k)\mathbf{R}^{-1}(k)[\mathbf{z}(k) - \mathbf{C}(k)\mathbf{x}^*(k|k)]]\end{aligned} \tag{3.87}$$

Mimicking (3.65–3.67), we obtain

$$\mathbf{x}^*(k|k) = \mathbf{A}(k-1)\mathbf{x}^*(k-1|k-1) + \mathbf{v}(k-1) + \mathbf{K}(k)\mathbf{C}^T(k)\mathbf{R}^{-1}(k)$$
$$(\mathbf{z}(k) - \mathbf{C}(k)(\mathbf{A}(k-1)\mathbf{x}^*(k-1|k-1) + \mathbf{v}(k-1))) \quad (3.88)$$

Now again

$$\mathbf{x}^*(k|n) = \mathbf{A}(k-1)\mathbf{x}^*(k-1|n) + \mathbf{v}(k-1) + \mathbf{Q}(k-1)\boldsymbol{\lambda}(k-1) \quad (3.89)$$

But from hypothesis (3.79)

$$\mathbf{x}^*(k-1|n) = \mathbf{x}^*(k-1|k-1) + \mathbf{K}(k-1)\mathbf{A}^T(k-1)\boldsymbol{\lambda}(k-1) \quad (3.90)$$

Using (3.90) in (3.89), we have

$$\mathbf{x}^*(k|n) = \mathbf{A}(k-1)\mathbf{x}^*(k-1|k-1)$$
$$+ [\mathbf{A}(k-1)\mathbf{K}(k-1)\mathbf{A}^T(k-1) + \mathbf{Q}(k-1)]\boldsymbol{\lambda}(k-1) \quad (3.91)$$

Using (3.80) we obtain,

$$\mathbf{x}^*(k|n) = \mathbf{A}(k-1)\mathbf{x}^*(k-1|k-1) + \mathbf{v}(k-1) + \mathbf{P}(k)\boldsymbol{\lambda}(k-1) \quad (3.92)$$

But

$$\boldsymbol{\lambda}(k-1) = \mathbf{A}^T(k)\boldsymbol{\lambda}(k) + \mathbf{C}^T(k)\mathbf{R}^{-1}(k)(\mathbf{z}(k) - \mathbf{C}(k)\mathbf{x}^*(k|n)) \quad (3.93)$$

Using (3.93) in (3.92), we obtain

$$\mathbf{x}^*(k|n) = \mathbf{A}(k-1)\mathbf{x}^*(k-1|k-1) + \mathbf{v}(k-1) + \mathbf{P}(k)\mathbf{A}^T(k)\boldsymbol{\lambda}(k)$$
$$+ \mathbf{P}(k)\mathbf{C}^T(k)\mathbf{R}^{-1}(k)(\mathbf{z}(k) - \mathbf{C}(k)\mathbf{x}^*(k|n)) \quad (3.94)$$

By rearranging as before and following the definitions, we get

$$\mathbf{x}^*(k|n) = \mathbf{x}^*(k|k) + \mathbf{K}(k)\mathbf{A}^T(k)\boldsymbol{\lambda}(k) \quad (3.95)$$

which is what we hypothesized. Thus, by induction, this is true for all k. We can now obtain the filtering equations:

$$\mathbf{x}^*(k|k) = \mathbf{A}(k-1)\mathbf{x}^*(k-1|k) + \mathbf{v}(k-1) + \mathbf{Q}(k-1)\boldsymbol{\lambda}(k-1) \quad (3.96)$$

But from (3.95)

$$\mathbf{x}^*(k-1|k) = \mathbf{x}^*(k-1|k-1) + \mathbf{K}(k-1)\mathbf{A}^T(k-1)\boldsymbol{\lambda}(k-1) \quad (3.97)$$

Now substituting,

$$\mathbf{x}^*(k|k) = \mathbf{A}(k-1)\mathbf{x}^*(k-1|k-1) + \mathbf{v}(k-1)$$
$$+ [\mathbf{A}(k-1)\mathbf{C}(k-1)\mathbf{A}^T(k-1) + \mathbf{Q}(k-1)]\boldsymbol{\lambda}(k-1) \quad (3.98)$$

Now $\boldsymbol{\lambda}(k)$ equals zero since $k = n$. Therefore,

$$\boldsymbol{\lambda}(k-1) = \mathbf{C}^T(k)\mathbf{R}^{-1}(k)[\mathbf{z}(k) - \mathbf{C}(k)\mathbf{x}^*(k|k)] \quad (3.99)$$

Using (3.90) in (3.98) and (3.99) in (3.98), we have

$$\mathbf{x}^*(k|k) = \mathbf{A}(k-1)\mathbf{x}^*(k-1|k-1) + \mathbf{v}(k-1)$$
$$+ \mathbf{P}(k)\mathbf{C}^T(k)\mathbf{R}^{-1}(k)[\mathbf{z}(k) - \mathbf{C}(k)\mathbf{x}^*(k|k)] \quad (3.100)$$

Solving for $\mathbf{x}^*(k|k)$ as before, we obtain

$$\mathbf{x}^*(k|k) = \mathbf{A}(k-1)\mathbf{x}^*(k-1|k-1) + \mathbf{v}(k-1) + \mathbf{K}(k)\mathbf{C}^T(k)\mathbf{R}^{-1}(k) \\ [\mathbf{z}(k) - \mathbf{C}(k)\cdot[\mathbf{A}(k-1)\mathbf{x}^*(k-1|k-1) + \mathbf{v}(k-1)]] \quad (3.101)$$

Now (3.101), (3.80), and (3.91) fully define the filter.

Let us now return to the nonlinear case. Recall that $\mathbf{v}(k)$ was given by (3.41). Therefore,

$$\mathbf{v}(k-1) = \mathbf{f}(\mathbf{x}^*(k-1|k-1)) \\ - \mathbf{A}(\mathbf{x}^*(k-1|k-1))\mathbf{x}^*(k-1|k-1) \quad (3.102)$$

and

$$\mathbf{A}(k-1)\mathbf{x}^*(k-1|k-1) + \mathbf{v}(k-1) = \mathbf{f}(\mathbf{x}^*(k-1|k-1)) \quad (3.103)$$

Therefore, (3.101) becomes

$$\mathbf{x}^*(k|k) = \mathbf{f}(\mathbf{x}^*(k-1|k-1)) \\ + \mathbf{K}(k)\mathbf{C}^T(k)\mathbf{R}^{-1}(k)[\mathbf{z}(k) - \mathbf{C}(k)\mathbf{f}(\mathbf{x}^*(k-1|k-1))] \quad (3.104)$$

which is the nonlinear filtering equation. $\mathbf{K}(k)$ and $\mathbf{P}(k)$ are obtained from (3.80) and (3.81).

The implementation of this filter is quite simple. All that it requires is an evaluation of the nonlinearity and does not include the sensitivity matrix discussed in the last two sections. For this reason one might suspect that the results may not be as accurate as those including more knowledge of the nonlinearity. Also, the chief drawback of this method is that one has no knowledge of how well he is performing with regard to the covariance of the estimates. There does not seem to be a simple answer to this problem.

Again the results reduce in the linear case to the discrete-time Kalman filter of Chapter 4 Section 4.3, as they should, providing a simple check on the validity of the equations.

If we were to return to the representation theorem of Bucy, we could follow through with a similar analysis for the continuous filtering equations. Again it would be a modal technique for Gaussian and a least-squares technique for anything else. As mentioned, this analysis was carried out by Bryson and Frazier before Bucy gave the theoretical justifications to their cost criterion. The results of Bryson and Frazier are then justified and provide one with a continuous-time filter. If the reader is interested in other approaches of this form, he should look at the work of Mowery, Neal, and Detchmendy and Sridhar, for example, and see how modified cost criteria yield slightly different, but generally consistent, results.

Treatment of measurement nonlinearities is covered in Neal [2]; it follows a similar linearization routine. In general, for scientific purposes, as was stated, a linear measurement is attempted to avoid later reduction problems.

The case of nonlinear measurements is important in the communication context, as is evidenced in the work of Snyder. Therefore, the results provided in this section should cover the range of general scientific interest.

6.4 FILTER INACCURACIES

The previous three sections were devoted to a study of the structure of various filter structures. At each step we found it helpful to linearize the equations and deal with a simpler set of problems. However, when this is done certain effects may occur that will make the resulting filter perform poorly. In this section we examine a class of these effects, namely, divergence and stability. We will deal mainly with the linear discrete-time system because of its computational importance.

The problem of divergence concerns the effects of such things as model uncertainties on the filtering equations and the resulting diverging estimate from the true state. Stability concerns the time behavior of the estimate equation itself and is considered in a purely deterministic fashion.

There seem to be no general techniques for analyzing the effects of divergence and of stability. The divergence problem especially may take so many diverse forms, each with its own characteristics, that a general solution, if it exists, may be inadequate, and cataloging particular solutions would be time-and space-consuming to the extent of utter boredom. For this reason, we have chosen to indicate a single method of approach and rely upon the user's ingenuity to see him through the maze of confusion and obtain a good feeling for his particular filter. This is a serious problem, though, since in general we cannot be always certain of our models. In this section we will catalog several of these uncertainties and give references that will give insight into the evaluation of their effects.

The ideas that we have presented in this and the previous chapters should provide the reader with the tools necessary for the definition, solution, and implementation of the estimation problem. What we shall do in the final section of this chapter is to discuss some further applications of the tools that were developed. Undoubtedly the reader may be cognizant of many more, so that this list is merely meant to show the breath of applications.

Let us first reintroduce the linear estimator equations for the Kalman solution (See Chapter 4, Section 4.3). They are

$$\begin{aligned}\hat{\mathbf{x}}(k+1) = {} & [\mathbf{I} - \mathbf{K}(k+1)\mathbf{C}(k+1)]\boldsymbol{\Phi}(k+1,k)\hat{\mathbf{x}}(k) \\ & + \mathbf{K}(k+1)\mathbf{z}(k+1)\end{aligned} \tag{4.1}$$

$$\begin{aligned}\mathbf{P}(k+1) = {} & [\mathbf{I} - \mathbf{K}(k+1)\mathbf{C}(k+1)]\boldsymbol{\Phi}(k+1,k)\mathbf{P}(k)\boldsymbol{\Phi}^T(k+1,k) \\ & [\mathbf{I} - \mathbf{K}(k+1)\mathbf{C}(k+1)]^T + [\mathbf{I} - \mathbf{K}(k+1)\mathbf{C}(k+1)]\mathbf{Q}(k) \\ & [\mathbf{I} - \mathbf{K}(k+1)\mathbf{C}(k+1)]^T + \mathbf{K}(k+1)\mathbf{R}(k+1)\mathbf{K}^T(k+1)\end{aligned} \tag{4.2}$$

and

$$\mathbf{K}(k+1) = \boldsymbol{\Phi}(k+1,k)\mathbf{P}(k)\boldsymbol{\Phi}^T(k+1,k)\mathbf{C}^T(k+1) \cdot [\mathbf{C}(k+1)\boldsymbol{\Phi}(k+1,k)\mathbf{P}(k)\boldsymbol{\Phi}^T(k+1,k)\mathbf{C}^T(k+1) + \mathbf{C}(k+1)\mathbf{Q}(k)\mathbf{C}^T(k+1) + \mathbf{R}(k+1)]^{-1} \tag{4.3}$$

where the system model is

$$\mathbf{x}(k+1) = \boldsymbol{\Phi}(k+1,k)\mathbf{x}(k) + \mathbf{u}(k) \tag{4.4}$$

and

$$\mathbf{z}(k+1) = \mathbf{C}(k+1)\mathbf{x}(k+1) + \mathbf{v}(k+1) \tag{4.5}$$

and the covariance matrices are

$$E[\mathbf{u}(k)\mathbf{u}^T(k)] = \mathbf{Q}(k) \tag{4.6}$$

$$E[\mathbf{v}(k)\mathbf{v}^T(k)] = \mathbf{R}(k) \tag{4.7}$$

and

$$E[\mathbf{x}(0)\mathbf{x}^T(0)] = \mathbf{P}(0) \tag{4.8}$$

and $\mathbf{x}(0)$, $\mathbf{u}(k)$, and $\mathbf{v}(k)$ are all zero mean independent Gaussian random variables. In general, we assume that we know the following about the model:

1. $\mathbf{K}(k+1)$ assumes perfect knowledge of $\boldsymbol{\Phi}(k+1,k)$, $\mathbf{C}(k+1)$, $\mathbf{Q}(k)$, and $\mathbf{R}(k+1)$.
2. $\mathbf{P}(k)$ assumes perfect knowledge of all of that for $\mathbf{K}(k)$ and also $\mathbf{P}(0)$.
3. $\hat{\mathbf{x}}(k)$ assumes all for $\mathbf{P}(k)$ and $\mathbf{K}(k+1)$.

Now the following things may not be known:

1. *Model inaccuracies*:
   a. The components of $\boldsymbol{\Phi}(k+1,k)$ may not be known perfectly; that is,

$$\boldsymbol{\Phi}(k+1,k) = \boldsymbol{\Phi}^*(k+1,k) + \delta\boldsymbol{\Phi}(k+1,k) \tag{4.9}$$

   where $\boldsymbol{\Phi}(k+1,k)$ is the true transition matrix and $\boldsymbol{\Phi}^*(k+1,k)$ is an approximation to it. The term $\delta\boldsymbol{\Phi}(k+1,k)$ is assumed to be additive. Indeed, it may not. It may be possible to consider $\delta\boldsymbol{\Phi}(k+1,k)$ to be merely a noise term and use $\boldsymbol{\Phi}^*(k+1,k)$ with just extra noise driving the system.
   b. The state variables may only be known in part. For example the model as given in (4.4) may be

$$\begin{bmatrix}\mathbf{x}_1(k+1)\\ \mathbf{x}_2(k+1)\end{bmatrix} = \begin{bmatrix}\boldsymbol{\phi}_{11}(k+1,k) & \boldsymbol{\phi}_{12}(k+1,k)\\ \boldsymbol{\phi}_{21}(k+1,k) & \boldsymbol{\phi}_{22}(k+1,k)\end{bmatrix}\begin{bmatrix}\mathbf{x}_1(k)\\ \mathbf{x}_2(k)\end{bmatrix} + \begin{bmatrix}\mathbf{u}_1(k)\\ \mathbf{u}_2(k)\end{bmatrix} \tag{4.10}$$

   and all we know is $\boldsymbol{\phi}_{22}(k+1,k)$ and $\boldsymbol{\phi}_{11}$, $\boldsymbol{\phi}_{12}$, $\boldsymbol{\phi}_{21}$ are unknown. Here $\mathbf{x}_1(k)$ and $\mathbf{x}_2(k)$ are *vectors* and not individual components.

In this case, an optimum filter requires knowledge of all the state variables. For example, we may want to estimate $\mathbf{x}_1(k+1)$ and all we know is the propogation of $\mathbf{x}_2(k+1)$. An important example this is in atmosphere studies where $\mathbf{x}_2$ may be intensities and $\mathbf{x}_1$ are particle densities. We may want to estimate both $\mathbf{x}_1$ and $\mathbf{x}_2$ based upon only measurements of the $\mathbf{x}_2$.

c. The measurement matrix may not be accurately known. In general, this may not be too serious a problem, since the designer has a great deal of control over its construction.

2. *Noise inaccuracies*:

a. We may be uncertain about the accuracy of the initial covariance: As stated such inaccuracies may influence $\mathbf{P}(k)$, but if $\mathbf{P}(k)$ is stable, lack of $\mathbf{P}(0)$ certainty may disappear as a serious defect.

b. $\mathbf{Q}(k)$ and $\mathbf{R}(k)$ inaccuracies may be quite common. The system noises may be self-induced by the designer to compensate for inaccuracies in his model, for linearization approximation, or for actual phenomenological noises. In general, this noise is the most difficult to model since most scientific experiments are of an exploratory nature to begin with and the phenomena are being investigated, leaving perturbing effects until last. The measurement noise is, in general, the easiest to model and monitor. Examples of phenomenon noise and measurement noise are given by a seismic example where the phenomenon noise may result from seismometers and other measuring equipment.

3. *Roundoff errors*:

These are computational errors associated with the actual machine computation of the estimates. They may be quite serious if we are seeking a highly accurate solution. They are briefly discussed in Bucy and Joseph (pp. 174–176).

We shall now proceed to analyze a particular example from those defined above and see what effects play a dominant role. In particular, we shall study a combination of (1b) and (2b). Using (4.10) if we only know $\boldsymbol{\phi}_{22}(k+1, k)$, then we model the state system by

$$\mathbf{x}_2(k+1) = \boldsymbol{\phi}_{22}(k+1, k)\mathbf{x}_2(k) + \mathbf{f}_2(k) + \mathbf{u}_2(k) \tag{4.11}$$

where $\mathbf{f}_2(k)$ is some deterministic function used to show our lack of knowledge of $\boldsymbol{\phi}_{21}(k+1, k)$ and $\mathbf{x}_1(k)$ and their effect on $\mathbf{x}_2(k+1)$. Likewise, $\mathbf{u}_2^*(k)$ is an approximate noise term used to model the noise in the model. The measurement equation measures only the effects of $\mathbf{x}_2(k+1)$ and does not infer any knowledge of $\mathbf{x}_1(k+1)$.

$$\mathbf{z}(k+1) = \mathbf{C}(k+1)\mathbf{x}_2(k+1) + \mathbf{v}(k+1) \tag{4.12}$$

The actual model for $\mathbf{x}_2(k+1)$ is given by

$$\mathbf{x}_2(k+1) = \phi_{22}(k+1, k)\mathbf{x}_2(k) + \mathbf{f}_2(k) + \mathbf{u}_2(k) \tag{4.13}$$

where

$$\mathbf{f}_2(k) = \phi_{21}(k+1, k)\mathbf{x}_1(k) \tag{4.14}$$

Now $\mathbf{f}_2(k)$ can be written with a deterministic part, its mean, and a random portion given by

$$\mathbf{u}_3(k) = \mathbf{f}_2(k) - E[\mathbf{f}_2(k)] \tag{4.15}$$

which clearly has zero mean. Thus, the approximate model in (4.11) is an attempt to compensate for the actual dynamics, which we either do not know or we deliberately neglect for a simpler structure.

One way in which this type of modeling error frequently arises is in the definition of dynamical systems based upon the measurement of the power spectrum. Recall that the scalar system

$$\dot{x}_1(t) = -\alpha x_1(t) + u(t) \tag{4.16}$$

has a power spectrum of the form

$$S_{x_1}(\omega) = \frac{C}{\omega^2 + \alpha^2} \tag{4.17}$$

where C is an appropriate scaling constant. The inverse can also be true: that is, we can model the process $x_1(t)$ in state variable form if we know the spectrum $S_{x_1}(\omega)$. This can be extended to higher-order spectra by noting that the state variable system

$$\dot{x}_1 = -x_2 \tag{4.18}$$

$$\dot{x}_2 = -\alpha x_2 + u(t) \tag{4.19}$$

has a power spectrum for x_1 of

$$S_{x_1}(\omega) = \frac{C}{\omega^2(\omega^2 + \alpha_2)} \tag{4.20}$$

Similarly, given $S_{x_1}(\omega)$, we could construct a nonunique state variable realization. This procedure is called spectral factorization (see Brockett or Van Trees [1]). Thus, given any power spectrum, we could construct a suitable state variable realization where the number of state variables is related to the number of poles in the power spectrum. Now, if a process $x_1(t)$ has a power spectrum

$$S_{x_1}(\omega) = \frac{C}{(\omega^2 + \alpha^2)(\omega^2 + \beta^2)} \tag{4.21}$$

we know that a two-state-variable system would be necessary to realize this. But, if in measuring $S_{x_1}(\omega)$ we measure $S^*_{x_1}(\omega)$ where

$$S_{x_1}^*(\omega) = \frac{C}{(\omega^2 + \alpha^2)} \tag{4.22}$$

then our state variable realization will be in error. Such a measurement could result if $\alpha \gg \beta$. Thus, the real state variable model may be

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \mathbf{u}(t) \tag{4.23}$$

and our guessed state variable model is

$$\dot{x}_1^* = -\alpha x_1^* + u^*(t) \tag{4.24}$$

This represents only one way in which this model may result in an erroneous system. It should be pointed out though that this analysis assumes that although there is uncertainty in the state system equation, it is known that the measurement contains only the state variables represented by the limited system knowledge. For example, in the discussion above, we know the measurement is of $x_1(t)$ and is given by

$$z(t) = C(t)x_1(t) + v(t) \tag{4.25}$$

But our error occurs in assuming that $x_1(t)$ can be represented by a single-pole spectrum rather than a multiple-pole spectrum. Now we shall suppress the 2's in (4.11) and consider it as the state equation. The covariance matrices are then given by

$$E[\mathbf{u}(k)\mathbf{u}^T(k)] = \mathbf{Q}(k) \tag{4.26}$$

and

$$E[\mathbf{v}(k)\mathbf{v}^T(k)] = \mathbf{R}(k) \tag{4.27}$$

But $\mathbf{f}(k)$, $\mathbf{Q}(k)$, and $\mathbf{R}(k)$ are assumed unknown. Now we shall assume some $\mathbf{f}^*(k)$, $\mathbf{Q}^*(k)$ and $\mathbf{R}^*(k)$ for these values and they will differ from the true values. Now, using $\mathbf{u}^*(k)$ and $\mathbf{v}^*(k)$ to represent the approximate noises associated with the appropriate covariance matrices, we have for a system model

$$\mathbf{x}^*(k+1) = \boldsymbol{\Phi}(k+1, k)\mathbf{x}^*(k) + \mathbf{f}^*(k) + \mathbf{u}^*(k) \tag{4.28}$$

$$\mathbf{z}^*(k+1) = \mathbf{C}(k+1)\mathbf{x}^*(k+1) + \mathbf{v}^*(k+1) \tag{4.29}$$

This measurement equation is also hypothetical, since it is based upon the assumption that $\mathbf{z}(k+1)$ depends linearly upon $\mathbf{x}^*(k+1)$, the hypothetical system model.

The covariances associated with this hypothetical model are defined as

$$E[\mathbf{u}^*(k)\mathbf{u}^{*T}(k)] = \mathbf{Q}^*(k) \tag{4.30}$$

$$E[\mathbf{v}^*(\mathbf{k})\mathbf{v}^{*T}(k)] = \mathbf{R}^*(k) \tag{4.31}$$

We now want to obtain a Kalman filter to be used to obtain an estimate $\hat{\mathbf{x}}^*(k)$ of $\mathbf{x}^*(k)$ that we shall use for $\mathbf{x}(k)$. If we actually received $\mathbf{z}^*(k)$, this would be trivial. Yet we receive not $\mathbf{z}^*(k)$ but $\mathbf{z}(k)$. Thus, we shall substitute $\mathbf{z}(k)$ for $\mathbf{z}^*(k)$, just as we shall reverse the process by substituting $\hat{\mathbf{x}}(k)$ for $\hat{\mathbf{x}}^*(k)$. Then the filtering equation becomes

$$\hat{\mathbf{x}}^*(k+1) = [\mathbf{I} - \mathbf{K}^*(k+1)\mathbf{C}(k+1)][\boldsymbol{\phi}(k+1,k)\hat{\mathbf{x}}^*(k) + \mathbf{f}^*(k)] + \mathbf{K}^*(k+1)\mathbf{z}(k+1) \tag{4.32}$$

where

$$\begin{aligned}\mathbf{P}^*(k+1) = {} & [\mathbf{I} - \mathbf{K}^*(k+1)\mathbf{C}(k+1)]\boldsymbol{\Phi}(k+1,k)\mathbf{P}^*(k)\boldsymbol{\Phi}^T(k+1,k) \\ & [\mathbf{I} - \mathbf{K}^*(k+1)\mathbf{C}(k+1)]^T + [\mathbf{I} - \mathbf{K}^*(k+1)\mathbf{C}(k+1)]\mathbf{Q}^*(k) \\ & [\mathbf{I} - \mathbf{K}^*(k+1)\mathbf{C}(k+1)]^T + \mathbf{K}^*(k+1)\mathbf{R}^*(k+1)\mathbf{K}^{*T}(k+1)\end{aligned} \tag{4.33}$$

and

$$\begin{aligned}\mathbf{K}^*(k+1) = {} & \boldsymbol{\Phi}(k+1,k)\mathbf{P}^*(k)\boldsymbol{\Phi}^T(k+1,k) \\ & \cdot\mathbf{C}^T(k+1)[\mathbf{C}(k+1)\boldsymbol{\Phi}(k+1,k)\mathbf{P}^*(k)\boldsymbol{\Phi}^T(k+1,k)\mathbf{C}^T(k+1) \\ & + \mathbf{C}(k+1)\mathbf{Q}^*(k)\mathbf{C}^T(k+1) + \mathbf{R}^*(k+1)]^{-1}\end{aligned} \tag{4.34}$$

Now we would like to obtain a performance measure for this estimate. To do so, let us define

$$\bar{\mathbf{P}}(k) = E[(\mathbf{x}(k) - \hat{\mathbf{x}}^*(k))(\mathbf{x}(k) - \hat{\mathbf{x}}^*(k))^T] \tag{4.35}$$

We shall call this a covariance matrix and obtain a propagation equation for it. To evaluate this, we will substitute the values of $\mathbf{x}(k)$ and $\hat{\mathbf{x}}^*(k)$ into the expression Now

$$\bar{\mathbf{P}}(k+1) = E[(\mathbf{x}(k+1) - \hat{\mathbf{x}}^*(k+1))(\mathbf{x}(k+1) - \hat{\mathbf{x}}^*(k+1))^T] \tag{4.36}$$

and define the matrix $\mathbf{D}(k+1)$ as

$$\mathbf{D}(k+1) \triangleq \mathbf{I} - \mathbf{K}^*(k+1)\mathbf{C}^*(k+1) \tag{4.37}$$

Then

$$\begin{aligned}\hat{\mathbf{x}}^*(k+1) = {} & \mathbf{D}(k+1)\boldsymbol{\Phi}(k+1,k)\hat{\mathbf{x}}^*(k) \\ & + \mathbf{D}(k+1)\mathbf{f}^*(k) + \mathbf{K}^*(k+1)\mathbf{z}(k+1)\end{aligned} \tag{4.38}$$

and the actual propagation equation for $\mathbf{x}(k+1)$ is given by

$$\mathbf{x}(k+1) = \boldsymbol{\Phi}(k+1,k)\mathbf{x}(k) + \mathbf{f}(k) + \mathbf{u}(k) \tag{4.39}$$

where we recall that $\mathbf{f}(k)$ represents the effect of the remaining state variables on $\mathbf{x}(k+1)$.

Define $\tilde{\mathbf{x}}(k+1)$ as

$$\tilde{\mathbf{x}}(k+1) = \mathbf{x}(k+1) - \hat{\mathbf{x}}^*(k+1) \tag{4.40}$$

Then using (4.38) and (4.39), we have

$$\tilde{\mathbf{x}}(k+1) = \boldsymbol{\Phi}(k+1,k)\mathbf{x}(k) + \mathbf{u}(k) + \mathbf{f}(k) - \mathbf{D}(k+1) \cdot [\boldsymbol{\Phi}(k+1,k)\tilde{\mathbf{x}}^*(k) + \mathbf{f}^*(k)] - \mathbf{K}^*(k+1)\mathbf{z}(k+1) \tag{4.41}$$

But the real $\mathbf{z}(k+1)$ is given by

$$\mathbf{z}(k+1) = \mathbf{C}(k+1)\,\mathbf{x}(k+1) + \mathbf{v}(k+1) \tag{4.42}$$

Therefore,

$$\tilde{\mathbf{x}}(k+1) = \mathbf{D}(k+1)[\boldsymbol{\Phi}(k+1,k)\tilde{\mathbf{x}}(k) + \mathbf{f}(k) - \mathbf{f}^*(k)] + \mathbf{u}(k) - \mathbf{K}^*(k+1)\,\mathbf{v}(k+1) \tag{4.43}$$

Thus,

$$\tilde{\mathbf{P}}(k+1) = E[\tilde{\mathbf{x}}(k+1)\tilde{\mathbf{x}}^T(k+1)] \tag{4.44}$$

which, using (4.43) and realizing that $\tilde{\mathbf{x}}(k)$, $\mathbf{u}(k)$, and $\mathbf{v}(k+1)$ are all statistically independent and $\mathbf{u}(k)$ and $\mathbf{v}(k+1)$ are zero mean, yields

$$\begin{aligned}\tilde{\mathbf{P}}(k+1) &= \mathbf{D}(k+1)\boldsymbol{\Phi}(k+1,k)\tilde{\mathbf{P}}(k)\boldsymbol{\Phi}^T(k+1,k)\,\mathbf{D}^T(k+1)\\ &+ \mathbf{D}(k+1)\boldsymbol{\Phi}(k+1,k)E[\mathbf{x}(k) - \tilde{\mathbf{x}}^*(k)](\mathbf{f}(k) - \mathbf{f}^*(k))^T\mathbf{D}^T(k+1)\\ &+ \mathbf{D}(k+1)(\mathbf{f}(k) - \mathbf{f}^*(k))E[\mathbf{x}(k) - \tilde{\mathbf{x}}^*(k)]^T\boldsymbol{\Phi}^T(k+1,k)\mathbf{D}^T(k+1)\\ &+ \mathbf{D}(k+1)(\mathbf{f}(k) - \mathbf{f}^*(k))(\mathbf{f}(k) - \mathbf{f}^*(k))^T\mathbf{D}^T(k+1)\\ &+ \mathbf{K}^*(k+1)\mathbf{R}(k+1)\,\mathbf{K}^{*T}(k+1) + \mathbf{Q}(k)\end{aligned} \tag{4.45}$$

We can simplify (4.45) by noting that (4.38) can be written as

$$\tilde{\mathbf{x}}^*(k+1) = \mathbf{D}(k+1)[\boldsymbol{\Phi}(k+1,k)\tilde{\mathbf{x}}^*(k) + \mathbf{f}(k) + [\mathbf{f}^*(k) - \mathbf{f}(k)] + \mathbf{K}^*(k+1)\mathbf{z}(k+1)] \tag{4.46}$$

Now (4.46) is linear, so we can break it up into two parts:

$$\tilde{\mathbf{x}}^*(k+1) = \tilde{\mathbf{x}}_1^*(k+1) + \tilde{\mathbf{x}}_2^*(k+1) \tag{4.47}$$

where

$$\tilde{\mathbf{x}}_1^*(k+1) = \mathbf{D}(k+1)[\boldsymbol{\Phi}(k+1,k)\tilde{\mathbf{x}}_1^*(k) + \mathbf{f}(k)] + \mathbf{K}^*(k+1)\mathbf{z}(k+1) \tag{4.48}$$

and

$$\mathbf{x}_2^*(k+1) = \mathbf{D}(k+1)[\boldsymbol{\Phi}(k+1,k)\tilde{\mathbf{x}}_2^*(k) + [\mathbf{f}^*(k) - \mathbf{f}(k)]] \tag{4.49}$$

with

$$\tilde{\mathbf{x}}_1^*(0) = E[\mathbf{x}(0)] \tag{4.50}$$

and

$$\tilde{\mathbf{x}}_2^*(0) = 0 \tag{4.51}$$

Thus $\tilde{\mathbf{x}}_1^*(k)$ is the estimate of $\tilde{\mathbf{x}}(k)$ for which we have perfect knowledge of $\mathbf{f}(k)$ but inexact knowledge of the measurement and system noise matrices. Therefore,

$$E[\hat{\mathbf{x}}_1^*(k+1)] = \mathbf{D}(k+1)[\boldsymbol{\Phi}(k+1,k)\,E[\hat{\mathbf{x}}_1^*(k)] + \mathbf{f}(k)] + \mathbf{K}^*(k+1)E[\mathbf{C}(k+1)\mathbf{x}(k+1) + \mathbf{v}(k+1)] = \mathbf{D}(k+1)[\boldsymbol{\Phi}(k+1,k)\,E[\hat{\mathbf{x}}_1^*(k)] + \mathbf{f}(k)] + \mathbf{K}^*(k+1)\mathbf{C}(k+1)E[\mathbf{x}(k)] \tag{4.52}$$

Now this can be solved recursively from zero to show that

$$E[\hat{\mathbf{x}}_1^*(k+1)] = E[\mathbf{x}(k+1)] \tag{4.53}$$

which means that $\mathbf{x}_1^*(k+1)$ is an unbiased estimate of $\mathbf{x}(k+1)$. Now one should note that (4.49) and (4.51) imply that $\mathbf{x}_2^*(k)$ is *deterministic*, so that

$$E[\mathbf{x}(k) - \hat{\mathbf{x}}^*(k)] = E[\mathbf{x}(k) - \hat{\mathbf{x}}_1^*(k) - \mathbf{x}_2^*(k)] = E[\mathbf{x}(k) - \hat{\mathbf{x}}_1^*(k)] - \hat{\mathbf{x}}_2^*(k) = -\hat{\mathbf{x}}_2^*(k) \tag{4.54}$$

which is then the bias on the estimate.

It is an *unknown* bias since $\mathbf{f}(k)$ is not known. Note that had $\mathbf{f}(k)$ been known then $\mathbf{x}_2^*(k)$ would be zero. Let us now define two more variables:

$$\Delta\mathbf{m}(k) = E[\mathbf{x}(k) - \hat{\mathbf{x}}^*(k)] \tag{4.55}$$

and

$$\Delta\mathbf{f}(k) = \mathbf{f}(k) - \mathbf{f}^*(k) \tag{4.56}$$

Thus, (4.49) can immediately be rewritten as

$$\Delta\mathbf{m}(k+1) = \mathbf{D}(k+1)[\boldsymbol{\Phi}(k+1,k)\Delta\mathbf{m}(k) + \Delta\mathbf{f}(k)] \tag{4.57}$$

with

$$\Delta\mathbf{m}(0) = \mathbf{0} \tag{4.58}$$

Using this in (4.45), we obtain

$$\begin{aligned}\bar{\mathbf{P}}(k+1) = {} & \mathbf{D}(k+1)\boldsymbol{\Phi}(k+1,k)\bar{\mathbf{P}}(k)\boldsymbol{\Phi}^T(k+1,k)\mathbf{D}^T(k+1) \\ & + \mathbf{D}(k+1)\boldsymbol{\Phi}(k+1)\Delta\mathbf{m}\,\Delta\mathbf{f}^T\mathbf{D}^T(k+1) \\ & + \mathbf{D}(k+1)\Delta\mathbf{f}\,\Delta\mathbf{m}^T\boldsymbol{\Phi}^T(k+1,k)\mathbf{D}^T(k+1) \\ & + \mathbf{D}(k+1)\Delta\mathbf{f}\,\Delta\mathbf{f}^T\,\mathbf{D}^T(k+1) \\ & + \mathbf{K}^*(k+1)\mathbf{R}(k+1)\mathbf{K}^{*T}(k+1) + \mathbf{Q}(k)\end{aligned} \tag{4.59}$$

Now recall that $\mathbf{K}^*(k+1)$ depends on the estimated $\mathbf{R}^*(k)$ and $\mathbf{Q}^*(k)$, and we have in (4.59) the actual values. Now we still do not know $\Delta\mathbf{f}$, since we do not know $\mathbf{f}(k)$. Thus, the only appropriate question we can ask is, Under what circumstances does $\bar{\mathbf{P}}(k+1)$ remain bounded? We do not know what it is but by its form and having only a gross knowledge of $\mathbf{f}(k)$, $\mathbf{Q}(k)$, and $\mathbf{R}(k)$, we can, by stability arguments, give some qualitative statements as to its behavior. In genral, this is all we will be able to do with problems of this sort. Unless we specify more about the form of the unknowns—for example, that they are random variables—this is the end.

We shall analyze the stability of this system in two parts. First, we shall look at the homogeneous part and, then, at the inhomogeneous section. Using the discrete-time Lyapanov theory, we shall show under what conditions (4.59) is u.a.s.i.l.

The homogeneous part of (4.59) is given by

$$\bar{\mathbf{P}}(k+1) = \mathbf{B}(k)\bar{\mathbf{P}}(k)\mathbf{B}^T(k); \qquad \bar{\mathbf{P}}(0) = \mathbf{P}^*(0) \tag{4.60}$$

where

$$\mathbf{B}(k) = \mathbf{D}(k+1)\boldsymbol{\Phi}(k+1,k) = (\mathbf{I} - \mathbf{K}^*(k+1)\mathbf{C}^*(k+1))\boldsymbol{\Phi}(k+1,k) \tag{4.61}$$

We now present several definitions and lemmas preparatory to proving the desired stability.

DEFINITION 4.1. A matrix $\mathbf{P}(k)$ is said to be *bounded* if there exists a finite $\alpha_1 \geqq 0$ such that

$$\mathbf{P}(k) = \max_{i,j} |P_{i,j}(k)| < \alpha_1; \qquad \forall k \tag{4.62}$$

LEMMA 4.1. Equation (4.60) is uniformly asymptotically stable in the large (u.a.s.i.l.) if there exists a finite α_2 and a λ such that

$$\bar{\mathbf{P}}(k) < \alpha_2 e^{-\lambda_1 k}; \qquad \alpha_2 > 0, \lambda_1 > 0 \tag{4.63}$$

Proof. See Theorem 4.13 of Chapter 2. ∎

If we now consider the following equation

$$\mathbf{x}(k+1) = \mathbf{B}(k)\mathbf{x}(k); \qquad \mathbf{x}_0 = \mathbf{x}(0) \tag{4.64}$$

where $\mathbf{x}(k+1)$ is an arbitrary vector, then we can state the following lemma.

LEMMA 4.2. Equation (4.60) is u.a.s.i.l. if (4.64) is u.a.s.i.l. in the vector sense; that is, if

$$\mathbf{x}^T(k)\mathbf{x}(k) \leqq \alpha_3 e^{-\lambda_3 k}; \qquad \forall k \tag{4.65}$$

with $\alpha_3, \lambda_3 > 0$, then $\bar{\mathbf{P}}(k+1)$ is u.a.s.i.l.

The reason for this lemma is that the stability analysis of Chapter 2 was for vector equations and equation (4.60) is a matrix equation. This lemma thus demonstrates the equivalence of the stability of a vector equation and a matrix equation.

Proof. We can begin by finding the solution to (4.60)

$$\bar{\mathbf{P}}(1) = \mathbf{B}(0)\bar{\mathbf{P}}(0)\mathbf{B}^T(0) \tag{4.66}$$

$$\bar{\mathbf{P}}(2) = \mathbf{B}(1)\bar{\mathbf{P}}(1)\mathbf{B}^T(1) = \mathbf{B}(1)\mathbf{B}(0)\bar{\mathbf{P}}(0)\mathbf{B}^T(0)\mathbf{B}(1) \tag{4.67}$$

and, in general,

$$\bar{\mathbf{P}}(k) = \left(\prod_{i=0}^{k-1}\mathbf{B}(i)\right)\mathbf{P}(0)\left(\prod_{i=0}^{k-1}\mathbf{B}(i)\right)^T \tag{4.68}$$

Likewise,

$$\mathbf{x}(k) = \prod_{i=0}^{k-1} \mathbf{B}(i)\mathbf{x}(0) \tag{4.69}$$

Then, if $\mathbf{x}(k)$ is u.a.s.i.l., it implies that for a finite $\mathbf{x}(0)$

$$\left\| \prod_{i=0}^{k-1} \mathbf{B}(i) \right\| < \alpha_0 e^{-\lambda_0 k} \tag{4.70}$$

for $\alpha_0, \lambda_0 > 0$. Therefore, (4.70) implies (4.65), which proves the lemma. ∎

Now if we can show that (4.64) is u.a.s.i.l., or at least under which conditions it is u.a.s.i.l., then we will have solved the problem. This fact is stated in the following theorem, whose proof we divert to Appendix C.

THEOREM 4.1

Let $\mathbf{x}^*(k+1)$ be given by

$$\mathbf{x}^*(k+1) = \boldsymbol{\Phi}(k+1, k)\mathbf{x}^*(k) + \mathbf{f}^*(k) + \mathbf{u}^*(k) \tag{4.71}$$

and let a measurement $\mathbf{z}^*(k+1)$ be given by

$$\mathbf{z}^*(k+1) = \mathbf{C}(k+1)\mathbf{x}^*(k+1) + \mathbf{v}^*(k+1) \tag{4.72}$$

where $\mathbf{R}^*(k+1)$ is the covariance of $\mathbf{v}^*(k+1)$. If there exists positive nonzero constants $\delta_1, \delta_2, \delta_3, \delta_4, \delta_5$ such that for all k

$$\delta_1 \mathbf{I} < \boldsymbol{\Phi}(k+1, k)\boldsymbol{\Phi}^T(k+1, k) < \delta_2 \mathbf{I} \tag{4.73}$$

$$\mathbf{C}(k)\mathbf{C}^T(k) < \delta_3 \mathbf{I} \tag{4.74}$$

$$\delta_4 \mathbf{I} < \mathbf{R}^*(k) < \delta_5 \mathbf{I} \tag{4.75}$$

and if there exist positive nonzero finite constants $\alpha_4, \alpha_5, \alpha_6, \alpha_7$ such that for some N and some $n \leqq N - 1$

$$\alpha_5 \mathbf{I} \leqq \sum_{i=N-n}^{N-1} \boldsymbol{\Phi}(k, i+1)\mathbf{Q}^*(i)\boldsymbol{\Phi}(k, i+1) \leqq \alpha_4 \mathbf{I} \tag{4.76}$$

$$\alpha_7 \mathbf{I} \leqq \sum_{i=N-n}^{N} \boldsymbol{\Phi}^T(i, k)\mathbf{C}^T(i)\mathbf{R}^{*-1}(i)\mathbf{C}(i)\boldsymbol{\Phi}(i, k) \leqq \alpha_6 \mathbf{I} \tag{4.77}$$

then

$$\mathbf{x}(k+1) = \mathbf{B}(k)\mathbf{x}(k) \tag{4.78}$$

is u.a.s.i.l. for all $k \geqq N$.

However, by definition $\mathbf{x}(k+1)$ in the above is also given as

$$\mathbf{x}(k+1) = (\mathbf{I} - \mathbf{K}(k+1)\mathbf{C}(k+1))\boldsymbol{\Phi}(k+1, k)\mathbf{x}(k) \tag{4.79}$$

and represents the unforced portion of the estimate equation. Thus, by showing that this is u.a.s.i.l., we are showing that the resulting Kalman filter is also u.a.s.i.l. This was first suggested by Kalman [1] and finally cor-

rectly proven by Bucy [3]. The proof of the u.a.s.i.l. of this equation employing Lyapanov theory appears in Appendix C.

The two conditions (4.76) and (4.77) appearing in the previous theorem have the appearance of the controllability and observability matrices for deterministic systems developed in Chapter 2. They are in fact called stochastic controllability and observability matrices. Specifically:

DEFINITION 4.2. The matrix $\mathbf{M}_s(N, N-k)$ is called the *observability matrix* for state $x(N)$, given $k+1$ measurements and is given by

$$\mathbf{M}_s(N, N-k) = \sum_{i=N-k}^{N} \boldsymbol{\Phi}^T(i, k)\mathbf{C}^T(i)\mathbf{R}^{-1}(i)\mathbf{C}(i)\boldsymbol{\Phi}(i, k) \tag{4.80}$$

Similarly, we can define the stochastic controllability matrix:

DEFINITION 4.3. The matrix $\mathbf{W}_s(N, N-k)$ is called the *stochastic controllability matrix* for the state $\mathbf{x}(N)$, given k measurments, and is given by

$$\mathbf{W}_s(N, N-k) = \sum_{i=N-k}^{N-1} \boldsymbol{\Phi}(k, i+1)\mathbf{Q}(i)\boldsymbol{\Phi}^T(k, i+1) \tag{4.81}$$

Note that both $\mathbf{W}_s(k, N-k)$ and $\mathbf{M}_s(k, N-k)$ are $n \times n$ matrices; thus. we can say that the state $\mathbf{x}(k)$ is N measurement observable or controllable if the rank of the corresponding stochastic observability or controllability matrix is of rank n. The reason for saying this can be substantiated by the following corollary, which bounds $\mathbf{P}(k)$, the covariance matrix of the estimate.

COROLLARY 4.1. Let $\hat{\mathbf{x}}(k+1)$ be the estimate of the state of system $\mathbf{x}(k+1)$ at time $(k+1)$ T, given N measurements in the past. $\hat{\mathbf{x}}(k+1)$ is given by

$$\begin{aligned}\hat{\mathbf{x}}(k+1) = {} & \boldsymbol{\Phi}(k+1, k)\hat{\mathbf{x}}(k) + \mathbf{K}(k+1) \\ & [\mathbf{z}(k+1) - \mathbf{C}(k+1)\boldsymbol{\Phi}(k+1, k)\mathbf{x}(k)]\end{aligned} \tag{4.82}$$

Let $\mathbf{P}(k+1)$ the covariance matrix be given by

$$\mathbf{P}(k+1) = E[(\mathbf{x}(k+1) - \hat{\mathbf{x}}(k+1))(\mathbf{x}(k+1) - \hat{\mathbf{x}}(k+1))^T] \tag{4.83}$$

Then

$$\begin{aligned}[A\mathbf{M}_s(k, k-N) + \mathbf{W}_s^{-1}(k, k-N)]^{-1} \leqq \mathbf{P}(k+1) \leqq {} & \mathbf{M}_s^{-1}(k, k-N) \\ & + A\mathbf{W}_s(k, k-N)\end{aligned} \tag{4.84}$$

where A equals $(\alpha_4\,\alpha_6/\alpha_5\,\alpha_7)n^2$.

This corollary follows directly from the proof of the previous theorem and appears in Appendix C. It provides us with bounds on the covariance matrix in terms of the stochastic observability and controllability of the discrete-time system. Results similar to this appear in Sorenson [2,3] and is also inherently in the work of Deyst and Price. Similar results for the continuous system were first presented by Kalman [3] in 1963. A more recent discussion

for bounding the covariance for continuous-time nonlinear systems is presented by Gilman.

These results are quite powerful in determining the well-posed nature of experiments where the system is a constant random parameter, but the measurements are more complex. In that case,

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{0}; \qquad \mathbf{x}(0) = \mathbf{x}_0 \tag{4.85}$$

so that $\mathbf{Q}(t)$ is identically 0 for all cases. Thus,

$$\mathbf{W}_s(k, k - N) = \mathbf{0}; \qquad \forall k, N \tag{4.86}$$

which implies that

$$\mathbf{0} \leqq \mathbf{P}(k) \leqq \mathbf{M}_s^{-1}(k, k - N) \tag{4.87}$$

That is, the covariance is determined by the observability matrix. Furthermore,

$$\mathbf{M}_s(k, k - N) = \sum_{i=k-N}^{k} \mathbf{C}^T(i)\mathbf{R}^{-1}(i)\mathbf{C}(i) \tag{4.88}$$

This result is true as long as $\mathbf{M}_s(k, k - N)$ makes sense, that is, for positive definite $\mathbf{R}(i)$. ~~Carly~~, if $\mathbf{M}_s(k, k - N)$ is of maximal rank (i.e.,n), then as the noise covariance decreases, so does the upper bound on $\mathbf{P}(k)$. This is intuitively pleasing and also allows us to evaluate sampling techniques for different measurements, that is, $\mathbf{C}(i)$. The choice of $\mathbf{C}(i)$, which is an $m \times n$ matrix may be determined by m, the number of sensors, for example. Thus, using this bound, we can perform a trade-off analysis on the number of samples, the number of measurements, and the amount of noise we are willing to tolerate for a desired performance level.

Other approaches to this problem of stochastic controllability and observability have been discussed by Aoki (pp. 197–222) and by Jaswinski [2, pp. 231–255]. Their considerations parallel those contained here. The original presentation of these concepts is Kalman [3], and they were done by him for the continuous case.

There are issues that we have not covered because they essentially involve different approaches to the same problem. One approach would be to analyze the covariance equation and determine conditions for its stability. A second approach would be to consider the estimate equation and the state equations and obtain the propagation equation for the error $\tilde{\mathbf{x}}(k + 1)$. This is

$$\begin{aligned}\tilde{\mathbf{x}}(k+1) &= \mathbf{x}(k+1) - \hat{\mathbf{x}}(k+1) = \boldsymbol{\Phi}(k+1, k)\tilde{\mathbf{x}}(k) - \mathbf{K}(k+1) \\ &\quad \mathbf{C}(k+1)\boldsymbol{\Phi}(k+1, k)\tilde{\mathbf{x}}(k) + \mathbf{K}(k+1)\mathbf{w}(k+1)\end{aligned} \tag{4.89}$$

The homogeneous part of this equation is identical to the equation that we have already analyzed. Yet this equation differs in that it is driven by a ran-

dom variable (discrete random process). Issues concerning the stability of stochastic systems are not yet fully understood, although Kushner[5] presents a Lyapanov theory that is useful in many cases. A simplified discussion of these issues is also in Aoki (Chapter 8). A general review of the divergence issues arising with Kalman filters is presented in Fitzgerald.

In a similar fashion we note that (4.59) can be written as

$$\bar{\mathbf{P}}(k+1) = \mathbf{B}(k)\mathbf{P}(k)\mathbf{B}^T(k) + \mathbf{F}(k) \tag{4.90}$$

where $\mathbf{F}(k)$ is given by

$$\begin{aligned}\mathbf{F}(k) = {} & \mathbf{B}(k)\Delta\mathbf{m}\,\Delta\mathbf{f}^T\mathbf{D}^T(k+1) + \mathbf{D}(k+1)\,\Delta\mathbf{f}\,\Delta\mathbf{m}^T\mathbf{B}^T(k) \\ & + \mathbf{D}(k+1)\Delta\mathbf{f}\,\Delta\mathbf{f}^T\,\mathbf{D}^T(k+1) \\ & + \mathbf{K}(k+1)\mathbf{R}(k+1)\mathbf{K}^T(k+1) + \mathbf{Q}(k)\end{aligned} \tag{4.91}$$

Recall that $\mathbf{K}(k+1)$ depends upon $\mathbf{P}^*(k)$ but not $\bar{\mathbf{P}}(k)$ and all other terms are clearly independent of $\bar{\mathbf{P}}(k)$. Thus, this is the forcing term on the covariance equation we have obtained for the system error. In general, we do not know $\Delta\mathbf{f}$ or $\Delta\mathbf{m}$, but as we have said, we do assume knowledge of their boundedness. Thus, we assume that $\mathbf{F}(k)$ is bounded. The following theorem then gives the conditions under which $\bar{\mathbf{P}}(k+1)$ is bounded.

THEOREM 4.2

If $\bar{\mathbf{P}}(k+1)$ is given by

$$\bar{\mathbf{P}}(k+1) = \mathbf{B}(k)\bar{\mathbf{P}}(k)\mathbf{B}^T(k) + \mathbf{F}(k) \tag{4.92}$$

where $\mathbf{B}(k)$ is in (4.61) and $\mathbf{F}(k)$ is bounded in the sense that

$$\|\mathbf{F}(k)\| = \max_{i,j}|F_{ij}(k)| < \beta_1 < \infty \tag{4.93}$$

and if there exists an $N > 0$ such that

$$\|\bar{\mathbf{P}}(N)\| = \max_{i,j}|P_{ij}(N)| < \beta_2 < \infty \tag{4.94}$$

then there exists a positive finite nonzero constant β_3 such that

$$\|\bar{\mathbf{P}}(k)\| = \max_{i,j}|\bar{P}_{ij}(k)| < \beta_3 < \infty \tag{4.95}$$

for all $k > N$.

Proof. This theorem is an immediate result of Lemma 4.2, Theorem 4.1, and Theorem 4.2 of Chapter 2 in which we said that for a homogeneous system that was uniformly asymptotically stable then bounded inputs led to bounded outputs. ∧

This theorem then states that if the system is stochastically observable and controllable and if $\Delta\mathbf{f}$ and $\Delta\mathbf{m}$ are bounded, the variance $\bar{\mathbf{P}}(k)$ is bounded. This implies that without perfect knowledge we can, with increasing k, approach the desired value within some reasonable statistical bound, albeit an unknown bound. Price notes that the homogeneous part is u.a.s.i.l.

based upon conditions of the assumed model and not upon the actual model, which in fact we do not know. Furthermore, the necessity of having process noise is evident from the observability condition.

6.5 EXTENSIONS AND CONCLUSIONS

In this chapter we have considered three distinct, but interrelated, classes of problems. The first was for continuous-time continuous state system with continuous-time measurements; the second was for discrete-time measurements with continuous-time state; and the third was for both discrete-time system and measurement. We will now review those results and make comparisons.

The most important case is that of a linear system with linear Gaussian measurements. For this we obtained the Kalman-Bucy equations that gave the exact MMSE estimate. They are

$$\frac{d\hat{\mathbf{x}}}{dt} = \mathbf{A}(t)\hat{\mathbf{x}}(t) + \mathbf{P}(t)\mathbf{C}^T(t)\mathbf{R}^{-1}(t)[\mathbf{z}(t) - \mathbf{C}(t)\hat{\mathbf{x}}(t)] \tag{5.1}$$

$$\frac{d\mathbf{P}(t)}{dt} = \mathbf{A}(t)\mathbf{P}(t) + \mathbf{P}(t)\mathbf{A}^T(t) + \mathbf{Q}(t) - \mathbf{P}(t)\mathbf{C}^T(t)\mathbf{R}^{-1}(t)\mathbf{C}(t)\mathbf{R}(t) \tag{5.2}$$

with $\mathbf{x}(0)$ and $\mathbf{P}(0)$ being given. From Chapter 4 we had obtained the discrete-time version using the projection principle in Hilbert spaces. This was

$$\hat{\mathbf{x}}(k+1) = \boldsymbol{\Phi}(k+1,k)\hat{\mathbf{x}}(k) + \mathbf{K}(k+1)\mathbf{C}^T(k+1)\mathbf{R}^{-1}(k+1)$$
$$[\mathbf{z}(k+1) - \mathbf{C}(k+1)\boldsymbol{\Phi}(k+1,k)\hat{\mathbf{x}}(k)] \tag{5.3}$$

$$\mathbf{P}(k+1) = \boldsymbol{\Phi}(k+1,k)\,\mathbf{K}(k)\boldsymbol{\Phi}^T(k+1,k) + \mathbf{Q}(k) \tag{5.4}$$

$$\mathbf{K}(k+1) = [\mathbf{I} + \mathbf{P}(k+1)\mathbf{C}^T(k+1)\mathbf{R}^{-1}(k+1)\mathbf{C}(k+1)]^{-1}\mathbf{P}(k+1) \tag{5.5}$$

where $\mathbf{P}(0)$ and $\mathbf{x}(0)$ are assumed known. The computational aspects of this appear in Meditch [2, pp. 182–185]. The equivalence of these two forms is shown in Problem 4.18

For the case of nonlinear measurements, the continuous case (first-order approximation) yields

$$\frac{d\mathbf{x}^*}{dt} = \mathbf{f}(\mathbf{x}^*, t) + \mathbf{P}^*(t)\mathbf{C}(\mathbf{x}^*, t)\mathbf{R}^{-1}(t)[\mathbf{z}(t) - \mathbf{C}(\mathbf{x}^*, t)\mathbf{x}^*(t)] \tag{5.6}$$

$$\frac{d\mathbf{P}^*}{dt} = \mathbf{A}(\mathbf{x}^*, t)\mathbf{P}^*(t) + \mathbf{P}^*(t)\mathbf{A}^T(\mathbf{x}^*, t) + \mathbf{Q}(t) - \sum_{i=0}^{n} \mathbf{P}^*(t)\mathbf{G}_i(\mathbf{x}^*, t)\mathbf{P}^*(t) \tag{5.7}$$

where $\mathbf{x}^*(0)$ and $\mathbf{P}^*(0)$. The discrete-time analogue of this is obtained from the linearized MAP estimate. Namely,

$$\mathbf{x}^*(k+1) = \mathbf{f}(\mathbf{x}^*(k)) + \mathbf{K}(k+1)\mathbf{C}^T(k+1)\mathbf{R}^{-1}(k+1) [\mathbf{z}(k+1) - \mathbf{C}(k+1)\mathbf{f}(\mathbf{x}^*(k))] \tag{5.8}$$

$$\mathbf{P}(k+1) = \mathbf{A}(k)\mathbf{K}(k)\mathbf{A}(k) + \mathbf{Q}(k) \tag{5.9}$$

$$\mathbf{K}(k+1) = [\mathbf{I} + \mathbf{P}(k+1)\mathbf{C}^T(k+1)\mathbf{R}^{-1}(k+1)\mathbf{C}(k+1)]^{-1}\mathbf{P}(k+1) \tag{5.10}$$

where $\mathbf{A}(k+1)$ takes the place of $\Phi(k+1, k)$. The covariance equation in this filter does not contain measurements directly, so it represents the discrete version of the extended Kalman filter. This filter is defined with $\mathbf{G}_0(x^*, t) = \mathbf{R}^{-1}(t)$, all other $\mathbf{G}_i(x^*, t)$ being identically zero.

The third filter developed was for discrete measurements and continuous-time systems. It was given by

$$\hat{\mathbf{x}}(k) = \hat{\mathbf{x}}(k|k-1) + \mathbf{K}(k)[\mathbf{z}(k) - \mathbf{C}(k)\hat{\mathbf{x}}(k|k-1)] \tag{5.11}$$

$$\mathbf{K}(k) = \mathbf{M}(k)\mathbf{C}^T(k)[\mathbf{C}(k)\mathbf{M}(k)\mathbf{C}^T(k) + \mathbf{R}(k)]^{-1} \tag{5.12}$$

$$\mathbf{P}(k) = \mathbf{M}(k) - \mathbf{M}(k)\mathbf{C}^T(k)[\mathbf{C}(k)\mathbf{M}(k)\mathbf{C}^T(k) + \mathbf{R}(k)]^{-1}\mathbf{C}(k)\mathbf{M}(k) \tag{5.13}$$

$$\dot{\mathbf{M}}(t) = \mathbf{A}(\hat{x}(k-1))\mathbf{M}(t) + \mathbf{M}(t)\mathbf{A}^T(\hat{\mathbf{x}}(k-1)) + \mathbf{Q}(t) \tag{5.14}$$

for $(k-1)\,T, kT$, where $\mathbf{M}((k-1)T) = \mathbf{P}(k-1)$, $\hat{\mathbf{x}}(k|k-1)$ is the solution of

$$\dot{\mathbf{x}}^*(t) = \mathbf{f}(\mathbf{x}^*(t)) + \tfrac{1}{2}\sum_{i=1}^{n} \gamma_i \operatorname{tr}[\mathbf{F}_i[\hat{\mathbf{x}}(k-1)]\mathbf{M}(t)] \tag{5.15}$$

at $t = kT$, where $\mathbf{x}^*((k-1)\,T) = \hat{\mathbf{x}}(k-1)$.

These five different filters, the first two exact and the last three approximate represent only a few of the many possible forms available. In Section 6.1 alone we present several variations, the extended Kalman filter being the simplest. Schwartz and Stear compare the usefulness of these different types of filters for two simple scalar problems.

In a similar fashion we can write the estimation equations for the case of Poisson measurements. Unfortunately, as we mentioned, there is no exact standard model to judge linearized estimates against. Simulations have been done by Snyder [4–6] for biomedical purposes, by J. R. Clark for optical communication analysis, and by McGarty [3] for meteorological data-gathering.

The estimation problem always entails the solution of this covariance equation. This equation is called the *Riccati equation* and techniques are available for its solution. Discussion of the matrix Riccati equation are contained in Ogata and in the papers by Levin, Reid [1,2], Wonham [4], and Coles. Also, for the case where $t_0 \to -\infty$, the steady-state Riccati equation is to be solved. This solution for the linear time-invariant case is called *spectral factorization* and is discussed by Brockett.

There are other techniques for solving these equations. The use of quasi-moment functions has been discussed by Culver [1,2]; Fisher; Kuznetsov,

Srinivasan

Stratonovich, and Tikhonov [1,2]; and ~~Srivisrasan~~. We outline this method in Problem 6.18.

We can also extend the analysis to the case of having colored measurement noise. This has been done by Bryson and Johanson and in Van Trees [1]. It essentially requires augmenting the state variables.

Finally, it is worth mentioning the other methods that have been used to obtain the same results. The first among those as Kailath's innovations process approach. The linear problem is solved in Kailath [2], while the nonlinear problem is in Frost and Kailath. The innovation in the linear case is the received signal less the estimate of the measurement gain times the estimate of the state. This process is a white noise process, and since the results for white noise processes are trivial, the result for the estimate follows simply (see Problem 6.21). J. R. Clark and Frost have carried over an innovations type of analysis for Poisson measurements.

Another approach is the integral equation approach that uses Gaussian assumptions. This is discussed in Van Trees [1], where the Karhunen-Loeve expansion plays a vital role. However, this approach does not yield a recursive approach directly unless used with an invariant embedding technique (see Van Trees [2]).

This completes our discussion of estimator structures. The next chapter will discuss several extensions that broaden the use of the material presented.

6.6. PROBLEMS

6.1. Consider the linear vector Markov process $\mathbf{x}(t)$ generated by

$$d\mathbf{x}(t) = \mathbf{A}(t)\mathbf{x}(t)\,dt + d\mathbf{n}_g(t) + d\mathbf{n}_p(t)$$

Let $\mathbf{n}_g(t)$ be a vector Wiener process with

$$E[\mathbf{n}_g(t)\mathbf{n}_g^T(s)] = \mathbf{Q}\min(t, s)$$

and let $\mathbf{n}_p(t)$ be a generalized Poisson process with rate λ and amplitudes having probability density $p_\alpha(\alpha)$.

(a) The measurements are given by

$$d\mathbf{y}(t) = \mathbf{C}(t)\mathbf{x}(t)\,dt + d\mathbf{w}(t)$$

where $\mathbf{w}(t)$ is a Wiener process with

$$E[\mathbf{w}(t)\mathbf{w}^T(s)] = \mathbf{R}min(t, s)$$

Find the propagation equations for $\hat{\mathbf{x}}(t)$ and $\mathbf{P}(t)$. (They should be exact.)

(b) Now let the measurement be a simple vector Poisson process

$$d\mathbf{y}(t) = d\mathbf{N}(t)$$

with rate parameter

$$\lambda(\mathbf{x}, t) = \mathbf{C}(t)\mathbf{x}(t)$$

Find the linearized propagation equations for $\hat{\mathbf{x}}(t)$ and $\mathbf{P}(t)$.

6.2. The estimation equations in Theorem 1.2 for the case of simple Poisson measurements considered only retaining first-order terms. Obtain a set of estimate equations when the nonlinearities are expanded out to second order. Use the factoring of the covariance where necessary.

6.3. The more general model of a Markov process consists of having the state given by

$$d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}, t)\, dt + \boldsymbol{\sigma}(\mathbf{x}, t)\, d\mathbf{n}_g(t) + \boldsymbol{\beta}(\mathbf{x}, t)\, d\mathbf{n}_p(t)$$

where $\boldsymbol{\sigma}(\mathbf{x}, t)$ and $\boldsymbol{\beta}(\mathbf{x}, t)$ satisfy suitable regularity conditions. Likewise, the measurements in the Gaussian case can be generalized to

$$d\mathbf{y}(t) = \mathbf{h}(\mathbf{x}, t)\, dt + \boldsymbol{\gamma}(\mathbf{x}, t)\, d\mathbf{w}(t)$$

Here $\mathbf{n}_g(t)$ is an $(r \times 1)$-vector Wiener process, so that $\boldsymbol{\sigma}(\mathbf{x}, t)$ is an $n \times r$ matrix. $\mathbf{n}_p(t)$ is $s \times 1$ generalized Poisson process, and thus, $\boldsymbol{\beta}(\mathbf{x}, t)$ is an $n \times s$ matrix. Likewise, $\mathbf{w}(t)$ is a $(p \times 1)$-vector Wiener process and $\boldsymbol{\gamma}(\mathbf{x}, t)$ is $m \times p$.

(a) Obtain the Kushner-Stratonovich equation for this system.

(b) Obtain linearized estimation equations for $\mathbf{x}^*(t)$ and $\mathbf{P}^*(t)$.

(c) Comment on the simplifications.

6.4. Let x and y be scalars and let

$$dx = -ax\, dt + Q^{1/2}\, dn_g \qquad [x(t_0) = x_0]$$
$$dy = bx\, dt + R^{1/2}\, dw \qquad [P(t_0) = P_0]$$

where

$$E[n_g(t)n_g(s)] = E[w(t)w(s)] = \min(t, s)$$

(a) Write the estimation equations for $\hat{x}(t)$ and $P(t)$.

(b) Solve for $\hat{x}(t)$ in closed form.

(c) Solve for $P(t)$ in closed form.

(d) Assume that $t_0 \to -\infty$; find the steady-stated value of $P(t)$.

6.5. A random $n \times 1$ vector $\mathbf{x}$ is to be estimated by means of a measurement of the form

$$\mathbf{z}(t) = \mathbf{C}(t)\mathbf{x} + \mathbf{w}(t)$$

where $\mathbf{C}(t)$ is an $m \times 1$ vector and $\mathbf{w}(t)$ is a Gaussian white noise process with covariance $\mathbf{R}\delta(t - s)$. The vector $\mathbf{x}$ is known to be a positive random vector with zero probability of it being less than zero. A vector quantity

$$y_i = lnx_i$$

is defined so that y_i is found to be almost Gaussian.

(a) Obtain a nonlinear estimation formulation for this problem.

(b) If y_i is Gaussian, what are the statistics of the x_i?
(c) Under what conditions is this system observable?

6.6. A one-dimensional random process is given by

$$\dot{x}(t) = -kx(t) + u(t) \qquad [x(t_0) = x_0]$$
$$z(t) = Cx(t) + v(t)$$

(a) Write the covariance equation and the estimation equation.
(b) Assume that $t_0 \to -\infty$ and that t is large. Solve the covariance equation and the estimate equation in closed form.

6.7. An $n \times 1$ random parameter $\mathbf{x}$ is to be estimated by means of the following scheme. An $m \times 1$ ($m < n$) vector $\mathbf{z}(t)$ is given by

$$\mathbf{z}(t) = \mathbf{C}(t)\mathbf{x}(t) + \mathbf{w}(t); \qquad E[\mathbf{w}(t)\mathbf{w}^T(s)] = \mathbf{R}\delta(t - s)$$

(a) Write the estimation equation for this system.
(b) Let $n = 1 = m$ and $\mathbf{C}(t) = C_1$. Find $\mathbf{P}(t)$.
(c) If $m < n$, what conditions must $\mathbf{C}(t)$ satisfy for $\mathbf{P}(t)$ to be stable.

6.8. A random process $\mathbf{x}(t)$ is generated by

$$\dot{\mathbf{x}} = \begin{bmatrix} 0 & 1 \\ -3 & -2 \end{bmatrix}\mathbf{x} + \mathbf{u}$$

where $\mathbf{u}$ has covariance $\mathbf{I}\delta(t)$. The measurement equation is

$$\mathbf{z}(t) = \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}\mathbf{x} + \mathbf{w}$$

where $\mathbf{w}$ has covariance $\mathbf{I}\delta(t)$.

(a) Write the estimate and covariance equations for this filter.
(b) Find $P_{11}(t)$ and $P_{22}(t)$ analytically.

6.9. Phase modulation entails the transmission of some signal $a(t)$ by means of modulating the phase of a carrier signal. The message $a(t)$ is assumed to be a stationary Gauss-Markov process generated by

$$\dot{x}(t) = Ax(t) + u(t)$$

where

$$a(t) = C(t)x(t)$$

The received signal is

$$z(t) = C \sin(\omega_0 t + a(t)) + v(t)$$

(a) Obtain the linearized estimation equations.
(b) Assume $t_0 \to -\infty$. Realize the estimation of $a(t)$ in a feedback configuration. This is called the phase-lock loop.

6.10: The matrix Riccati equation is given by

$$\dot{\mathbf{P}}(t) = \mathbf{A}(t)\mathbf{P}(t) + \mathbf{P}(t)\mathbf{A}^T(t) - \mathbf{P}(t)\mathbf{C}^T(t)\mathbf{R}^{-1}(t)\mathbf{C}(t)\mathbf{P}(t) + \mathbf{Q}(t)$$

(a) Show that $\mathbf{P}(t)$ is equal to

$$\mathbf{P}(t) = \mathbf{X}(t)\mathbf{Y}^{-1}(t)$$

where $\mathbf{X}(t)$ satisfies

$$\dot{\mathbf{X}}(t) = \mathbf{Q}(t)\mathbf{Y}(t) + \mathbf{A}(t)\mathbf{X}(t)$$

and $\mathbf{Y}(t)$ satisfies

$$\dot{\mathbf{Y}}(t) = -\mathbf{A}^T(t)\mathbf{Y}(t) + \mathbf{C}^T(t)\mathbf{R}^{-1}(t)\mathbf{C}(t)\mathbf{X}(t)$$

(b) Let $\mathbf{Q}(t) = \mathbf{0}$ and let $\mathbf{A}$, $\mathbf{C}$, $\mathbf{R}$ be time-independent. Solve for $\mathbf{P}(t)$.

6.11. For the system

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\,dt + \boldsymbol{\sigma}(\mathbf{x}, t)\,d\mathbf{n}_g$$
$$d\mathbf{y} = \mathbf{h}(\mathbf{x}, t)\,dt + \boldsymbol{\beta}(\mathbf{x}, t)\,d\mathbf{w}$$

evaluate the linearized filter.

6.12. For the system

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\,dt + d\mathbf{n}_g$$
$$d\mathbf{y} = \mathbf{h}(\mathbf{x}, t)\,dt + d\mathbf{w}$$

evaluate the linear filter by assuming all moments greater than the second are zero (Bass, Norum, and Schwartz).

6.13. A model for an optical communication channel with log normal fading is given by letting the measurement be a Poisson counting process with rate

$$\lambda(x, t) = \beta s(t) \exp\{2[x(t) - P_0]\}$$

where $x(t)$ is to be estimated. Let

$$dx(t) = -kx(t)\,dt + \sqrt{2P_0 k}\,dw(t)$$

Evaluate $\hat{x}(t)$ and $P(t)$ for this system. Obtain $P(t)$ for large t.

6.14. Repeat Problem 6.13 but now let

$$\lambda(x, t) = \beta s(t) \exp\{2[x(t) - P_0]\} + \lambda_0$$

6.15. Consider the two systems given by

$$\Sigma_1 : \begin{cases} \dot{x}_1 = A_1 x_1 + B_1 u \\ y = C_1 x_1 \end{cases}$$
$$\Sigma_2 : \begin{cases} \dot{x}_2 = A_2 x_2 + B_2 u \\ y = C_2 x_2 \end{cases}$$

Let the measurement be

$$z(t) = y(t) + v(t)$$

where $u(t)$, $v(t)$ are zero mean Gaussian white noise processes with covariances $Q(t)\delta(t - s)$ and $R(t)\delta(t - s)$, respectively. Assume

$$x_1(t) = M(t)x_2(t)$$
$$x_2(t) = M^{-1}(t)x_1(t)$$

Show that the covariances $P_1(t)$ and $P_2(t)$ are given by

$$P_1(t) = M(t)P_2(t)M^T(t)$$

and the estimate $\hat{x}_1(t)$ is

$$\hat{x}_1(t) = M(t)\hat{x}_2(t)$$

6.16. A stochastic process is generated by the system

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\,dt + d\mathbf{n}$$

where $\mathbf{n}$ is a Wiener process of covariance $\mathbf{Q}$. The measurements are taken at discrete instants, so that

$$\mathbf{z}(t_k) = \mathbf{h}(\mathbf{x}(t_k), t_k) + \mathbf{v}(t_k)$$

where $\{t_k\}$ are the measurement times and $\mathbf{v}(t_k)$ is a Gaussian random noise sequence with zero mean and

$$E[\mathbf{v}(t_k)\mathbf{v}^T(t_j)] = \mathbf{R}(t_k)\delta_{kj}$$

(a) Let $p_{\mathbf{x}}(\mathbf{u}, t_k|\mathbf{z}(t_0) \cdots \mathbf{z}(t_j)) = p_{\mathbf{x}}(\mathbf{u}, t_k|Z_k)$. show that

$$p_{\mathbf{x}}(\mathbf{u}, t_k|Z_k) = \frac{H(t_k, \mathbf{u}, \mathbf{z}(t_k))p_{\mathbf{x}}(\mathbf{u}, t_k|Z_{k-1})}{\int H(t_k, \mathbf{v}, \mathbf{z}(t_k))p_{\mathbf{x}}(\mathbf{v}, t_k|Z_{k-1})\,d\mathbf{v}}$$

where

$$H(t_k, \mathbf{u}, \mathbf{z}(t_k)) = \exp\{-\tfrac{1}{2}[\mathbf{z}(t_k) - \mathbf{h}(\mathbf{u}, t_k)]^T[\mathbf{R}^{-1}(t_k)[\mathbf{z}(t_k) - \mathbf{h}(\mathbf{u}, t_k)]\}$$

(b) Using Ito's lemma show that

$$\frac{d\hat{\mathbf{x}}(t)}{dt} = E[\mathbf{f}(\mathbf{x}(t), t)] + [E[\mathbf{x}(t)\mathbf{h}^T(\mathbf{x}, t)] - \hat{\mathbf{x}}(t)E[\mathbf{h}(\mathbf{x}(t), t)]]\mathbf{R}^{-1}(t) \cdot [\mathbf{z}(t) - E[\mathbf{h}(\mathbf{x}, t)]]$$

and

$$\frac{dP_{ij}}{dt} = [E[\mathbf{x}(t)\mathbf{f}^T(\mathbf{x}, t)] - \hat{\mathbf{x}}(t)E[\mathbf{f}^T(\mathbf{x}, t)] + E[\mathbf{f}(\mathbf{x}, t)\mathbf{x}^T(t)] - E[\mathbf{f}(\hat{\mathbf{x}}, t)]\hat{\mathbf{x}}(t) - \mathbf{Q}(t)]_{ij} - [E[x_i(t)\mathbf{h}^T(\mathbf{x}, t)] - \hat{x}_i(t)E[\mathbf{h}^T(\mathbf{x}, t)]]\mathbf{R}^{-1}(t) \cdot [E[x_j(t)\mathbf{h}(\mathbf{x}, t)] - \hat{x}_j(t)E[\mathbf{h}(\mathbf{x}, t)]] + [E[x_i(t)x_j(t)\mathbf{h}^T(\mathbf{x}, t)] - E[x_i(t)x_j(t)]E[\mathbf{h}^T(\mathbf{x}, t)] - \hat{x}_i(t)E[x_j(t)\mathbf{h}^T(\mathbf{x}, t)] - \hat{x}_j(t)E[x_i(t)\mathbf{h}^T(\mathbf{x}, t)] + 2\hat{x}_i(t)\hat{x}_j(t)E[\mathbf{h}^T(\mathbf{x}, t)]]\mathbf{R}^{-1}(t)[\mathbf{z}(t) - E[\mathbf{h}(\mathbf{x}, t)]]$$

(c) Obtain a set of linearized equations for this problem by means of Taylor-series expansion. Compare this to the case with continuous measurements.

6.17 A laser system transmits one of two signals through a turbulent channel. For a given signal m_i the number of counts given the field incident E_i on a detector is a Poisson random variable with

$$P[N(t) = k | m_i, E_i] = \frac{[\lambda |E_i|^2]^k}{K!} \exp[-\lambda |E_i|^2]$$

where

$$|E_i|^2 = [E_{x_i}^2 + E_{y_i}^2]$$

where E_{x_i} and E_{y_i} are the x and y components of the electric field corresponding to meassage m_i. Since the transmission medium is turbulent, E_{x_i} and E_{y_i} are found to be zero mean independent Gaussian random variables, identically distributed with variance σ_i^2.

(a) Find the probability density function of $|E_i|^2$.

(b) Find the probability that $N(t) = k$, given that message m_i was sent.

(c) Evaluate the a posteriori probability density of $|E_i|^2$ given that $N(t) = k$. This is

$$p_{|E_i|^2}(\beta | N(t) = k, m_i)$$

(d) Find the maximum a posteriori estimate of $|E_i|^2$ given m_i, $N(t) = k$.

6.18. Consider any probability density function $p_x(u, t | \mathscr{B}_t)$. Let $m(t)$ and $\sigma^2(t)$ be the conditional mean and variance of $x(t)$. and let $\mathscr{B}_t$ be any arbitrary σ-field generated by measurements on $x(t)$. If $x(t)$ were Gaussian, then

$$p_{x,g}(u, t | \mathscr{B}_t) = \frac{1}{\sqrt{2\pi\sigma^2(t)}} \exp\left(-\frac{[u - m(t)]^2}{2\sigma^2(t)}\right)$$

(a) Show that $p_x(u, t | \mathscr{B}_t)$ can be written as

$$p_x(u, t | \mathscr{B}_t) = p_{x,g}(u, t | \mathscr{B}_t) \sum_{k=0}^{\infty} \frac{1}{k!} \frac{b_k(t)}{\sigma^k(t)} H_k\left[\frac{(u - m(t))}{\sigma(t)}\right]$$

where $H_k(u)$ is a Hermite polynomial. (This is called the Edgeworth expansion.)

(b) Show that quasi-moment functions b_k are given by

$$b_k(t) = \sigma^k(t) \int_{-\infty}^{\infty} p_x(u, t | \mathscr{B}_t) H_k\frac{(u - m(t))}{\sigma(t)} dy$$

$$= \sigma^k(t) E\left[H_k\left(\frac{x(t) - m(t)}{\sigma(t)}\right) | \mathscr{B}_t\right]$$

(c) Using the Fokker-Planck equation for the scalar system

$$dx = f(x, t)\, dt + dw$$
$$dy = h(x, t)\, dt + du$$

where $E[w^2(t)] = t$, $E[u^2(t)] = t$, and $x(0) = 0$, $\mathscr{B}_t = O_{t,t_0}$, find the propagation equation for $b_k(t)$. *Hint.* Use the orthogonality properties of Hermite polynomials.

(d) Let $M_g(u, t|\mathscr{B}_t)$ be the characteristic function of $p_{x,g}(u, t|\mathscr{B}_t)$ and $M(u, t|\mathscr{B}_t)$ the characteristic function of $p_x(u, t|\mathscr{B}_t)$. Define

$$k(u, t) = \frac{M(u, t|\mathscr{B}_t)}{M_g(u, t|\mathscr{B}_t)}$$

Let

$$k(u, t) = \sum_{n=0}^{\infty} K_n(t)u^n \frac{(j)^n}{n!}$$

where

$$K_n(t) = \frac{1}{(j)^n} \frac{\partial^n K(u, t)}{\partial u^n}$$

Show that

$$K_n(t) = \frac{b_k(t)}{\sigma^k(t)}$$

(e) Using the above result, let

$$C_k(t) = E[(x(t) - m(t))^k|\mathscr{B}_t]$$

Find $C_k(t)$ in terms of $K_k(t)$ for $k = 1,2,3,4$.

Relate this to the result of part (c).

6.19. Repeat Problem 6.18 for the case of a vector state and a vector measurement. (*Hint.* See Fisher.)

6.20. The discrete-time system was given by

$$\mathbf{x}(k + 1) = \boldsymbol{\Phi}(k + 1, k)\mathbf{x}(k) + \mathbf{n}(k)$$
$$\mathbf{z}(k + 1) = \mathbf{C}(k + 1)\mathbf{x}(k + 1) + \mathbf{w}(k + 1)$$

where $\mathbf{n}(k)$ has covariance $\mathbf{Q}(k)$ and $\mathbf{w}(k)$ has covariance $\mathbf{R}(k)$. The estimate equations are

$$\hat{\mathbf{x}}(k + 1) = \boldsymbol{\Phi}(k + 1, k)\hat{\mathbf{x}}(k) + \mathbf{K}(k + 1)[\mathbf{z}(k + 1) - \mathbf{C}(k + 1)\boldsymbol{\Phi}(k + 1, k)\hat{\mathbf{x}}(k)]$$
$$\mathbf{K}(k + 1) = \mathbf{M}(k + 1)\mathbf{C}^T(k + 1)[\mathbf{C}(k + 1)\mathbf{M}(k + 1)\mathbf{C}(k + 1) + \mathbf{R}(k + 1)]$$
$$\mathbf{M}(k+1) = \boldsymbol{\Phi}(k + 1, k)\mathbf{P}(k)\boldsymbol{\Phi}^T(k + 1, k) + \mathbf{Q}(k)$$
$$\mathbf{P}(k + 1) = [\mathbf{I} - \mathbf{K}(k + 1)\mathbf{C}(k + 1)]\mathbf{M}(k + 1)$$

Show that if Δt is the sample time in the above equations that as $\Delta t \to 0$, the linear continuous-time filter can be obtained.

6.21. (Kailath [2]) A continuous-time linear time-varying system

$$\dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{u}(t)$$
$$\mathbf{z}(t) = \mathbf{C}(t)\mathbf{x}(t) + \mathbf{v}(t)$$

has

$$E[\mathbf{u}(t)\mathbf{u}^T(s)] = \mathbf{Q}(t)\delta(t - s)$$
$$E[\mathbf{v}(t)\mathbf{v}^T(s)] = \mathbf{R}(t)\delta(t - s)$$

(a) Let $\mathbf{v}(t) = \mathbf{z}(t) - \mathbf{C}(t)\hat{\mathbf{x}}(t)$. Show that $\mathbf{v}(t)$ is a white noise process, that is,

$$E[\mathbf{v}(t)\mathbf{v}(s)] = \mathbf{R}(t)\delta(t - s)$$

This is called the innovation process.

(b) Assume that $\mathbf{v}(t)$ has been given. Since the system is linear, we know that

$$\hat{\mathbf{x}}(t) = \int_0^t \mathbf{h}(t, s)\mathbf{z}(s)\, ds$$

so that $\mathbf{v}(t)$ is a linear functional of the measurement. Show that we can write

$$\hat{\mathbf{x}}(t) = \int_0^t \mathbf{g}(t, s)\mathbf{v}(s)\, ds \qquad (*)$$

and use the *orthogonality principle* to show that

$$\mathbf{g}(t, s) = E[\mathbf{x}(t)\mathbf{v}^T(s)]\mathbf{R}^{-1}(s)$$

(c) Let

$$\mathbf{K}(t) = E[\mathbf{x}(t)\mathbf{v}^T(t)]\mathbf{R}^{-1}(t)$$

and show by using $(*)$ that

$$\dot{\hat{\mathbf{x}}}(t) = \mathbf{A}(t)\hat{\mathbf{x}}(t) + \mathbf{K}(t)\mathbf{v}(t)$$

(d) Show that

(i) $\mathbf{K}(t) = \mathbf{P}(t)\mathbf{C}^T(t)\mathbf{R}^{-1}(t)$

and using the results of Problem 6.20 show

(ii) $\dfrac{d\mathbf{P}(t)}{dt} = \mathbf{A}(t)\mathbf{P}(t) + \mathbf{P}(t)\mathbf{A}^T(t) - \mathbf{K}(t)\mathbf{R}(t)\mathbf{K}^T(t) + \mathbf{Q}(t)$

6.22. (Kailath and Geesey) The K-B filter assumes that a state model is available for $\mathbf{x}(t)$. However, in many cases only $\mathbf{K}_x(t, s)$, the covariance of $\mathbf{x}(t)$, is available. That is,

$$E[(\mathbf{x}(t) - \bar{\mathbf{x}}(t))(\mathbf{x}(s) - \bar{\mathbf{x}}(s))^T] = \mathbf{K}_x(t, s)$$

Furthermore, assume that $\mathbf{K}_x(t, s)$ can be written as

$$\mathbf{K}_x(t, s) = \begin{cases} \sum_{i=1}^{n} \alpha_i(t)\beta_i(s) & (t \geqq s) \\ \sum_{i=1}^{n} \alpha_i(s)\beta_i(t) & (t < s) \end{cases}$$
$$= \alpha(t \wedge s)\beta(t \wedge s)$$

where α_i and β_i are finite valued matrices and

$$t \vee s = \max(t, s)$$
$$t \wedge s = \min(t, s)$$

(a) Let $\Phi(t, s)$ be defined by

$$\frac{d\Phi(t, s)}{dt} = \mathbf{A}(t)\Phi(t, s) \qquad [\Phi(t, t) = \mathbf{I}]$$

Let

$$\mathbf{M}(t) = \alpha(t)\Phi(t_0, t) \quad ; \quad \mathbf{N}(t) = \Phi(t, t_0)\beta(t)$$

Show that

$$\mathbf{K}_x(t, s) = \mathbf{M}(t \vee s)\Phi(t \vee s, t \wedge s)\mathbf{N}(t \wedge s)$$

(b) Let $\mathbf{z}(t)$ be a process with covariance

$$\mathbf{K}_z(t, s) = \mathbf{C}(t)\mathbf{K}_x(t, s)\mathbf{C}^T(s) + \mathbf{I}$$

Show that $\mathbf{z}(t)$ can be written in the form

$$\dot{\varphi}(t) = \mathbf{F}(t)\varphi(t) + \bar{\mathbf{K}}(t)\mathbf{v}(t) \qquad [\varphi(t_0) = \mathbf{0}]$$
$$\mathbf{z}(t) = \mathbf{M}(t)\varphi(t) + \mathbf{v}(t)$$

where $\mathbf{v}(t)$ is white Gaussian and $\mathbf{F}(t)$ arbitrary and

$$\bar{\mathbf{K}}(t) = \mathbf{N}(t) - \Sigma(t)\mathbf{M}^T(t)$$

where

$$\Sigma(t) = E[\varphi(t)\varphi^T(t)]$$

Show that $\Sigma(t)$ satisfies

$$\dot{\Sigma}(t) = \mathbf{F}\Sigma + \Sigma\mathbf{F}^T + [\mathbf{N} - \mathbf{P}\mathbf{M}^T][\mathbf{N} - \mathbf{P}\mathbf{M}^T]$$

(c) Show that $\mathbf{v}(t)$, the "innovations," is given by

$$\mathbf{v}(t) = \mathbf{z}(t) - \mathbf{M}(t)\varphi(t)$$
$$\dot{\varphi}(t) = \mathbf{F}(t)\varphi(t) + \bar{\mathbf{K}}(t)[\mathbf{z}(t) - \mathbf{M}(t)\varphi(t)] \qquad [\varphi(t_0) = \mathbf{0}]$$

(d) Since $\mathbf{F}(t)$ is arbitrary, use $\mathbf{F}(t) = \mathbf{0}$ and the results of c and d to obtain the estimation equations for $\mathbf{x}(t)$ that has covariance

$$\mathbf{K}_x(t, s) = a_1 \exp(-\lambda_1|t - s|) + a_2 \exp(-\lambda_2|t - s|)$$

(e) Develop an equivalent state variable model for (d).

6.23. A scalar Markov process $x(t)$ is given by

$$\dot{x}(t) = -\alpha(t) + \dot{w}(t)$$

where $a > 0$ and $\dot{w}(t)$ is a white noise (Gaussian) process with spectral height Q. Measurements $z(t)$ are also available for this process from t_0 to time $t > t_1$. Let

$$\Sigma(t) = E[(x(t) - \hat{x}(t))^2]$$

and

$$z(t) = x(t) + \dot{u}(t) \qquad (E[\dot{u}(t)\dot{u}(s)] = R\delta(t - s))$$

and

$$P(t) = E[(x(t) - \hat{x}(t))^2 | O_{t_0,t}]$$

a. Find $\Sigma(t)$as a function of t and $\Sigma(t_0)$. Obtain $\Sigma(t)$ for $t \gg t_0$.

b. Find $P(t)$ as a function of t and $\Sigma(t_0)$. [Note: $P(t_0) = \Sigma(t_0)$.] Obtain $P(t)$ for $t \gg t_0$.

c. Sketch $\Sigma(t)$ and $P(t)$ versus t.

d. Evaluate $P(t)$, $t \gg t_0$ as $R \to 0$.

6.24. The discrete estimation equations are given by

$$\hat{\mathbf{x}}(k+1) = \Phi(k+1, k)\hat{\mathbf{x}}(k) + \mathbf{K}(k+1)[\mathbf{z}(k+1) - \mathbf{C}(k+1)\Phi(k+1, k)\hat{\mathbf{x}}(k)]$$

where

$$\mathbf{K}(k+1) = \mathbf{P}(k+1)\mathbf{C}^T(k+1)\mathbf{R}^{-1}(k+1)$$

Assume that $\mathbf{K}^*(k+1)$ is the optimal value of $\mathbf{K}(k+1)$ but that we use $\mathbf{K}(k+1)$ that is

$$\mathbf{K}(k+1) = \mathbf{K}^*(k+1) + \delta\mathbf{K}^*(k+1)$$

where $\delta\mathbf{K}^*(k+1)$ is a zero mean Gaussian matrix where

$$E[\delta\mathbf{K}^*_{ij}(k+1)\delta\mathbf{K}_{lm}(k+1)] = \Sigma_{ij,lm}(k+1)$$

(a) Obtain an equation for $\tilde{\mathbf{P}}(k+1)$ where

$$\tilde{\mathbf{P}}(k+1) = E[(\mathbf{x}(k+1) - \hat{\mathbf{x}}'(k+1))(\mathbf{x}(k+1) - \hat{\mathbf{x}}'(k+1))^T]$$

where $\hat{\mathbf{x}}'(k+1)$ is $\hat{\mathbf{x}}(k+1)$ using $\mathbf{K}(k+1)$ and not $\mathbf{K}^*(k+1)$.

(b) Show that this estimate equation is stable.

6.25. Assume a system is given by

$$\mathbf{x}(k+1) = \mathbf{x}(k) \qquad [\mathbf{x}(0) = \mathbf{x}_0]$$
$$\mathbf{z}(k) = \mathbf{C}(k)\mathbf{x}(k) + \mathbf{w}(k)$$

where $\mathbf{w}(k)$ has covariance $R(k)$. Now in the real model there exists a random bias $\mathbf{b}(k)$ in the measurement equation such that the actual measurement is

$$\mathbf{z}(k) = \mathbf{C}(k)\mathbf{x}(k) + \mathbf{b}(k) + \mathbf{w}(k)$$

Let

$$\mathbf{P}(k+1) = E[(\mathbf{x}(k+1) - \hat{\mathbf{x}}(k+1))(\mathbf{x}(k+1) - \hat{\mathbf{x}}(k+1))^T]$$
$$\mathbf{Q}(k+1) = E[(\hat{\mathbf{x}}^*(k+1) - \hat{\mathbf{x}}(k+1))(\hat{\mathbf{x}}^*(k+1) - \hat{\mathbf{x}}(k+1))^T]$$

where $\hat{\mathbf{x}}^*(k+1)$ is the estimate of $\mathbf{x}(k+1)$, assuming the unbiased measurement model.

(a) Let $\mathbf{P}'(k+1) = E[(\hat{\mathbf{x}}^*(k+1) - \mathbf{x}(k+1))(\hat{\mathbf{x}}(k+1) - \mathbf{x}(k+1))^T]$. Show that

$$\mathbf{P}'(k+1) = \mathbf{P}(k+1) + \mathbf{Q}(k+1)$$

and that $\mathbf{Q}(k+1)$ is given by

$$\begin{aligned}\mathbf{Q}(k+1) = \mathbf{Q}(k) &+ \mathbf{P}(k+1)\mathbf{C}^T(k+1)\mathbf{R}^{-1}(k+1)\mathbf{C}(k+1)\mathbf{Q}(k)\\ &+ \mathbf{Q}(k)\mathbf{C}^T(k+1)\mathbf{R}^{-1}(k+1)\mathbf{C}^T(k+1)\mathbf{P}(k+1)\\ &- \mathbf{P}(k+1)\mathbf{C}^T(k+1)\mathbf{R}^{-1}(k+1)\mathbf{Q}(k+1)\mathbf{R}^{-1}(k+1)\\ &\mathbf{C}(k+1)\mathbf{P}(k+1) + \mathbf{P}(k+1)\mathbf{C}^T(k+1)\mathbf{R}^{-1}(k+1)\\ &\mathbf{S}(k+1)\mathbf{R}^{-1}(k+1)\mathbf{C}(k+1)\mathbf{P}(k+1)\end{aligned}$$

where

$$\mathbf{S}(k+1) = E[\mathbf{b}(k+1)\mathbf{b}^T(k+1)]$$

(b) Evalute the continuous-time version of this estimation problem to show that

$$\begin{aligned}\frac{d\mathbf{Q}(t)}{dt} &= \mathbf{P}(t)\mathbf{C}^T(t)\mathbf{R}^{-1}(t)\mathbf{C}(t)\mathbf{Q}(t)\\ &\quad + \mathbf{Q}(t)\mathbf{C}^T(t)\mathbf{R}^{-1}(t)\mathbf{C}(t)\mathbf{P}(t)\\ &\quad + \mathbf{P}(t)\mathbf{C}^T(t)\mathbf{R}^{-1}(t)\mathbf{S}(t)\mathbf{R}^{-1}(t)\mathbf{C}(t)\mathbf{P}(t)\\ \frac{d\mathbf{P}(t)}{dt} &= -\mathbf{P}(t)\mathbf{C}^T(t)\mathbf{R}^{-1}(t)\mathbf{C}(t)\mathbf{P}(t)\end{aligned}$$

and that

$$\mathbf{P}'(t) = \mathbf{Q}(t) + \mathbf{P}(t)$$

where

$$\mathbf{P}'(t) = E[(\hat{\mathbf{x}}^*(t) - \mathbf{x}(t))(\hat{\mathbf{x}}^*(t) - \mathbf{x}(t))^T]$$

6.26. Let $\mathbf{z}(k)$ represent a measurable output given by

$$\mathbf{z}(k) = \mathbf{C}(k)\mathbf{x} + \mathbf{w}(k)$$

where $\mathbf{x}$ is a random vector to be estimated and $\mathbf{w}(k)$ is a zero mean Gaussian sequence with

$$E[\mathbf{z}(k)\mathbf{z}^T(j)] = \mathbf{R}(k)\delta_{kj}$$

(a) Show that the maximum likelihood estimate is after n measurement

$$\hat{\mathbf{x}}(n) = \left[\sum_{i=1}^{n} \mathbf{C}^T(i)\mathbf{R}^{-1}(i)\mathbf{C}(i)\right]^{-1} \sum_{i=1}^{n} \mathbf{C}^T(i)\mathbf{R}^{-1}(i)\mathbf{z}(i)$$

(b) Suppose $\mathbf{R}(i)$ is not exactly known. Let

$$\mathbf{R}'(i) = \mathbf{R}(i) + \delta\mathbf{R}(i)$$

where $\delta\mathbf{R}(i)$ is a small perturbation about the real value $\mathbf{R}(i)$. Assume $\delta\mathbf{R}(i)$has zero mean entries and are statistically independent for each i. Show that

$$\begin{aligned}\hat{\mathbf{x}}(n) + \delta\hat{\mathbf{x}}(n) &= \left[\sum_{i=1}^{n} \mathbf{C}^T(i)\mathbf{R}^{-1}(i)\mathbf{C}(i) - \sum_{i=1}^{n} \mathbf{C}^T(i)\mathbf{R}^{-1}(i)\delta\mathbf{R}(i)\mathbf{R}^{-1}(i)\mathbf{C}(i)\right]^{-1} \\ &\cdot \left[\sum_{i=1}^{n} \mathbf{C}^T(i)\mathbf{R}^{-1}(i)\mathbf{z}(i) - \sum_{i=1}^{n} \mathbf{C}^T(i)\mathbf{R}^{-1}(i)\delta\mathbf{R}(i)\mathbf{R}^{-1}(i)\mathbf{z}(i)\right]\end{aligned}$$

Hint. Show

$$[\mathbf{R}'(i)]^{-1} \approx \mathbf{R}^{-1}(i) - \mathbf{R}^{-1}(i)\delta\mathbf{R}(i)\mathbf{R}^{-1}(i)$$

(c) Show that $\delta\hat{\mathbf{x}}$ can be written as

$$\delta\hat{\mathbf{x}} = -\mathbf{P}_n\left[\sum_{i=1}^{n} \mathbf{C}^T(i)\mathbf{R}^{-1}(i)\delta\mathbf{R}(i)\mathbf{R}^{-1}(i)[\mathbf{z}(i) - \mathbf{C}(i)\hat{\mathbf{x}}(n)]\right]$$

where

$$\mathbf{P}_n = E\left[(\mathbf{x} - \hat{\mathbf{x}}(n))(\mathbf{x} - \hat{\mathbf{x}}(n))^T\right]$$

(d) Show that $\delta\hat{\mathbf{x}}$ is zero mean and

$$E[\delta\hat{\mathbf{x}}(n)\delta\hat{\mathbf{x}}^T(n)] = \sum_{i=1}^{n} \mathbf{P}_n\mathbf{C}^T(i)\mathbf{R}^{-1}(i)\mathbf{D}(i)\mathbf{R}^{-1}(i)\mathbf{C}(i)\mathbf{P}_n$$

where the rsth element of $\mathbf{D}(i)$ is given by

$$\begin{aligned}\mathbf{D}_{rs}(i) &= \mathrm{tr}\left[[\mathbf{R}^{-1}(i)(\mathbf{R}(i) - \mathbf{C}^T(i)\mathbf{P}_n\mathbf{C}(i))\mathbf{R}^{-1}(i)][\mathbf{P}(i)]_{rs}\right] \\ &= E\left[[\delta\mathbf{R}(i)\mathbf{R}^{-1}(i)[\mathbf{R}(i) - \mathbf{C}(i)\mathbf{P}_n\mathbf{C}^T(i)\delta\mathbf{R}^{-1}(i)]\mathbf{R}(i)]_{rs}\right]\end{aligned}$$

where $[\mathbf{P}(i)]_{rs}$ are the covariance components of $\delta\mathbf{R}(i)$.

6.27. A set of discrete measurements $\mathbf{z}(k-N) \cdots \mathbf{z}(k)$ are made and are of the form

$$\mathbf{z}(i) = \mathbf{C}(i)\mathbf{x}(i) + \mathbf{w}(i)$$

where $\mathbf{w}(i)$ are zero mean independent Gaussian random variables with covariance

$$E[\mathbf{w}(i)\mathbf{w}^T(i)] = \mathbf{R}(i)$$

The parameter $\mathbf{x}(k)$ is to be estimated using these measurements and $\mathbf{x}(k)$ is given by

$$\mathbf{x}(k) = \boldsymbol{\Phi}(k, k-1)\mathbf{x}(k-1)$$

with $\mathbf{x}(k-N)$ being an unknown random quantity.

(a) Evaluate the joint probability density of $\mathbf{z}(k) \cdots \mathbf{z}(k-N)$.

(b) The maximum likelihood estimate is that value of $\mathbf{x}(k)$ that maximizes the density

$$p(\mathbf{z}(k-\mathrm{N}), \cdots, \mathbf{z}(k)|\mathbf{x}(k))$$

Show that $\hat{\mathbf{x}}(k)$, the maximum likelihood estimate of $\mathbf{x}(k)$ for this model, is given by

$$\hat{\mathbf{x}}(k) = \bar{\mathbf{P}}(k) \sum_{i=k-N}^{k} \boldsymbol{\Phi}^T(i, k)\mathbf{C}^T(i)\mathbf{R}^{-1}(i)\mathbf{z}(i)$$

where

$$\bar{\mathbf{P}}(k) = \left[\sum_{i=k-N}^{k} \boldsymbol{\Phi}^T(i, k)\mathbf{C}^T(i)\mathbf{R}^{-1}(i)\mathbf{C}(i)\boldsymbol{\Phi}(i, k) \right]^{-1}$$

6.28. Let $x(t)$ be a Gaussian random process on $[0, T]$ with zero mean and covariance

$$E[x(t)x(u)] = K_x(t, u)$$

Assume that $x(t)$ can be written as a series

$$x(t) = \sum_{i=1}^{\infty} x_i \varphi_i(t)$$

where $\{\varphi_i(t)\}$ is a set of orthornormal functions on $[0, T]$, that is,

$$\int_0^T \varphi_i(t)\varphi_j(t)\, dt = \delta^{ij}$$

(a) Find x_i in terms of $x(t)$, $\varphi_i(t)$ and T.

(b) Let λ_i be

$$\lambda_i = E[x_i^2]$$

Show it is sufficient for $\{\varphi_i(t)\}$ to be the solution to the integral equation

$$\lambda_i \varphi_i(t) = \int_0^T K_x(t, u)\varphi_i(u)\, du$$

for x_i and x_j to be independent random variables.

(c) Show that with this representation

$$\lim_{N \to \infty} E\left\{ \left[x(t) - \sum_{i=1}^{N} x_i \varphi_i(t) \right]^2 \right\} = 0$$

Use the Chebychev inequality to interpret your results.

(d) Let $n_w(t)$ be a zero mean white Gaussian noise power spectrum $N_0/2$. Find the $\{n_i\}$ such that

$$n_w(t) = \sum_{i=1}^{\infty} n_i \varphi_i(t)$$

(e) Let a be a Gaussian random variable with mean $\bar{a}$ and variance σ_a^2. Let $s(t, a)$ be defined on $t \in [0, T]$ and be a function of a. The signal $r(t)$

$$r(t) = s(t, a) + n_w(t)$$

is received on $[0, T]$. Let

$$r_k(t) = \sum_{i=1}^{K} r_i \varphi_i(t)$$

where r_i are the projections of $r(t)$ on $\varphi_i(t)$. Show that the a that maximizes the a posteriori density function

$$p_{a|r_1,\cdots,r_K}(A|r_1, \cdots, r_K)$$

is given by the solution to

$$\hat{a} = a + \frac{2\sigma_a^2}{N_0} \sum_{i=1}^{K} (r_i - s_i(\alpha)) \left[\frac{\partial s_i(\alpha)}{\partial \alpha} \right]_{\alpha=\hat{a}}$$

(f) Show that as $K \to \infty$, the estimate a that maximizes the a posteriori density is

$$\hat{a} = a + \frac{2\sigma_a^2}{N_0} \int_0^T [(r(t) - s(t, \alpha)) \frac{\partial}{\partial \alpha} s(t, \alpha)]_{\alpha=\hat{a}} \, dt$$

6.29. Let a signal source generate M message $\{m_k\}, k = 1, \cdots, M$. For each message m_k there corresponds a unique signal vector $\mathbf{s}_k$. The messages are sent over a random Gaussian channel with additive Gaussian noise so that the received vector $\mathbf{r}$ is

$$\mathbf{r} = \mathbf{z}_k + \mathbf{n}$$

where $\mathbf{n}$ is a zero mean $n \times 1$ Gaussian vector with covariance $\mathbf{K}_n$

$$\mathbf{K}_n = E[\mathbf{n}\ \mathbf{n}^T]$$

and $\mathbf{z}_k$ is

$$\mathbf{z}_k = \mathbf{A}\mathbf{s}_k$$

where $\mathbf{A}$ is an $n \times n$ matrix composed of zero mean Gaussian random variables all independent of the noise vector $\mathbf{n}$.

Clearly $\mathbf{z}_k$ and $\mathbf{r}$ are both zero mean Gaussian random variables. Define

$$\mathbf{K}_{z_k} \triangleq E[\mathbf{z}_k \mathbf{z}_k^T]$$

Then if m_k was sent, the received signal covariance is

$$\mathbf{K}_r = \mathrm{E}[\mathbf{r}\,\mathbf{r}^T] = \mathbf{K}_{z_k} + \mathbf{K}_n$$

(a) Assume that message k was sent. Show that the minimum mean square estimate of $\mathbf{z}_k$, given $\mathbf{r}$, is

$$\mathbf{z}_k = \mathbf{H}_k \mathbf{r}$$

where

$$\mathbf{H}_k = \mathbf{K}_{z_k}[\mathbf{K}_{z_k} + \mathbf{K}_n]^{-1}$$

(b) Show that if m_k was sent that

$$[\mathbf{K}_r]^{-1} = \mathbf{K}_n^{-1} - \mathbf{K}_n^{-1}\mathbf{K}_{z_k}$$

Hint. $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$ if $\mathbf{A}^{-1}$ and $\mathbf{B}^{-1}$ exist. Assume they do.

(c) What is the decision rule for this signal scheme so that the minimum error probability is obtained for equally likely signals m_k? Show that the decision rule reduces too.
Choose m_k if

$$\mathbf{r}^T\mathbf{K}_n^{-1}\hat{\mathbf{z}}_k > \mathbf{r}^T\mathbf{K}_n^{-1}\hat{\mathbf{z}}_i \qquad (\forall i \neq k)$$

(d) Sketch an implementation of the detection structure for $\mathbf{K}_n = \mathbf{I}$, the identity matrix, and comment on its interpretation.

CHAPTER 7
CONCLUSIONS

Throughout the previous chapters we have tried to bring out the relationship between theory and applications despite the fact that at times one had to be discussed without the benefit of the other. In this final chapter we wish to extend our comments concerning theory and applications into areas that for want of space could not be discussed at length in the body of the text. In the first section we discuss several applications of the theory. These areas of applications are in a state of varying degrees of development. Some like the aerospace area have been rather elegantly developed, leading to extraordinary results in the ability to locate, direct, and project various objects. Other areas are new for the theory and thus few results are available. Such areas are those such as biomedical systems or the earth sciences, where a significant change of vocabulary is necessary in order to even understand the theory. However, rapid advances are being made even in these areas.

The second section discusses several areas of theoretical extensions. These are areas that in general require extensions to the present theory. What we have done is to outline briefly these areas, indicating what extensions have to be made and referencing to the work that has been done in these areas. Like any list, it is not all-inclusive but acts merely as an indication to what can be achieved with the theory.

7.1 APPLICATIONS

In the previous chapters we discussed at length the problem of nonlinear estimation. In this section we intend to extend that discussion to seven areas in which the methods developed herein have been used. Each of the areas represent a problem in which the quantity sought after, to be estimated, can be generated by a Markov process and the measurements then depend upon this process.

7.1.1 Aerospace Systems

The initial use of recursive filtering was in the aerospace sciences. Battin

attributes it initially to Gauss, who employed recursive least-squares methods to estimate planet trajectories. The first area in aerospace use is in space navigation. The prime reference in this area is the work by Battin. Bucy and Joseph also discuss the use of nonlinear filtering to navigation. Their discussion is more consistent with the nonlinear estimation approach rather than the linearized extended Kalman filter approach of Battin. Mowry also discusses applications to space navigation, giving examples relating to constant velocity tracking, reentry, and angle tracking. Ohap and Stubbered using a different ad hoc technique also make applications to the navigation problem.

The analytical structure of the navigation problem is one in which the position and velocity of the spacecraft are governed by the inverse-square laws of gravitational attraction. In general, these laws are well defined by the classical two-body problem. The effects of the other heavenly bodies then act as random or unknown forces that perturb the motion of the spacecraft. The measurements made to ascertain the position of the vehicle are in general nonlinear in terms of this state. For example, range measurements and angle measurements may be trigonometrically related to the desired quantities. To avoid many difficulties, the equations are quite often linearized about nominal trajectories with the resulting problem becoming a linear one (see Battin).

It should be clear that the techniques used in space navigation have a carry-over into many other fields of navigation.

7.1.2 Biomedical Systems

With many of the recent advances in biomedical engineering, many physiological systems have been identified and modeled. The physician may desire to monitor the state of his patient, the state being defined as the variables associated with these systems. Such systems are continually undergoing random perturbations. Measurements made on these systems tend to be extremely noisy. For example, in the cardiovascular system, the total volume rate output of a patient may be desired. The heart has a certain model based upon its mechanical structure. To measure this state, electrodes are placed on the chest of the patient. When the muscles contract, they emit an electrical discharge, which can be related to the state of the heart. This measurement is very noisy because of the motions of the patient as well as poor electrode contact. Thus, such a model falls quite readily within the context discussed.

Snyder has also been involved in this field but his interest has been in processes governed by Poisson processes. This occurs in the area of obtaining information from radioactive tracers in the field of nuclear medicine, where the tracers are placed in the circulatory system and the measurements are counts measured by a Geiger counter (see Snyder [4,5]).

7.1.3 Meteorology

In the area of meterology many problems requiring filtering arise. Prime among these are those in the areas of weather forecasting and data-reduction of synoptic data. The motion of storm fronts, their positions, and their velocities may be sought. Information concerning them may come from radar data or from direct probes. This, then, defines both the system and the measurement.

These techniques have also been applied to the estimation of the density of the constituents of the upper atmosphere (30 km to 120 km). Using measurements of scattered light, McGarty [2] has obtained inversion procedures for estimating the density of aerosols, neutral constituents, and ozone. The technique assumes that the particles obey piecewise hydrostatic relations. Using the radiative transfer relationships for scattered light, one can define a nonlinear measurement system. The measurements for this system are the outputs of photomultiplier tubes on a satellite, which may in general be either continuous or discrete. The state is assumed to be, as a result geometric and physical considerations, a random parameter. Such a state is generated by the state equation

$$\frac{d\mathbf{x}(t)}{dt} = 0 \tag{1.1}$$

where $\mathbf{x}_0 = \mathbf{x}(t_0)$ is assumed to be a random variable with known statistics. Thus, $\mathbf{x}(t)$ is constant for all t and equals $\mathbf{x}(t_0)$, which means that $\mathbf{x}(t)$ is a random parameter.

One class of measurements is continuous-time signals that are additively disturbed by white Gaussian measurement noise. It is shown in McGarty [2] that using these measurements obtained from satellites, one can deduce the structure of the upper atmosphere by indirect measurements. An interesting phenomenon occurs when the measurement probe is of low intensity (e.g. starlight). In that case the measurements are impulses with Poisson rates governed by a rate parameter nonlinearly related to the state to be estimated. With these types of measurements a different type of estimation scheme must be used, but fundamentally the approaches are the same.

7.1.4 System Identification

The term "system identification" implies that by making measurements on a system it may be possible to determine its structure. For example, if we have an $n \times 1$ linear time-varying state system but do not know the matrix $\mathbf{A}(t)$, then the process of obtaining $\mathbf{A}(t)$ is called *state identification*. We may know that $a_{ij}(t)$, the ij entry of $\mathbf{A}(t)$, is a random process. We may, furthermore, know its mean and correlation. Then it is possible to construct an augmented state variable $\mathbf{x}^*(t)$ that is

$$\mathbf{x}^*(t) = \begin{bmatrix} a_{11}(t) \\ \vdots \\ a_{nn}(t) \\ x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix} \tag{1.2}$$

This defines $n^2 + n$ nonlinear state system since

$$\dot{\mathbf{a}}(t) = \mathbf{B}(t)\,\mathbf{a}(t) + \mathbf{w}_1(t); \qquad \mathbf{a}(t) = \begin{bmatrix} a_{11} \\ \vdots \\ a_{nn} \end{bmatrix} \tag{1.3}$$

and

$$\dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{w}_2(t) \tag{1.4}$$

so that the augmented state vector is governed by a nonlinear state equation

$$\dot{\mathbf{x}}^*(t) = \mathbf{f}(\mathbf{x}^*, t) + \mathbf{w}(t) \tag{1.5}$$

Thus, given the measurement $\mathbf{z}(t)$, we desire $\hat{a}_{ij}(t)$. Given $\hat{a}_{ij}(t)$, we have *identified* the system.

The application of identification theory is discussed in many references. In general, a global discussion is impossible. One approach similar to the one discussed here is in Kashyap.

7.1.5 Communication Systems

The use of the state variable approach to communication systems has proliferated in the past few years. In the work of Snyder [3], Van Trees [1–3], and Baggeroer [2] many communication systems are analyzed via this technique. The main tool of the communication engineers is the power spectra (Wozencraft and Jacobs). A message to be transmitted can be considered to be characterized by a power spectrum. This is given by the Fourier transform of a stationary correlation matrix. The correlation matrix is defined as:

$$\begin{aligned} \mathbf{K}_x(\tau) &= E[\mathbf{x}(t)\mathbf{x}^T(t+\tau)] \\ &= \begin{cases} \boldsymbol{\Phi}(t+\tau, t)\,\mathbf{P}(t); & \tau \leqq 0 \\ \mathbf{P}(\tau + t)\,\boldsymbol{\Phi}^T(t, t+\tau); & \tau \geqq 0 \end{cases} \end{aligned} \tag{1.6}$$

If the signal is a zero mean process, this completely defines the signal. Many analogue messages can be characterized by the Gaussian approximation.

Thus, in general, the message to be transmitted can be expressed by a linear time-invariant state system:

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{w}(t) \tag{1.7}$$

This system will generate a stationary Gauss-Markov process with a power-spectral density given by a rational polynomial in frequency.

The output of the communication system is a modulated version of the message. For amplitude modulations the received signal can be given by

$$\mathbf{z}(t) = \mathbf{C}(t)\,\mathbf{x}(t) + \mathbf{v}(t) \tag{1.8}$$

where $\mathbf{C}(t)$ is

$$\mathbf{C}(t) = [\sin\ 2\pi f_c t \,:\, 0 \cdots : 0] \tag{1.9}$$

and f_c is the carrier frequency.

For phase modulation (PM) the received signal may be

$$\mathbf{z}(t) = \mathbf{C}\sin[2\pi f_0 t + \mathbf{b}^T\mathbf{x}(t)] + \mathbf{v}(t) \tag{1.10}$$

where $\mathbf{b}$ is an $n \times 1$ vector with constant values. In this the received signal $\mathbf{z}(t)$ is nonlinearly related to the message. Snyder [1–3] carries out the analysis for several systems.

7.1.6 Pattern Recognition

The problem of recognizing a pattern can be phrased as a statistical decision problem parameterized, subject to certain constraints. For example, we may wish to distinguish between two patterns with different means or centers on some multidimensional space. As time evolves, we are provided with more and more points in this multidimensional space, and from it, we are to deduce which of possibly many patterns may exist. If it is possible to classify the patterns by a finite set of parameters that obeys some statistical model and if it is possible to give some statistical structure to the measurements, then a well-posed estimation problem is defined. The work in this area has been surveyed by Ho and Agrawala, who point out the clear usefulness of the types of estimation schemes we developed.

7.1.7 Process Control

In the chemical industry such devices as heat exchangers and distillation columns (and in the nuclear industry, nuclear reactors) present clear examples of uncertain systems subject to control via noisy measurements. The systems are capable of being described by differential state equations with suitable disturbances (see Gould). Likewise, the measurements are usually quite noisy. It is then necessary to estimate the state of such systems for the purpose of controlling them. This concept of coupling estimation and control is discussed by Wonham [2] and by Fleming. A recent example of the use of estimation and control of nuclear reactors using modern estimation theory is given by Moore.

These seven areas only briefly outline some of the applications to which the techniques of recursive estimation can be applied. There are clearly many others, as can be attested to by the increasing body of literature in these areas.

7.2 THEORETICAL EXTENSIONS

We shall complete our discussion of estimation theory in this section by discussing several extensions to the theory developed in the text. Each area is described and the pertinent literature discussed. In some of these extensions a considerably large body of new theory is necessary to understand them fully (for example, stochastic control), but they do follow directly from the general theory we have developed.

7.2.1. Distributed Systems

The systems considered thus far have only contained a single independent variable, time. In many instances the parameter to be estimated depends not only on time but on spatial variables. Such systems are governed by partial-differential equations and the state vector $\mathbf{x}(\mathbf{p}, t)$ is a function of both time t and space $\mathbf{p}$. The propagation of electromagnetic fields and the temperature within a heat exchanger are but two examples of physical situations wherein the state is defined by a distributed parameter system. The standard system for a distributed parameter model is given by

$$\frac{\partial}{\partial t}\,\mathbf{x}(\mathbf{p}, t) = \mathscr{L}_p\,\mathbf{x}(\mathbf{p}, t) + \mathbf{u}(\mathbf{p}, t) \tag{2.1}$$

where $\mathscr{L}_p$ is a spatial operator on the coordinates $\mathbf{p}$ and $\mathbf{u}(\mathbf{p}, t)$ is a temporally white and spatially colored noise field with covariance

$$E[\mathbf{u}(\mathbf{p}, t)\mathbf{u}^T(\mathbf{p}', t')] = \mathbf{Q}(\mathbf{p}, \mathbf{p}', t)\,\delta(t - t') \tag{2.2}$$

where $\mathbf{Q}(\mathbf{p},\mathbf{p}',t)$ is an $n \times n$ positive definite matrix. The solution to this equation is $\mathbf{x}(\mathbf{p},t)$, which is a random field. Random fields were first discussed by Levy [1, 2] and the structure of Gaussian random fields can be found in Wong [1,2], Dudley [1], McKean [1], and Yaglom [1,2]. The Markov nature of the field is not well defined, because of a poor sense of causality in distributed systems (see Wong [2]).

Associated with the distributed state equation is a measurement equation

$$\mathbf{z}(\mathbf{p}, t) = \mathbf{C}(\mathbf{p}, t)\,\mathbf{x}(\mathbf{p}, t) + \mathbf{v}(\mathbf{p}, t) \tag{2.3}$$

where $\mathbf{C}(\mathbf{p},t)$ is an $m \times n$ spatiotemporal matrix and $\mathbf{v}(\mathbf{p},t)$ is spatiotemporal white noise, that is,

$$E[\mathbf{v}(\mathbf{p}, t)\mathbf{v}^T(\mathbf{p}', t')] = \mathbf{R}(\mathbf{p}, t)\delta(\mathbf{p} - \mathbf{p}')\delta(t - t') \tag{2.4}$$

Using a linear spatiotemporal filter and the projection theorem, Tzafestas and Nightengale [1] have formally shown that $\hat{\mathbf{x}}(\mathbf{p},t)$, the linear minimum mean square error estimate, is given by

$$\frac{\partial \hat{\mathbf{x}}}{\partial t}(\mathbf{p}, t) = \mathscr{L}_p \hat{\mathbf{x}}(\mathbf{p}, t) + \int_D \mathbf{P}(\mathbf{p}, s, t)\, \mathbf{C}^T(\mathbf{s}, t)\, \mathbf{R}^{-1}(\mathbf{s}, t) [\mathbf{z}(\mathbf{s}, t) - \mathbf{C}(\mathbf{s}, t)\, \hat{\mathbf{x}}(\mathbf{s}, t)]\, d\mathbf{s} \tag{2.5}$$

where $\mathbf{P}(\mathbf{p}, \mathbf{s}, t)$ is a covariance matrix generated by the space-time equation

$$\frac{\partial \mathbf{P}}{\partial t}(\mathbf{p}, \mathbf{s}, t) = \mathscr{L}_p \mathbf{P}(\mathbf{p}, \mathbf{s}, t) + \mathbf{P}(\mathbf{p}, \mathbf{s}, t)\, \mathscr{L}_s^T + \mathbf{Q}(\mathbf{p}, \mathbf{s}, t) - \int_D \mathbf{P}(\mathbf{p}, \mathbf{r}, t) \mathbf{C}^T(\mathbf{r}, t) \mathbf{R}^{-1}(\mathbf{r}, t) \mathbf{C}(\mathbf{r}, t) \mathbf{P}(\mathbf{r}, \mathbf{s}, t)\, d\mathbf{r} \tag{2.6}$$

where $\mathscr{L}_s^T$ is the adjoint operator of $\mathscr{L}_p$.

These estimation equations are the spatial analogues of the continuous-time lumped parameter Kalman-Bucy equations. Results on simulations using these equations appear in Tzafestas and Nightingale [2, 3]. Approaches using other techniques have been given by Meditch [3], using a minimization of a functional; Kushner [6]; Balakrishnan; and Falb. A thesis by Bensoussan has reviwed the area and has presented the results in a well posed mathematical framework.

These techniques have been used in Van Trees [3] to obtain optimum detector-estimator structures for signals that suffer delay and Doppler distortion.

7.2.2 Smoothing

The estimation problem, also called the *filtering problem*, obtains an estimate of the state of a system at time t_i, given measurements from time t_0 to time t. If, however, we want to estimate the state at some time t, given measurements from $t_0 < t$ to $t_1 > t$, then this is called *smoothing*. It can easily be shown that the optimal estimate in this case is

$$\hat{\mathbf{x}}(t|t_1) = E[\hat{\mathbf{x}}(t)|O_{t_0, t_1}] \tag{2.7}$$

where $O_{t_0, t}$ is the minimum σ-field generated by the observations over the interval $[t_0, t_1]$. Using this, Frost and Kailath and Frost [1] have shown that if t_0 and t_1 are fixed, then $\mathbf{x}(t|t_1)$, the optimal smoothed estimate, is generated by

$$\frac{d\hat{\mathbf{x}}}{dt}(t|t_1) = \mathbf{A}(t)\hat{\mathbf{x}}(t|t_1) + \mathbf{Q}(t)\mathbf{P}^{-1}(t)[\hat{\mathbf{x}}(t|t_1) - \hat{\mathbf{x}}(t)] \tag{2.8}$$

and $\mathbf{P}^{-1}(t)$ is generated by

$$\dot{\mathbf{P}}^{-1}(t) = -\mathbf{P}^{-1}(t)\mathbf{A}(t) - \mathbf{A}^T(t)\mathbf{P}^{-1}(t) - \mathbf{P}^{-1}(t)\mathbf{Q}(t)\mathbf{P}^{-1}(t) + \mathbf{C}^T(t)\mathbf{R}^{-1}(t)\mathbf{C}(t) \tag{2.9}$$

Clearly, the smoothing operation requires the estimation (filtered estimate) of $\mathbf{x}(t)$, given $O_{t_0, t}$. The covariance of the smoothed estimate $\Sigma(t|t_1)$ is

$$E[(\hat{\mathbf{x}}(t|t_1) - \mathbf{x}(t))(\hat{\mathbf{x}}(t|t_1) - \mathbf{x}(t))^T] = \Sigma(t|t_1) \tag{2.10}$$

This is generated by

$$\dot{\Sigma}(t|t_1) = (\mathbf{A}(t) + \mathbf{Q}(t)\mathbf{P}^{-1}(t))\Sigma(t|t_1) + \Sigma(t|t_1)(\mathbf{A}(t) + \mathbf{Q}(t)\mathbf{P}^{-1}(t))^T - \mathbf{Q}(t) \tag{2.11}$$

The paper by Kailath and Frost also discusses the filtering equation for the case where $\mathbf{x}(t)$ is to be estimated with t_0 fixed and t_1 increasing. This is called *fixed-point smooting*.

The discrete-time version of smoothing for both fixed data and fixed point has been discussed by Meditch [1,2]. Meditch [2] provides an extensive discussion of both the derivation and use of such smoothing routines. Smoothing generally improves the performance of estimates, but the price paid is an increased amount of computation and a delay of $t_1 - t$ units of time.

7.2.3 Detection Theory

Detection theory concerns the choosing of one of many hypotheses based upon a set of observations. The simplest detection problem is the binary hypothesis testing problem, where there are two hypotheses H_0 and H_1. Under H_0, the received signal $\mathbf{z}(t)$ contains only white noise, whereas under hypothesis H_1, $\mathbf{z}(t)$ contains white noise plus a random process $\mathbf{x}(t)$. Thus,

$$H_0: d\mathbf{z}(t) = d\mathbf{v}(t) \tag{2.12}$$

$$H_1: d\mathbf{z}(t) = d\mathbf{x}(t) + d\mathbf{v}(t) \tag{2.13}$$

Now $\mathbf{x}(t)$ is assumed to be generated by

$$d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}, t)dt + d\mathbf{n}(t) \tag{2.14}$$

where $\mathbf{n}(t)$ is a Wiener process. Souslin and Stratonovich; Stratonovich and Souslin; Souslin; Kailath [3–5]; and Duncan [1,2] have shown that the likelihood ratio Λ for this system is given by

$$\Lambda(T) = \exp\left[\int_0^T \hat{\mathbf{x}}_1(t)d\mathbf{z}(t) - \tfrac{1}{2}\int_0^T \hat{\mathbf{x}}_1^2(t)dt\right] \tag{2.15}$$

where the first integral is to be interpreted in the Ito sense and where

$$\hat{\mathbf{x}}_1(t) = E[\mathbf{x}(t)|O_{t_0,t}, H_1] \tag{2.16}$$

which is the conditional mean of $\mathbf{x}(t)$ given the minimum σ-field generated by the observation process assuming hypothesis H_1. This result was derived for Gaussian processes by Schweppe [1] and Souslin and Stratonovich. The simplest and most palatable approach is contained in Kailath [9, 8] using the innovations approach. This is the most general result, since it allows for dependence between the state and the noise. This general result has not been obtained by noninnovations techniques.

The likelihood ratio $\Lambda(T)$ is used to test for the presence of either hypothesis. If $\Lambda(T)$ is greater than a threshold, then we choose H_1; otherwise, we choose H_0. The performance of this detection scheme is given by the probability of error and has been discussed in Evans [1,2].

7.2.4 Stochastic Control

Throughout our analysis we have assumed that the state equation was undriven. In many instances this is not the case, and in one case a control is applied in order for the system to evolve in a predescribed fashion. In the case of deterministic control theory, the choice of controls or driving functions to minimize or maximize given cost criteria is called *optimal control* (see Athans and Falb). For dynamical systems disturbed by stochastic signals, optimal control techniques are also available. Consider the following problem Let the state equation be formally written as

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}(t)\,\mathbf{x}(t) + \mathbf{m}(t) + \mathbf{u}(t) \tag{2.17}$$

where $\mathbf{u}(t)$ is white noise and $\mathbf{m}(t)$ is a deterministic control. The measurements are

$$\mathbf{z}(t) = \mathbf{C}(t)\,\mathbf{x}(t) + \mathbf{v}(t) \tag{2.18}$$

where $\mathbf{v}(t)$ is also white noise. Now the choice of $\mathbf{m}(t)$ is made such that the cost function

$$J[\mathbf{m}] = E\left[\int_0^T L[t, \mathbf{m}(t), \mathbf{x}(t)]\ dt\right] \tag{2.19}$$

is optimized.

Wonham [2] has shown that for this model it is possible to separate the control problem from the estimation problem; namely, what may be sought is a feedback control where

$$\mathbf{m}(t) = \phi[t, \mathbf{x}(t)] \tag{2.20}$$

where $\mathbf{x}(t)$ is the state. Because of the system structure, $\mathbf{x}(t)$ can only be estimated from $\mathbf{z}(t)$. Wonham has shown that there exist optimal feedback controls such that the optimum $\mathbf{m}(t)$, $\mathbf{m}^o(t)$ is

$$\mathbf{m}^o(t) = \phi[t, \hat{\mathbf{x}}(t)] \tag{2.21}$$

Thus, it is possible to separate the control and estimation problem. Other results in this area are in Wonham [3], Fleming, Kushner [7], and Meditch [2]. A considerable extension of the theory is necessary to obtain these results.

7.2.5 Set Theoretic Approaches

The model that has been proposed let both the system disturbance and the

measurement disturbance be random processes. An alternate approach is to consider the system

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x}(t) + \mathbf{u}(t) \tag{2.22}$$

and the measurement

$$\mathbf{z}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{v}(t) \tag{2.23}$$

to be such that $\mathbf{u}(t)$ and $\mathbf{v}(t)$ are not random, but unknown yet bounded, disturbances. That is, the $\mathbf{u}(t)$ belong to the set Ω_Q where

$$\Omega_Q = \{\mathbf{u}: \mathbf{u}^T\mathbf{Q}^{-1}\mathbf{u} \leqq 1; \forall t\} \tag{2.24}$$

is an ellipsoidal set.
Likewise, we assume all $\mathbf{v}(t)$ belong to Ω_R, where

$$\Omega_R = \{\mathbf{v}: \mathbf{v}^T\mathbf{R}^{-1}\mathbf{v} \leqq 1; \forall t\} \tag{2.25}$$

The state is said to be within an ellipsoid, a set defined by a quadratic criterion. The initial set is $S(t_0)$ and is defined by

$$S(t_0) = \{\mathbf{x}: (\mathbf{x} - \hat{\mathbf{x}}(t_0))^T \Sigma^{-1}(t_0)(\mathbf{x} - \hat{\mathbf{x}}(t_0)) \leqq 1\} \tag{2.26}$$

where $\Sigma(t_0)$ is an $n \times n$ positive definite matrix and $\mathbf{x}(t_0)$ is the center of the ellipsoid. Thus, at $t = t_0$ we assume $\mathbf{x}$ lies within this ellipsoid. As time progresses, we want to follow the state by means of similar ellipsoids. Schweppe [2] shows that at time t_k the bounding ellipsoid is given by $\bar{\Sigma}(t_k)$ where

$$\bar{\Sigma}(t_k) = \{\mathbf{x}: (\mathbf{x} - \hat{\mathbf{x}}(t_k))^T \Sigma^{-1}(t_k)(\mathbf{x} - \hat{\mathbf{x}}(t_k)) \leqq 1\} \tag{2.27}$$

Thus, $\mathbf{x}$ lies within this ellipsoid at time t_k. He shows how $\bar{\Sigma}(t_k)$ and $\hat{\mathbf{x}}(t_k)$ can be obtained recursively from the measurements. The resulting equations are quite similar to the Kalman discrete-time filter equations, as he notes.

This technique provides a different approach to the estimation problem, by eliminating the random nature and introducing deterministic uncertainty. More recent results in this area are given by Bertsekas and Rhodes [1,2], who discuss continuous-time structure, and by Schlaepfer, who uses this technique in distributed parameter systems.

This completes our development of the theory of nonlinear estimation. It has required the review of the state-space theory, a development of probability and stochastic process theory, an analysis of Hilbert spaces, and a careful study of propagation equations. It is a theory that can easily be applied to some difficult problems yet also arduously applied to some apparently unassuming structures. It provides a study of a struggle against the uncertainty of nature and insight into the nature of stochastic systems and state estimation.

APPENDIX A

EXISTENCE AND UNIQUENESS PROPERTIES OF DIFFERENTIAL EQUATIONS

The systems models developed in Chapter 2 were for deterministic differential equations where there was assumed to be arbitrary nonlinearities. However, restrictions on the types of functions that appear in the state equations must be made if the solutions are to make sense. Specifically, we are interested in two basic issues. The first is whether a solution even exists for the differential equation. To show this, we use the constructive approach by showing how a solution may be obtained. Second, we wish to show uniqueness, that is, whether a solution, if it does exist, is the only one. To do these two things we must limit the class of functions that we will use. These limits are discussed in this appendix.

The equation of interest is the following:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t); \qquad \mathbf{x}(t_0) = \mathbf{x}_0 \tag{A.1}$$

where the nonlinearity $\mathbf{f}(\mathbf{x}, t)$ is to be limited.

We shall ask the question. Does there exist a solution to (A.1), $\mathbf{x}(t)$ that passes through $\mathbf{x}_0$ at time t_0 and satisfies the differential equation throughout the rest of the trajectory.

We begin by defining the Lipschitz condition on a cylinder C.

DEFINITION A.1. Let $C(\lambda, \tau)$ be the set of points

$$C(\lambda, \tau) = \{\mathbf{x}, t: \|\mathbf{x} - \mathbf{x}_0\| \leqq \lambda; |t - t_0| \leqq \tau\} \tag{A.2}$$

and call C a cylinder of radius λ and length 2τ and center $\mathbf{x}_0$, t_0. And

$$\|\mathbf{x} - \mathbf{x}_0\| = ((\mathbf{x} - \mathbf{x}_0)^T(\mathbf{x} - \mathbf{x}_0))^{1/2} \tag{A.3}*$$

This is shown in Figure A.1.

DEFINITION A.2. A vector-valued function $\mathbf{f}(\mathbf{x},t)$ is said to satisfy a Lipschitz

*This is the norm derived from the inner product l^2 on an n dimensional euclidean space.

Figure A.1 Lipschitz conditions.

condition on C if there exists a constant k such that if $(\mathbf{x}_1, t)$ and $(\mathbf{x}_2, t)$ are any two points on C, then

$$\|\mathbf{f}(\mathbf{x}_1, t) - \mathbf{f}(\mathbf{x}_2, t)\| \leqq k\|\mathbf{x}_1 - \mathbf{x}_2\| \tag{A.4}$$

This is shown in Figure A.1. Note that for this function to be Lipschitz on the entire cylinder C it must satisfy (A.4) for every set of $(\mathbf{x}_i, \mathbf{x}_j)$ belonging to C. Yet C is a specific region of the state spece; thus, it is local over C. If this holds for all λ and τ, that is, for all possible cylinders—then the condition is a global Lipschitz condition.

The Lipschitz condition implies that in order for it to be satisfied, the derivative must change less slowly than some fraction of the change in state.

Example. If $\mathbf{f}(\mathbf{x}) = (\mathbf{x})^{1/2}$, then as $\mathbf{x}_1 \to 0$ and $\mathbf{x}_2 = 0$, there is no k to satisfy

$$\frac{(|\mathbf{x}|)^{1/2}}{|\mathbf{x}|} \leqq k \tag{A.5}$$

Now $\lim_{\mathbf{x}\to 0} \to \infty$ for k, so that this function is *not* Lipschitz.

We shall now follow a proof in Brockett for uniqueness and then the one in Ince for existence.

THEOREM A.1

Let there exist a solution $\mathbf{x}(t)$ given the system

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}, t); \mathbf{x}(t_0) = \mathbf{x}_0 \tag{A.6}$$

and assume that $\mathbf{f}(\mathbf{x}, t)$ is Lipschitz in a cylinder $C(\lambda, \tau)$. Then there exists at most one solution, $\phi(t, \mathbf{x}_0, t_0)$ on $C(\lambda,\tau)$, passing through the initial point.

Proof. Let us first note that we are initially assuming existence. We shall prove the theorem by assuming that *two* solutions exist and show that this leads to a contradiction. Let $\phi_1(t, \mathbf{x}_0, t_0)$ and $\phi_2(t, \mathbf{x}_0, t_0)$ be two solutions in $C(\lambda, \tau)$ passing through $\mathbf{x}_0$ at t_0. They must also satisfy

$$\dot{\phi}_1(t, \mathbf{x}_0, t_0) = \mathbf{f}(\phi_1, t) \tag{A.7}$$

$$\dot{\phi}_2(t, \mathbf{x}_0, t_0) = \mathbf{f}(\phi_2, t) \tag{A.8}$$

We also assumed that they were Lipschitz on $C(\lambda, \tau)$. This implies

$$\|\mathbf{f}(\mathbf{x}_2, t) - \mathbf{f}(\mathbf{x}_1, t)\| \leqq k\|\mathbf{x}_2 - \mathbf{x}_1\| \tag{A.9}$$

Now subtract (A.8) from (A.7) and obtain

$$\dot{\phi}_1(t, \mathbf{x}_0, t_0) - \dot{\phi}_2(t, \mathbf{x}_0, t_0) = \mathbf{f}(\phi_1, t) - \mathbf{f}(\phi_2, t) \tag{A.10}$$

Now recall that for any $\mathbf{x}$.

$$\frac{d}{dt}(\mathbf{x}^T\mathbf{x}) = \frac{d\mathbf{x}^T}{dt}\mathbf{x} + \mathbf{x}^T\frac{d\mathbf{x}}{dt} \tag{A.11}$$

Furthermore,

$$\|\mathbf{x}\|^2 = \mathbf{x}^T\mathbf{x} \tag{A.12}$$

Therefore,

$$\begin{aligned}
&\frac{d}{dt}(\|\phi_1(t, t_0, \mathbf{x}_0) - \phi_2(t, t_0, \mathbf{x}_0)\|^2) \\
&= \frac{d}{dt}(\phi_1(t, t_0, \mathbf{x}_0) - \phi_2(t, t_0, \mathbf{x}_0))^T(\phi_1(t, t_0, \mathbf{x}_0) - \phi_2(t, t_0, \mathbf{x}_0)) \\
&+ (\phi_1(t, t_0, \mathbf{x}_0) - \phi_2(t, t_0, \mathbf{x}_0))^T\frac{d}{dt}(\phi_1(t, t_0, \mathbf{x}_0) - \phi_2(t, t_0, \mathbf{x}_0))
\end{aligned} \tag{A.13}$$

Now let

$$\phi_1 = \phi_1(t, t_0, \mathbf{x}_0) \tag{A.14}$$

$$\phi_2 = \phi_2(t, t_0, \mathbf{x}_0) \tag{A.15}$$

Using (A.10) in (A.13) and realizing that if $\mathbf{a}^T\mathbf{b}$ is a scalar

$$\mathbf{a}^T\mathbf{b} = \mathbf{b}^T\mathbf{a} \tag{A.16}$$

we obtain for (A.13)

$$\frac{d}{dt}\left(\|\phi_1 - \phi_2\|^2\right) = 2(\phi_1 - \phi_2)^T(\mathbf{f}(\phi_1) - \mathbf{f}(\phi_2)) \tag{A.17}$$

Now using the Lispchitz inequality

$$\begin{aligned}(\phi_1 - \phi_2)^T(\mathbf{f}(\phi_1) - \mathbf{f}(\phi_2)) &\leqq k(\phi_1 - \phi_2)^T(\phi_1 - \phi_2) \\ &= k\|\phi_1 - \phi_2\|^2\end{aligned} \tag{A.18}$$

Thus, in (A.17) we have

$$\frac{d}{dt}\left(\|\phi_1 - \phi_2\|^2\right) \leqq 2k\|\phi_1 - \phi_2\|^2 \tag{A.19}$$

Now define

$$\sigma(t, \mathbf{x}_0, t_0) = \|\phi_1 - \phi_2\|^2 \tag{A.20}$$

And note that

$$\sigma(t_0, \mathbf{x}_0, t_0) = 0 \tag{A.21}$$

Using this in (A.19), we have

$$\frac{d}{dt}\sigma(t, \mathbf{x}_0, t_0) \leqq 2k\sigma(t, \mathbf{x}_0, t_0) \tag{A.22}$$

or rearranging

$$\frac{d}{dt}\sigma(t, x_0, t_0) - 2k\sigma(t, \mathbf{x}_0, t_0) \leqq 0 \tag{A.23}$$

But multiply both sides by $e^{-2k(t-t_0)}$,

$$\begin{aligned}&\left(\frac{d\sigma(t, \mathbf{x}_0, t_0)}{dt} - 2k\sigma(t, \mathbf{x}_0, t)\right)e^{-2k(t-t_0)} \\ &\quad = \frac{d}{dt}\left[\sigma(t, \mathbf{x}_0, t_0)e^{-2k(t-t_0)}\right]\end{aligned} \tag{A.24}$$

Therefore, (A.24) becomes

$$\frac{d}{dt}\left[\sigma(t, \mathbf{x}_0, t_0)e^{-2k(t-t_0)}\right] \leqq 0 \tag{A.25}$$

Now integrate both sides from t_0 to t and obtain

$$\sigma(t, \mathbf{x}_0, t_0)e^{-2k(t-t_0)} \leqq 0 \tag{A.26}$$

But $\sigma(t, \mathbf{x}_0, t_0)$ as defined in (A.20) is always positive so that the *only* solution is for

$$\sigma(t, \mathbf{x}_0, t_0) = 0 \tag{A.27}$$

which implies that (see Chapter 4, Section 4.1, for a discussion of the property of norms.;

$$\phi_1(t, \mathbf{x}_0, t_0) = \phi_2(t, \mathbf{x}_0, t_0) \tag{A.28}$$

or that the solution is unique. ∎

Now the a priori assumption of this previous proof was that indeed a solution existed. We now want to prove that such a solution exists. We shall do so by Picard's method of successive approximations. But, before doing so, we must introduce the concepts of continuity and convergence. These concepts are essential to an understanding not only of the existence problem but of such things as cost function.

DEFINITION A.3. A function $\mathbf{f}$ is *continuous* at a point $\mathbf{x}_0$ if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$|\mathbf{x} - \mathbf{x}_0| < \delta \tag{A.29}$$

implies

$$|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_0)| < \varepsilon \tag{A.30}$$

Let us now consider convergence. That is, if we have some set of functions $\{\mathbf{f}_n\}$, what do we mean by the convergence of this set?

DEFINITION A.4. Let $\{\mathbf{f}_n\}$ be a sequence of functions from a set X into a set Y on which we define a distance d. That is, let

$$\mathbf{f}_n : X \to Y \tag{A.31}$$

and let

$$d(\mathbf{f}_n(\mathbf{x}), \mathbf{g}(\mathbf{x})) = \|\mathbf{f}_n(\mathbf{x}) - \mathbf{g}(\mathbf{x})\| \tag{A.32}$$

Then $\{\mathbf{f}_n\}$ is said to converge uniformly to a function $\mathbf{g}(\mathbf{x})$

$$\mathbf{g} : X \to Y \tag{A.33}$$

if for every $\varepsilon > 0$ there exists an $n_0(\varepsilon)$ such that $n > n_0$ implies

$$\|\mathbf{f}_n(\mathbf{x}) - \mathbf{g}(\mathbf{x})\| < \varepsilon \tag{A.34}$$

for all $\mathbf{x} \in X$.

Now an important theorem will be stated that links the two concepts above.

THEOREM A.2.

Let $\{\mathbf{f}_n\}$ be a sequence of continuous functions from X into Y. If $\{\mathbf{f}_n\}$ converges uniformly to $\mathbf{g} : X \to Y$, then $\mathbf{g}$ is continuous.

Proof. See Lipschitz (p. 209). ∎

Thus, when we are dealing with sequences that converge uniformly to $\mathbf{g}$, the resulting $\mathbf{g}$ is continuous. This will be an important factor in our results.

With these ideas we can now proceed and prove the existence theorem. We will see in Appendix B that this idea will be carried over to the concept of existence of solutions to random process equations. To those familiar with functional analysis this will only require us to change the norm and thus the convergence criteria. The general technique follows directly from that observation.

THEOREM A.3

Given that $\mathbf{f}(\mathbf{x}, t)$ is bounded, globally Lipschitz, and continuous in t. Then given any $(\mathbf{x}_0, t_0)$, *there exists a uniuqe solution* of the differential equation passing through $(\mathbf{x}_0, t_0)$.

Proof. Now we are given

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t) \tag{A.35}$$

Let us rewrite this as an integral equation. Let $\mathbf{x}(t) = \phi(t)$; then

$$\phi(t) = \mathbf{x}_0 + \int_0^t \mathbf{f}(\phi(\xi), \xi)\ d\xi \tag{A.36}$$

Now (A.36) satisfies $\phi(t_0) = \mathbf{x}_0$ and, by differentiation, (A.35) Thus, $\phi(t)$ is a solution. Let us now approximate this solution. Let

$$\begin{aligned}
\phi_1(t) &= \mathbf{x}_0 + \int_{t_0}^t \mathbf{f}(x_0, \xi)\ d\xi \\
\phi_2(t) &= \mathbf{x}_0 + \int_{t_0}^t \mathbf{f}(\phi_1(\xi), \xi)\ d\xi \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\phi_n(t) &= \mathbf{x}_0 + \int_{t_0}^t \mathbf{f}(\phi_{n-1}(\xi), \xi)\ d\xi
\end{aligned} \tag{A.37}$$

Now let

$$\delta_{n+1}(t) = \|\phi_{n+1}(t) - \phi_n(t)\| \tag{A.38}$$

Using (A.37) in (A.38), we have

$$\delta_{n+1}(t) = \left\| \int_{t_0}^t [\mathbf{f}(\phi_n, \xi) - \mathbf{f}(\phi_{n-1}, \xi)]\ d\xi \right\| \tag{A.39}$$

Bringing the absolute values inside yields

$$\delta_{n+1}(t) \leqq \int_{t_0}^t \|\mathbf{f}(\phi_n, \xi) - \mathbf{f}(\phi_{n-1}, \xi)\|\ d\xi \tag{A.40}$$

Now, using the Lipschitz condition,

$$\delta_{n+1}(t) \leqq \int_{t_0}^t k\|\phi_n - \phi_{n-1}\|\ d\xi \tag{A.41}$$

but by (A.38)

$$\delta_{n+1}(t) \leqq k \int_{t_0}^t \delta_n(\xi)\ d\xi \tag{A.42}$$

Now, let us evaluate this bound. By definition

$$\phi_0 = \mathbf{x}_0 \tag{A.43}$$

so that

$$\delta_1(t) \leqq k \int_{t_0}^{t} \|\mathbf{f}(\phi_0, \xi) - \mathbf{f}(\phi_{-1}, \xi)\| \, d\xi \tag{A.44}$$

and again by definition

$$\mathbf{f}(\phi_{-1}, \xi) \equiv \mathbf{0} \tag{A.45}$$

Therefore,

$$\delta_1(t) \leqq k \int_{t_0}^{t} \|\mathbf{f}(\phi_0, \xi)\| \, d\xi \tag{A.46}$$

But $\mathbf{f}(\phi, t)$ is bounded in the interval $[t_0, t]$. Thus

$$\|\mathbf{f}(\phi_0, \xi)\| < M \tag{A.47}$$

we obtain

$$\delta_1(t) \leqq K \int_{t_0}^{t} M \, d\xi \tag{A.48}$$

or

$$\delta_1(t) < KM(t - t_0) \tag{A.49}$$

Then

$$\delta_2(t) \leqq K \int_{t_0}^{t} \delta_1(\xi) \, d\xi \tag{A.50}$$

Substituting (A.49),

$$\delta_2(t) \leqq K^2(t - t_0)^2 M/2 \tag{A.51}$$

Continuing, we can show that

$$\delta_n(t) \leqq M \frac{(K(t - t_0))^n}{n!} \tag{A.52}$$

which converges. Now

$$\phi_n(t) = \mathbf{x}_0 + \sum_{i=0}^{n-1} [\phi_{i-1}(t) - \phi_i(t)] \tag{A.53}$$

Furthermore, the series in (A.53) converges by virtue of (A.52). Then if this is true,

$$\mathbf{x}(t) = \lim_{n \to \infty} \phi_n(t) \tag{A.54}$$

exists and is a continuous function of t on the interval $[t_0, t]$. Thus, the series $\{\phi_n(t)\}$ is *uniformly convergent* to $\mathbf{x}(t)$ in the interval (Ince. p. 64). Now using this, we will show that $\mathbf{f}(\phi_n, t)$ is also uniformly convergent.

Now

$$\|\mathbf{f}(\mathbf{x}(t), t) - \mathbf{f}(\phi_n(t), t)\| \leqq K \|\mathbf{x}(t) - \phi_n(t)\| \tag{A.55}$$

by the Lipschitz condition. But $\{\phi_n(t)\}$ is uniformly convergent to $\mathbf{x}(t)$ on $[t_0, t]$. Therefore, by (A.55), so is $\mathbf{f}(\phi_n, t)$. Now we will let

$$\phi_n(t) = \mathbf{x}_0 + \int_{t_0}^{t} \mathbf{f}(\phi_{n-1}, \xi)\ d\xi \tag{A.56}$$

And taking the limit as $n \to \infty$, we observe that

$$\lim_{n\to\infty} \phi_n(t) = \mathbf{x}_0 + \lim_{n\to\infty} \int_{t_0}^{t} \mathbf{f}(\phi_{n-1}, \xi)\ d\xi \tag{A.57}$$

We would like to take the limit inside the integral. The following theorem allows us to do it (Rudin [2], p. 31).

THEOREM A.4

Suppose $(t - t_0) < \infty$, $t > -t_0$ and $\{\mathbf{f}_n\}$ is a sequence that is bounded and is uniformly convergent to $\mathbf{f}$ on $[t_0, t]$: then

$$\lim_{n\to\infty} \int_{t_0}^{t} \mathbf{f}_n\ d\xi = \int_{t_0}^{t} \mathbf{f}\ d\xi \tag{A.58}$$

Now, using this theorem and the fact that $\mathbf{f}(\phi_n, \xi)$ is uniformly convergent to $\mathbf{f}(\mathbf{x}(\xi), \xi)$, we have for (A.57)

$$\mathbf{x}(t) = \mathbf{x}(0) + \int_{t_0}^{t} \mathbf{f}(\mathbf{x}(\xi), \xi)\ d\xi \tag{A.59}$$

which shows that indeed a solution does exist. ∎

The restriction on continuity of $\mathbf{f}$ may be too strong, as is seen in Ince. Also $(t - t_0)$ may be infinite, but again the previous theorem would not hold. If the reader seeks more information, the above reference is useful. The point made by all of these discussions is that one must be careful in blindly solving the problem. Note that in the proof of existence, the series had to converge. If a programmer just stops computation after a thousand steps, he might get a numerical answer. Yet it might not be the true answer. Indeed, there may not even exist a solution! Thus, the purpose of the depth of coverage was to go through all the detail, clearly state all the assumptions, show the user that a great deal of thought has already gone into the questions of existence and uniqueness, and advise him strongly to give some thought to the subject himself.

APPENDIX B

EXISTENCE AND UNIQUENESS PROPERTIES OF STOCHASTIC DIFFERENTIAL EQUATIONS

Chapter 3 developed the idea of stochastic differential equations. In Appendix A we proved the existence and uniqueness of the solutions to the deterministic state equations. This appendix considers the problem of showing that solutions to stochastic differential equations exist and are unique under certain conditions on the functions appearing in the equations. We specifically consider the scalar stochastic differential equation

$$dx(t) = f(x(t))\ dt + \sigma(x(t))\ dw(t) \tag{B.1}$$

where $x(t)$ is the scalar process and $w(t)$ is a normalized Wiener process. We follow Doob [2] in this proof.

We first present several lemmas concerning inequalities, bounds, and convergence that will be necessary to prove the existence and uniqueness of the solution to the stochastic differential equation. In particular, the Borel-Cantelli lemma will be developed, and it will be used to demonstrate how to treat events within probability-1 context. The last lemma requires the definition of a semimartingale and produces an inequality termed the *semi-martingale* inequality.

LEMMA B.1. For any Riemann integrable function $f(t)$ defined on the interval $[a,b)$, we have

$$\left[\int_a^b |f(t)|\ dt\right]^2 \leqq (b-a)\int_a^b (f(t))^2\ dt \tag{B.2}$$

Proof. From Chapter 2, equation (4.4) we know that

$$\left[\sum_{i=1}^n |x_i|\right]^2 \leqq n \sum_{i=1}^n x_i^2 \tag{B.3}$$

Now let

$$x_i = f(t_i)\Delta_i \tag{B.4}$$

and let Δ_i be a set of positive measures. Then

$$[\sum_{i=1}^{n} |f(t_i)|\Delta_i]^2 \leqq n \sum_{i=1}^{n} (f(t_i))^2 \Delta_i^2 \tag{B.5}$$

But now

$$\sum_{i=1}^{n} n\Delta_i \equiv (b-a) \tag{B.6}$$

Thus,

$$[\sum_{i=1}^{n} |f(t_i)|\Delta_i]^2 \leqq \sum_{i=1}^{n} (f(t_i))^2 \Delta_i n\Delta_i \tag{B.7}$$

Now if $\Delta_i = \Delta$, which can be done because of the assumed smoothness of $f(t)$, we have

$$[\sum_{i=1}^{n} |f(t_i)|\Delta]^2 \leqq (b-a) \sum_{i=1}^{n} (f(t_i))^2 \Delta \tag{B.8}$$

and in the limit as $n \to \infty$, $\Delta \to 0$, we obtain

$$[\int_a^b |f(t)|dt]^2 \leqq (b-a)\int_a^b f^2(t)\ dt \blacksquare \tag{B.9}$$

LEMMA B.2. Let x and y be any two functions of time. Then

$$|x+y|^2 \leqq 2|x|^2 + 2|y|^2 \tag{B.10}$$

Proof. For any x and y,

$$|x+y| \leqq |x| + |y| \tag{B.11}$$

Thus, squaring both sides,

$$|x+y|^2 \leqq |x|^2 + |y|^2 + 2|x||y| \tag{B.12}$$

Now we also know that for any constant C_1 (positive or negative),

$$|x| - |y| \geqq C_1 \tag{B.13}$$

Then, squaring,

$$|x|^2 + |y|^2 - 2|x||y| \geqq C_1^2 \geqq 0 \tag{B.14}$$

Thus,

$$2|x||y| \leqq |x|^2 + |y|^2 \tag{B.15}$$

Now substituting into the original inequality we have

$$|x+y|^2 \leqq |x|^2 + |y|^2 + |x|^2 + |y|^2 \leqq 2|x|^2 + 2|y|^2 \blacksquare \tag{B.16}$$

The following lemma is the one that allows us to prove theorems with probability one. Historically, it was used in the proof of the strong law of large numbers. Comments on its use are contained in Feller [1, pp. 189–190], Breiman (pp. 41–42), or Hida (pp. 7–8).

LEMMA B.3. (Borel-Cantelli) Let $\{S_i\}_{i=1}^{\infty}$ be a sequence of events. Let

$$S = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} S_k = \limsup_k S_k \tag{B.17}$$

(i) If $\sum_{k=1}^{\infty} P[S_k] < \infty$, then $P(S) = 0$.

(ii) If the events in each finite subsequence of $S_1, S_2, \cdots$ are mutually independent and

$$\sum_{k=1}^{\infty} P(S_k) = \infty$$

then $P(S) = 1$.

Proof. Let Ω be the sample space upon which the events are defined. Then, clearly, S is the set of all $\omega \in \Omega$ that belong to infinitely many of the S_n. Hence, the occurrence of S is equivalent to the occurrence of infinitely many of the S_n.

Observe that for any $n > 0$, we have

$$S \in \bigcup_{k=n}^{\infty} S_k$$

Now, from the properties of the probability measure, we observe

$$0 \leqq P(S) \leqq P\left(\sum_{k=n}^{\infty} S_k\right) \leqq \sum_{k=n}^{\infty} P(S_k) \tag{B.18}$$

Thus, as $n \to \infty$, we see that since the series is convergent, we can bound $P(S)$ arbitrarily close to zero:

$$0 \leqq P(S) \leqq \varepsilon \tag{B.19}$$

by choosing n large enough in (B.18). Indeed, by making n large enough, we obtain as $n \to \infty$

$$\lim_{n \to \infty} P(S) \to 0 \tag{B.20}$$

which proves the first part of the theorem.

We shall reverse things to prove what is essentially an equivalent statement. Consider again the set of points $\omega \in \Omega$ such that the only events that belong to this set are those belonging to $S_1, \cdots, S_n$. Let G_n be the set of ω that is contained in *no* S_k for $k > n$. Then the set G_n does *not* belong to the union of the sets $S_1, \cdots, S_n$ and *must* therefore belong to the complement of the union, namely,

$$G_n \subset \left(\bigcup_{k=n}^{\infty} S_k\right)^c \tag{B.21}$$

or, by De Morgan's law

$$G_n \subset \bigcap_{k=n}^{\infty} S_k^c \tag{B.22}$$

Thus,

$$P(G_n) \leqq P\left[\bigcap_{k=n}^{\infty} S_k^c\right] \tag{B.23}$$

and by the independence assumption,

$$P(G_n) \leqq \prod_{k=n}^{\infty} P(S_k^c) = \prod_{k=n}^{\infty} (1 - P(S)) \tag{B.24}$$

Now, using the fact that

$$1 - x \leqq e^{-x}; \qquad x > 0 \tag{B.25}$$

we obtain

$$P(G_n) \leqq \exp\left[-\sum_{k=n}^{\infty} P(S_k)\right] \tag{B.26}$$

But for any finite n, $\sum_{k=n}^{\infty} P(S_k)$ diverges so that

$$P(G_n) \to 0 \tag{B.27}$$

Note that letting $n \to \infty$ is just redefining the same set of measure 0. Now we can say that since $P(G_n) = 0$, this implies that the probability of ω being contained in no set greater than S_n is 0. Thus, since this holds for all n, this implies that the probability that ω belongs to an infinite number of S_n is 1. Thus,

$$P(S) = 1 \qquad \blacksquare \tag{B.28}$$

This lemma is strong in the sense that it states that an event occurs with probability 1, meaning the events not in this class are almost never observed.

LEMMA B.4 (Chebyshev inequality). Let y be a random variable with zero mean and variance σ^2. Then,

$$P[|y| \geqq \varepsilon] \leqq \frac{\sigma^2}{\varepsilon^2} \tag{B.29}$$

The proof of this lemma can be found in any of the references on probability. This lemma is most useful in conjunction with the Borel-Cantelli lemma and will be employed extensively in the following theorems.

In Chapter 3 we introduced a stochastic process called a martingale. Such a process has an expected value conditioned on knowledge of the past, equal to the most recent value of the process. There is a generalization of this pro-

cess called a *semimartingale* where the conditional expectation is bounded by the latest knowledge of the state.

DEFINITION B.1. A process $x(t)$, $t \in [0,T]$, is called a *semimartingale* if $s < t$ and

$$E[x(t)|\mathscr{F}_s] \geqq x(s) \tag{B.30}$$

where $\mathscr{F}_s$ is the sub σ-field generated by $\{x(\xi);\ \xi \leqq s\}$. The process in the above with the inequality reversed is called a *lower semimartingale*. The following lemma about the semimartingale will be important.

LEMMA B.5. Let $\{x_j,\ 1 \leqq j \leqq n\}$ be a semimartingale and let λ be any real number. Then

$$\lambda P\left\{\max_j x_j \geqq \lambda\right\} \leqq E[|x_n|] \tag{B.31}$$

Proof. Let A be the event

$$A = \left\{\max_j x_j \geqq \lambda\right\} \tag{B.32}$$

Let B_k be the event for which x_k is the first x_j with $x_j \geqq \lambda$ and let B_1 be the event $\{x_1 \geqq \lambda\}$. That is,

$$B_k = \{x_j < \lambda;\ 1 \leqq j < k;\ x_k \geqq \lambda\} \tag{B.33}$$

Furthermore, note that the events B_k are disjoint and that $A = \bigcup_{k=1}^{n} B_k$. Thus, if P is the probability measure for the space on which x_j are defined, we have

$$\int_A x_n\, dP = \sum_k \int_{B_k} x_n\, dP \tag{B.34}$$

and since we have a semimartingale

$$\sum_k \int_{B_k} x_n\, dP \geqq \sum_k \int_{B_k} x_k\, dP \tag{B.35}$$

and since $x_k \geqq \lambda$, we have

$$\int_D |x_n|\, dP \geqq \sum_k \int_{B_k} x_k\, dP \geqq \lambda \sum_k P[B_k] = \lambda P[A] \tag{B.36}$$

which proves the lemma. ∎

This lemma has an immediate extension to continuous semimartingales. That is, if $x(t)$, $t \in [a,b]$ is a separable semimartingale, then for every $\varepsilon > 0$ we have $\varepsilon P[\sup_t |x(t)| \geqq \varepsilon] \leqq E[|x(b)|]$. See Doob [2, p. 353]. Also see Wong [2, p. 51].

We now want to prove the theorem on existence and uniqueness for stochastic differential equations. It is very long and employs the Borel-

Cantelli lemma quite extensively. It in essence states the stochastic analogue of the proof of existence and uniqueness of the differential equation as obtained in Appendix A. Again, it is a sufficiency proof requiring Lipschitz conditions.

THEOREM B.1

Let

$$dx(t) = f(x(t))\ dt + \sigma(x(t))\ dw(t) \tag{B.37}$$

where $t \in (0, T)$, be a stochastic differential equation where both $f(x(t))$ and $\sigma(x(t))$ satisfy the Lipschitz condition

$$|f(x_1) - f(x_2)| \leqq K|x_1 - x_2| \tag{B.38}$$

and

$$|\sigma(x_1) - \sigma(x_2)| \leqq K|x_1 - x_2| \tag{B.39}$$

and $w(t)$ is a normalized scalar Wiener process. Then there exists a unique solution.

Proof. We shall use a Picard iteration proof as was done for the deterministic equation in Chapter 2. We shall follow a four-part proof. First, we bound the mean square difference between two different approximations so that the difference forms a term in a convergent series. We then use this bound, the semimartingale inequality, and the Chebyschev inequality to show that both σ_n and f_n, the Picard approximations, go with probability one to σ and f. This will require the use of the Borel-Cantelli lemma. Third, we show that x_n is a l.i.m. evaluation of $x(t)$. The fourth point is to show that this convergence is uniform in t, and this will again require the Borel-Cantilli lemma. The second part of the proof is that of uniqueness and is relatively straightforward. Let

$$x_n(t) = x(0) + \int_0^t f(x_{n-1}(\xi))\ d\xi + \int_0^t \sigma(x_{n-1}(\xi))\ dw(\xi) \tag{B.40}$$

be a series approximation to the solution where the stochastic integral is interpreted in the Ito sense. Define

$$\Delta_n x(t) = x_n(t) - x_{n-1}(t) \tag{B.41}$$

$$\Delta_n f(x(t)) = f(x_n(t)) - f(x_{n-1}(t)) \tag{B.42}$$

$$\Delta_n \sigma(x(t)) = \sigma(x_n(t)) - \sigma(x_{n-1}(t)) \tag{B.43}$$

Thus, by the Lipschitz condition

$$|\Delta_n f(t)| \leqq K|\Delta_n x(t)| \tag{B.44}$$

$$|\Delta_n \sigma(t)| \leqq K|\Delta_n x(t)| \tag{B.45}$$

Now bound the mean square value of $\Delta_n x(t)$. That is, using (B.10) of Lemma B.2 in the difference form of (B.40), we obtain

$$E[[\Delta_n x(t)]^2]$$

$$\leqq 2E\left[\left|\int_0^t \Delta_{n-1} f(\xi)\ d\xi\right|^2\right] + 2E\left[\left|\int_0^t \Delta_{n-1}\sigma(\xi)\ dw(\xi)\right|^2\right] \tag{B.46}$$

Now let us use the bound given by (B.2) of Lemma B.1, which says

$$E\left[\left|\int_0^t \Delta_n f(\xi)\ d\xi\right|^2\right] \leqq tE\left[\int_0^t (\Delta_n f(\xi))^2\ d\xi\right] \tag{B.47}$$

Using Fubini's theorem, we can interchange the order of integration. Thus,

$$E\left[\left|\int_0^t \Delta_n f(\xi)\ d\xi\right|\right] \leqq T\int_0^t E[(\Delta_n f(\xi))^2]\ d\xi \tag{B.48}$$

Finally, use the Lipschitz condition to obtain

$$E\left[\left|\int_0^t \Delta_n f(\xi)\ d\xi\right|\right] \leqq K^2 T\int_0^t E[(\Delta_n x(\xi))^2]\ d\xi \tag{B.49}$$

We also want to evaluate

$$E\left[\left|\int_0^t \Delta_{n-1}\sigma(\xi)\ dw(\xi)\right|^2\right]$$

$$\leqq E\left[\int_0^t\int_0^t \left|\Delta_{n-1}\sigma(\xi)\right|\left|\Delta_{n-1}\sigma(\eta)\right|\left|dw(\xi)\right|\left|dw(\eta)\right|\right] \tag{B.50}$$

Now bring the expectation inside and recall that $w(t)$ is an independent increment process; we obtain

$$E[|dw(\xi)||dw(\eta)|] = \begin{cases} d\xi; & \xi = \eta \\ o(d\xi); & \eta \neq \xi \end{cases} \tag{B.51}$$

where we assume $dw(\xi)$ is independent of $\Delta_{n-1}\sigma(\xi)$.

Thus, we bound (B.50) as follows

$$E\left[\left|\int_0^t \Delta_{n-1}\sigma(\xi)\ dw(\xi)\right|^2\right] \leqq \int_0^t E[(\Delta_{n-1}\sigma(\xi))^2]\ dt \tag{B.52}$$

Using the Lipschitz condition, we obtain

$$E\left[\left|\int_0^t \Delta_{n-1}\sigma(\xi)\ dw(\xi)\right|^2\right] \leqq K^2\int_0^t E[(\Delta_{n-1}x(\xi))^2]\ dt \tag{B.53}$$

Thus, we obtain a bound on $\Delta_n x(t)$ as follows:

$$E[(\Delta_n x(t))^2] \leqq 2K^2(T+1)\int_0^t E[(\Delta_{n-1}x(\xi))^2]\ d\xi \tag{B.54}$$

Now, as in Appendix A we can obtain a bound by means of this recursive relationship, which is

$$E[(\Delta_n x(t))^2] \leqq \frac{C^n}{n!} \tag{B.55}$$

where C is some positive finite constant. Thus, the Picard iteration is convergent in mean square.

Let us now consider the following probability

$$P\Big[\max_{0\le t\le T}\Big|\int_0^t \Delta_n f(\xi)\ d\xi\Big| \ge 2^{-n}\Big] \tag{B.56}$$

The event

$$\Big|\int_0^t \Delta_n f(\xi)\ d\xi\Big| > 2^{-n} \tag{B.57}$$

is a subset of the event

$$\int_0^t \big|\Delta_n f(\xi)\big| d\xi \ge 2^{-n} \tag{B.58}$$

which is itself a subset of the event

$$\int_0^T K\big|\Delta_n x(\xi)\big| d\xi \ge 2^{-n} \tag{B.59}$$

Thus,

$$P\Big[\max_{0\le t\le T}\Big|\int_0^t \Delta_n f(\xi)\ d\xi\Big| \ge 2^{-n}\Big] \le P\Big[K\int_0^T \big|\Delta_n x(\xi)\big| d\xi \ge 2^{-n}\Big] \tag{B.60}$$

Using the Chebyshev inequality, we obtain

$$P\Big[K\int_0^T \big|\Delta_n x(\xi)\big| d\xi \ge 2^{-n}\Big] \le \frac{K^2 E\Big[\Big(\int_0^T \big|\Delta_n x(\xi)\big| d\xi\Big)^2\Big]}{4^{-n}} \tag{B.61}$$

But, using the bound on the expectation, we obtain

$$P\Big[\max_{0\le t\le T}\Big|\int_0^t \Delta_n f(\xi)\ d\xi\Big| \ge 2^{-n}\Big] \le \frac{4^n T K^2 C^n}{n!} \tag{B.62}$$

which is a general term in a convergent series. Thus, according to the Borel-Cantelli lemma, we know that

$$\max_{0\le t\le T}\Big|\int_0^t \Delta_n f(\xi)\ d\xi\Big| < 2^{-n} \tag{B.63}$$

with probability one for sufficiently large n. This follows from Borel-Cantelli lemma by noting that if we let S_n be the event, then since $\sum P[S_n] < \infty$, we know $P[S] = 0$. This implies $P[S^c] = 1$. But $P[S^c] = P\Big[\bigcap_{n=1}\bigcup_{k=n} S_k\Big]$ $= P\Big[\bigcup_{n=1}\bigcap_{k=n} \bar{S}_k\Big]$. (B.63) follows directly from this observation. Thus,

$$\lim_{n\to\infty}\int_0^t f_n(\xi)\ d\xi \to \int_0^t f(x(\xi))\ d\xi \tag{B.64}$$

Now let us obtain a similar result for the $\sigma(x(t))$ portion. We know that

$$\int_0^t \Delta\sigma(\xi)\ dw(\xi) \tag{B.65}$$

is a martingale. It is a simple matter to show that the square of the above quantity satisfies the semimartingale property. We shall now employ the semimartingale inequality that was presented in (B.31). Now

$$P\Big[\max_{0\le t\le T}\Big|\int_0^t \Delta_n\sigma(\xi)\ d\xi\Big| \ge 2^{-n}\Big] \le 4^n E\Big[\Big(\int_0^T \Delta_n\sigma(\xi)\ d\xi\Big)^2\Big] \tag{B.66}$$

by the inequality. What we have done is to let

$$x_j = \Big|\int_0^t \Delta_n\sigma(\xi)\ d\xi\Big|^2 \tag{B.67}$$

and $k = 4^{-n}$ and recall that

$$P\Big[\max_{0\le t\le T}\Big|\int_0^t \Delta_n\sigma(\xi)\ d\xi\Big| \ge 2^{-n}\Big] = P\Big[\max_{0\le t\le T}\Big|\int_0^t \Delta_n\sigma(\xi)\ d\xi\Big|^2 \ge 4^{-n}\Big] \tag{B.68}$$

and directly apply (B.31) of Lemma B.5. Taking the expectation yields for the right-hand side of (B.66)

$$4^n E\Big[\Big(\int_0^T \Delta_n\sigma(\xi)\ d\xi\Big)^2\Big] = 4^n\int_0^T E[(\Delta_n\sigma(\xi))^2]\ d\xi \tag{B.69}$$

Then, using the Lipschitz inequality, we have

$$4^n E\Big[\Big(\int_0^T \Delta_n\sigma(\xi)\ d\xi\Big)^2\Big] \le 4K^2\int_0^T E[(\Delta_n x(\xi))^2]\ d\xi \tag{B.70}$$

and finally, using the bound on the second moment of $\Delta_n x(\xi)$, we have

$$P\Big[\max_{0\le t\le T}\Big|\int_0^t \Delta_n\sigma(\xi)\ d\xi\Big| \ge 2^{-n}\Big] \le \frac{4^n K^2 C^n}{n!} \tag{B.71}$$

Thus, again using the Borel-Cantelli lemma, we see that

$$\max_{0\le t\le T}\Big|\int_0^t \Delta_n\sigma(\xi)\ dw(\xi)\Big| \le 2^{-n} \tag{B.72}$$

with probability one for a sufficiently large n. Therefore, this implies

$$\int_0^t \sigma(x_n(\xi))\ dw(\xi) \to \int_0^t \sigma(x(\xi))\ dw(\xi) \tag{B.73}$$

with probability one and uniformly on t.

We now want to prove the third fact, that of l.i.m. convergence. Now, for each $n > m$ and each t,

$$E[(x_n(t) - x_m(t))^2] = E\Big[\Big(\sum_{j=m+1}^{n} \Delta_j x(t)\Big)^2\Big] \le \frac{C_2}{2^m} \tag{B.74}$$

where C_2 is some finite positive constant. Thus, as $n \to \infty$, we have

$$E[(x(t) - x_m(t))^2] \le \frac{C_2}{2^n} \tag{B.75}$$

and thus the process is a limit in the mean equivalent.

It is a simple matter to show that

$$|x_n(t) - x_{n-1}(t)| \to 0 \tag{B.76}$$

with probability one and that it is uniform (Ito [2], pp. 195–196). We already know that

$$P[|x_n(t) - x_{n-1}(t)| > 2^{-n}] \tag{B.77}$$

can be obtained by means of the semimartingale inequality. That is,

$$P[|x_n(t) - x_{n-1}(t)| > 2^{-n}] < \frac{C2^n}{n!} \tag{B.78}$$

where C is a positive constant. Again the Borel-Cantelli lemma yields the fact that

$$|x_n(t) - x_{n-1}(t)| \to 0 \tag{B.79}$$

with probability one.

Now, to demonstrate uniqueness, we shall show that if there exist two solutions, then their mean square differences converge and thus produce a contradiction. Let $x(t)$ and $y(t)$ be two solutions. Then

$$\begin{aligned} E[(x(t) - y(t))^2] \leqq\ & 2E[(\int_0^t [f(x(\xi)) - f(y(\xi))]\ d\xi)^2] \\ & + 2E[(\int_0^t [\sigma(x(\xi)) - \sigma(y(\xi))]\ d\xi)^2] \end{aligned} \tag{B.80}$$

Again, using the now familiar inequalities,

$$\begin{aligned} E[(x(t) - y(t))^2] \leqq\ & 2\int_0^t E[[f(x(\xi)) - f(y(\xi))]^2]\ d\xi \\ & + 2\int_0^t E[[\sigma(x(\xi)) - \sigma(y(\xi))]^2]\ d\xi \end{aligned} \tag{B.81}$$

Using the Lipschitz conditions, we obtain

$$E[(x(t) - y(t))^2] \leqq 2K^2(1 + T)\int_0^t E[(x(\xi) - y(\xi))^2]\ d\xi \tag{B.82}$$

Then let $r(t)$ be

$$r(t) = E[(x(t) - y(t))^2] \tag{B.83}$$

Thus,

$$r(t) \leqq 2K(1 + T)\int_0^t r(\xi)\ d\xi \tag{B.84}$$

Now let us use (B.84) to bound $r(t)$ on the right; that is,

$$r(t) \leqq R\int_0^t dt\ C\int_0^{t_1} r(t_1)\ dt_1 \tag{B.85}$$

where we let $R = 2K(1 + T)$. Do this again n times

$$r(t) \leqq R^n \int_0^t dt \int_0^{t_1} dt_1 \cdots \int_0^{t_{n-1}} dt_{n-1} r(t_{n-1}) \tag{B.86}$$

But recall that $r(t_{n-1})$ is bounded by C. Thus,

$$r(t) \leqq R^n C \int_0^t dt \int_0^{t_1} dt_1 \cdots \int_0^{t_{n-1}} dt_{n-1} \tag{B.87}$$

the integral on the right is merely $t^n/n!$. Therefore,

$$r(t) \leqq R^n C \frac{t^n}{n!} \tag{B.88}$$

Thus,

$$r(t) \leqq \frac{(2K^2(1+T)^n)}{n!} t^n C \tag{B.89}$$

where C is a finite constant and n is a constant that we used in the integral bounds. As we let this $n \to \infty$, we find

$$r(t) \to 0$$

Thus, they are mean-square equivalent and therefore unique in this sense. ∎

Extensions of the previous theorem to the vector case are given in Gikhman and Skorokhod (pp. 391–403). Similarly, for the case where $w(t)$ is not just a Wiener process but an independent increment process of a more general nature (e.g., Poisson process), Skorokhod (pp. 42–73) discusses the necessary extensions.

APPENDIX C

STABILITY OF THE DISCRETE-TIME ESTIMATOR

The structure of the discrete-time linear filter was first developed in Section 4.3 and later elaborated on in Section 6.4. In that latter section we discussed the question of divergence and stability of the unforced discrete-time system. Specifically, we found that certain questions relating to divergence could be answered by considering the stability of the discrete-time estimate equation. However, in Section 2.4 we developed just such a set of conditions in the Lyapanov theory that could be used to answer such a question. In this appendix we shall use that theory and apply it to the discrete-time estimate equation. In so doing we will develop the discrete-time observability and controllability conditions for stochastic systems as discussed in Section 6.4.

Our approach will be to first consider a discrete-time system and develop a suboptimal filter using the maximum-likelihood technique. We will then use this system to obtain an upper bound on the covariance function in terms of the stochastic controllability and observability matrices. Then, considering a similar system, we show that we can also lower-bound the covariance matrix. These bounds are given in terms of the quadratic forms they generate. We finally consider a specific quadratic form for the system, specifically $\mathbf{x}^T(k)\mathbf{P}^{-1}(k)\mathbf{x}(k)$, and show that it is a time-decreasing form and thus a Lyapanov function. Thus, having found a Laypanov function for the system we can evoke the results of Section 2.4 and use it to show u.a.s.i.l. for the deterministic filter.

This development and analysis was discussed first by Kalman [3] in 1963 and by others, notably by Deyst and Price, by Bucy and Joseph, and—most recently—by Bucy [2–3]. Our method of analysis follows Bucy [3] except that it is for the discrete-time case.

DEFINITION C. 1. Let $\mathbf{A}$ be an $n \times n$ positive definite matrix. Let $\mathbf{x}$ be any $n \times 1$ vector. Then we say

$$\alpha \mathbf{I} \leqq \mathbf{A} \leqq \beta \mathbf{I} \tag{C.1}$$

if and only if for any $\mathbf{x}$

$$\alpha \mathbf{x}^T\mathbf{x} \leqq \mathbf{x}^T\mathbf{A}\mathbf{x} \leqq \beta \mathbf{x}^T\mathbf{x} \tag{C.2}$$

(Note that $\mathbf{x}$ can be normalized to unit length.)

The following theorem will now be proved. We first state it in its totality, then prove several interim lemmas and theorems, and finally prove this theorem.

THEOREM C.1.

Let $\hat{\mathbf{x}}(k+1)$ be the estimate of the state of the system $\mathbf{x}(k+1)$, where

$$\mathbf{x}(k+1) = \boldsymbol{\Phi}(k+1, k)\mathbf{x}(k) + \mathbf{u}(k) \tag{C.3}$$

$$\mathbf{z}(k) = \mathbf{C}(k)\mathbf{x}(k) + \mathbf{w}(k) \tag{C.4}$$

and $\mathbf{u}(k)$ and $\mathbf{w}(k)$ are independent white Gaussian sequences with covariances $\mathbf{Q}(k)$ and $\mathbf{R}(k)$, respectively. The estimate $\hat{\mathbf{x}}(k+1)$ is given by

$$\begin{aligned}\hat{\mathbf{x}}(k+1) = \boldsymbol{\Phi}(k+1, k)\hat{\mathbf{x}}(k) + \mathbf{K}(k+1)[\mathbf{z}(k+1) \\ - \mathbf{C}(k+1)\boldsymbol{\Phi}(k+1, k)\hat{\mathbf{x}}(k)]\end{aligned} \tag{C.5}$$

where $\mathbf{K}(k+1)$ is given in Section 6.4. The function

$$V_P(\mathbf{x}_*(k), k) = \mathbf{x}_*^T(k)\mathbf{P}^{-1}(k)\mathbf{x}_*(k) \tag{C.6}$$

where $\mathbf{P}^{-1}(k)$ is the covariance matrix of the discrete-time estimate equation, is a Lyapanov function for the system

$$\hat{\mathbf{x}}_*(k+1) = [\mathbf{I} - \mathbf{C}(k+1)\,\mathbf{K}(k+1)]\boldsymbol{\Phi}(k+1, k)\hat{\mathbf{x}}_*(k) \tag{C.7}$$

and this system is u.a.s.i.l. if the matrices

$$\mathbf{M}(N, N-n) = \sum_{k=N-n}^{N} \boldsymbol{\Phi}^T(k, N)\mathbf{C}^T(k)\mathbf{R}^{-1}(k)\mathbf{C}(k)\boldsymbol{\Phi}(k, N) \tag{C.8}$$

and

$$\mathbf{W}(N, N-n) = \sum_{k=N-n}^{N-1} \boldsymbol{\Phi}(N, k+1)\mathbf{Q}(k)\boldsymbol{\Phi}^T(N, k+1) \tag{C.9}$$

satisfy

$$\gamma \mathbf{I} \leqq \mathbf{W}(N, N-n) \leqq \delta \mathbf{I}; \qquad \gamma, \delta > 0 \tag{C.10}$$

$$\alpha \mathbf{I} \leqq \mathbf{M}(N, N-n) \leqq \beta \mathbf{I}; \qquad \alpha, \beta > 0 \tag{C.11}$$

for some N and $n < N - 1$ greater than 0.

The object then of the proof is to show that the function $V_P(\mathbf{x}_*(k), k)$ is a Lyapanov function. To do this, we must first evaluate its bounds and then consider its rate of change. We now consider what is called the maximum-likelihood estimate of the state. The maximum-likelihood estimate is that

value of x at time N that maximizes the joint probability density of the measurements conditioned on $\mathbf{x}(N)$; that is, if there are $n+1$ measurements, $\mathbf{z}(N),\cdots,\mathbf{z}(N-n)$, and $p_z(\mathbf{z}(N),\cdots,\mathbf{z}(N-n)|\mathbf{x}(N))$ is this conditional density, then $\mathbf{x}_{ML}(N)$, the maximum-likelihood estimate, is that value of $\mathbf{x}(N)$ that maximizes this density. This type of estimate uses no a priori knowledge of the stochastic nature of the state and thus performs in a suboptimum fahsion to the MMSE estimate.

THEOREM C.2

Consider the discrete-time system given by

$$\mathbf{x}(k+1) = \Phi(k+1,k)\mathbf{x}(k) \tag{C.12}$$

Let $\hat{\mathbf{x}}(N)$ be the maximum likelihood estimate of the state $\mathbf{x}(N)$, given measurements $\mathbf{z}(N),\cdots,\mathbf{z}(N-n)$ where for any k

$$\mathbf{z}(k) = \mathbf{C}(k)\mathbf{x}(k) + \mathbf{w}(k) \tag{C.13}$$

Then, if the system is stochastically observable.

$$\hat{\mathbf{x}}(N) = \mathbf{M}^{-1}(N, N-n)\sum_{k=N-n}^{N}\Phi^T(k,N)\mathbf{C}^T(k)\mathbf{R}^{-1}(k)\mathbf{z}(k) \tag{C.14}$$

Note that if $\mathbf{M}(N,N-n)$ is singular, then we can use the pseudoinverse of a matrix (see Problem 4.18) and still define an estimate. This is discussed in Kalman [3].

Proof. The joint probability desntiy of $\mathbf{z}(N),\cdots,\mathbf{z}(N-n)$ is given by

$$p_z(\mathbf{z}(N),\cdots,\mathbf{z}(N-n)|\mathbf{x}(N)) = C\exp\{-\sum_{k=N-n}^{N}[\mathbf{z}(k) - \mathbf{C}(k)\Phi(k,N)\mathbf{x}(N)]^T\mathbf{R}^{-1}(k)[\mathbf{z}(k) - \mathbf{C}(k)\Phi(k,N)\mathbf{x}(N)]\} \tag{C.15}$$

Now it is easily shown that the exponent can be written as

$$-\mathbf{x}^T(N)\mathbf{M}(N,N-n)\mathbf{x}(N) + 2\mathbf{x}^T(N)\sum_{k=N-n}^{N}\Phi^T(k,N)\mathbf{C}^T(k)\mathbf{R}^{-1}(k)\mathbf{z}(k) - \sum_{k=N-n}^{N}\mathbf{z}^T(k)\mathbf{R}^{-1}(k)\mathbf{z}(k) \tag{C.16}$$

Completing the square, this is equivalent to the quadratic form

$$-(\mathbf{x}(N) - \mathbf{M}^{-1}(N,N-n)\sum_{k=N-n}^{N}\Phi^T(k,N)\mathbf{C}^T(k)\mathbf{R}^{-1}(k)\mathbf{z}(k))^T \mathbf{M}(N,N-n)(\mathbf{x}(N) - \mathbf{M}^{-1}(N,N-n)\sum_{k=N-n}^{N}\Phi^T(k,N)\mathbf{C}^T(k)\mathbf{R}^{-1}(k)\mathbf{z}(k)) \tag{C.17}$$

Now the minimum of this exponent is obtained by equating it to zero, which implies the optimum value of $\mathbf{x}(N)$ is as given in the theorem. ∎

We now want to consider the use of this estimator for the system given by (C.3) and (C.4). Clearly, the optimum estimator (C.5), which has covariance $\mathbf{P}(k)$, will perform better in estimating $\mathbf{x}(N)$ than will estimator (C.14). The exact difference is expressed in the following lemma.

LEMMA C.1. Let $\bar{\mathbf{x}}(N)$ be the estimate of (C.5) for $\mathbf{x}(N)$ given by (C.3). Let $\tilde{\mathbf{x}}(N)$ be the estimate given $\mathbf{z}(N),\cdots,\mathbf{z}(N-n)$ and let $\mathbf{P}(N)$ be the resulting covariance. Then,

$$\mathbf{P}(N) \leqq \mathbf{M}^{-1}(N,N-n) + \mathbf{M}^{-1}(N,N-n)\sum_{k=N-n}^{N}\sum_{q=N-n}^{N}\boldsymbol{\Phi}^T(k,N)\mathbf{C}^T(k)\mathbf{R}^{-1}(k)$$
$$\mathbf{C}(k)\boldsymbol{\Phi}(k,N)\mathbf{W}(N,\max(N-q,k))\boldsymbol{\Phi}^T(q,N)\mathbf{C}^T(q)\mathbf{R}^{-1}(q)\mathbf{C}(q)\boldsymbol{\Phi}(q,N)]\mathbf{M}^{-1}(N,N-n) \tag{C.18}$$

Proof. Now $\mathbf{x}(N)$ is given by

$$\mathbf{x}(N) = \boldsymbol{\Phi}(N, N-1)\mathbf{x}(N-1) + \mathbf{n}(N-1) \tag{C.19}$$

It can then be shown that for any k we have

$$\mathbf{x}(k) = \boldsymbol{\Phi}(k, N)\mathbf{x}(N) + \sum_{j=k}^{N-1}\boldsymbol{\Phi}(k, j+1)\mathbf{n}(j) \tag{C.20}$$

The maximum-likelihood estimate is given by

$$\bar{\mathbf{x}}(N) = \mathbf{M}^{-1}(N, N-n)\sum_{k=N-n}^{N}\boldsymbol{\Phi}^T(k, N)\mathbf{C}^T(k)\mathbf{R}^{-1}(k)\mathbf{z}(k) \tag{C.21}$$

But

$$\mathbf{z}(k) = \mathbf{C}(k)\mathbf{x}(k) + \mathbf{w}(k) \tag{C.22}$$

Using (C.20) in (C.22) and then using that in (C.21), we obtain,

$$\mathbf{x}(N) = \mathbf{M}^{-1}(N, N-n)\sum_{k=N-n}^{N}\boldsymbol{\Phi}^T(k, N)\mathbf{C}^T(k)\mathbf{R}^{-1}(k)$$
$$[\mathbf{C}(k)\boldsymbol{\Phi}(k, N)\mathbf{x}(N) + \mathbf{C}(k)\sum_{j=k}^{N-1}\boldsymbol{\Phi}(k, j+1)\mathbf{n}(j) + \mathbf{w}(k)] \tag{C.23}$$

Thus, the maximum-likelihood error $\bar{\mathbf{x}}_{ML}(N)$ is given by

$$\bar{\mathbf{x}}(N) = \mathbf{M}^{-1}(N, N-n)\sum_{k=N-n}^{N}\boldsymbol{\Phi}^T(k, N)\mathbf{C}^T(k)\mathbf{R}^{-1}(k)$$
$$[\mathbf{C}(k)\sum_{j=k}^{N-1}\boldsymbol{\Phi}(k, j+1)\mathbf{n}(j) + \mathbf{w}(k)] \tag{C.24}$$

Then defining $\bar{\mathbf{P}}(N)$ as

$$\bar{\mathbf{P}}(N) = E[\bar{\mathbf{x}}(N)\bar{\mathbf{x}}^T(N)] \tag{C.25}$$

and taking the expectations in (C.24), we find that

$$\bar{\mathbf{P}}(N) = \mathbf{M}^{-1}(N, N-n) \sum_{k=N-n}^{N} \sum_{q=N-n}^{N} \boldsymbol{\Phi}^T(k, N)\mathbf{C}^T(k)\mathbf{R}^{-1}(k)\mathbf{C}(k)\boldsymbol{\Phi}(k, N)$$
$$\mathbf{W}(N, \max(N-q,k))\boldsymbol{\Phi}^T(q,N)\mathbf{C}^T(q)\mathbf{R}^{-1}(q)\mathbf{C}(q)\boldsymbol{\Phi}(q,N)\mathbf{M}^{-1}(N,N-n)$$
$$+ \mathbf{M}^{-1}(N,N-n) \tag{C.26}$$

Now since this is the covariance, using the maximum-likelihood estimator which is suboptimum. we have

$$\mathbf{P}(N) \leqq \bar{\mathbf{P}}(N) \tag{C.27}$$

which is the desired result. ∎

Before proceeding with the bounding procedure of $\mathbf{P}(N)$, we need to prove three lemmas that will be essential in reducing the bound in the above lemma to a useful form.

LEMMA C.2. Let $\mathbf{A}$ be a positive definite symmetric matrix. Then for all vectors $\boldsymbol{\alpha}$,

$$\boldsymbol{\alpha}^T\mathbf{A}\boldsymbol{\alpha} \leqq \boldsymbol{\alpha}^T\boldsymbol{\alpha}\mathrm{tr}\mathbf{A} \tag{C.28}$$

Proof. Let $\mathbf{P}'$ be a unitary matrix $\mathbf{P}'^T\mathbf{P}' = \mathbf{I}$ such that

$$\mathbf{P}'^T\mathbf{A}\mathbf{P}' = \boldsymbol{\Lambda} \tag{C.29}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix (see Moore, p. 254). Now let $\boldsymbol{\beta} = \mathbf{P}'^{-1}\boldsymbol{\alpha}$ (see Hildebrand, pp. 36–39) so that

$$\boldsymbol{\beta}^T\boldsymbol{\Lambda}\boldsymbol{\beta} = \sum_{i=1}^{n} \lambda_i\beta_i^2 \tag{C.30}$$

where λ_i are the eigenvalues of $\boldsymbol{\Lambda}$ and $\mathbf{A}$. Thus,

$$\begin{aligned}\boldsymbol{\beta}^T\boldsymbol{\Lambda}\boldsymbol{\beta} &\leqq \sum \beta_i^2 \max[\lambda_k] \\ &= \boldsymbol{\beta}^T\boldsymbol{\beta} \max[\lambda_k] \\ &\leqq \boldsymbol{\beta}^T\boldsymbol{\beta}\,\mathrm{tr}[\boldsymbol{\Lambda}] = \boldsymbol{\beta}^T\boldsymbol{\beta}\,\mathrm{tr}[\mathbf{A}]\end{aligned} \tag{C.31}$$

Since $\mathrm{tr}\,\mathbf{A} = \sum_{i=1}^{n} \lambda_i$ and $\lambda_i > 0$ for all i.

Now, since $\boldsymbol{\beta}^T\boldsymbol{\beta}$ equals $\boldsymbol{\alpha}^T\boldsymbol{\alpha}$, we have

$$\boldsymbol{\alpha}^T\mathbf{A}\boldsymbol{\alpha} = \boldsymbol{\beta}^T\boldsymbol{\Lambda}\boldsymbol{\beta} \leqq \boldsymbol{\alpha}^T\boldsymbol{\alpha}\,\mathrm{tr}\mathbf{A} \tag{C.32}$$

which proves the lemma. ∎

LEMMA C.3. Let $\mathbf{W}$ and $\mathbf{C}$ be positive definite matrices. Then,

$$\boldsymbol{\alpha}^T\mathbf{W}^{-1}\boldsymbol{\alpha} \leqq \boldsymbol{\alpha}^T\mathbf{C}\boldsymbol{\alpha}\,\mathrm{tr}\mathbf{W}^{-1}\mathbf{C}^{-1} \tag{C.33}$$

Proof. There exists an invertible $\mathbf{P}'$ such that

$$\mathbf{P}'^T\mathbf{C}\mathbf{P}' = \mathbf{I} \tag{C.34}$$

(see Moore, p. 262, Theorem 4.2). Now premultiply by $(\mathbf{P}^T)^{-1}$ to yield $\mathbf{CP}' = (\mathbf{P}^T)^{-1}$. Then postmultiply to yield $\mathbf{CP}'\mathbf{P}^{-1} = \mathbf{I}$ or equivalently $\mathbf{P}'\mathbf{P}^{-1}$ equals $\mathbf{C}^{-1}$. Now

$$\alpha^T\mathbf{W}^{-1}\alpha = \alpha^T\mathbf{W}^{-1}\mathbf{C}^{-1}\mathbf{C}\alpha \tag{C.35}$$

Choose $\beta = \mathbf{Q}\alpha$ such that $\alpha^T\mathbf{Q}^T\mathbf{Q}\alpha$ equals $\alpha^T\mathbf{C}\alpha$. This means that

$$\mathbf{Q}^T\mathbf{Q} = \mathbf{C} \tag{C.36}$$

or

$$\mathbf{C}^{-1} = \mathbf{Q}^{-1}(\mathbf{Q}^T)^{-1} \tag{C.37}$$

which means $\mathbf{Q} = \mathbf{P}'$. Thus,

$$\alpha^T\mathbf{W}^{-1}\alpha = \beta^T(\mathbf{Q}^T)^{-1}\mathbf{W}^{-1}\mathbf{C}^{-1}\mathbf{C}\mathbf{Q}^{-1}\beta \tag{C.38}$$

Now, using the previous lemma, we have

$$\alpha^T\mathbf{W}^{-1}\alpha \leqq \beta^T\beta\,\mathrm{tr}[(\mathbf{Q}^{-1})^T\mathbf{W}^{-1}\mathbf{C}^{-1}\mathbf{C}\mathbf{Q}^{-1}] \tag{C.39}$$

But since tr $\mathbf{AB}$ = tr $\mathbf{BA}$ (J. T. Moore, p. 134), we have

$$\alpha^T\mathbf{W}^{-1}\alpha \leqq \beta^T\beta\,\mathrm{tr}[\mathbf{W}^{-1}\mathbf{C}^{-1}\mathbf{C}\mathbf{Q}^{-1}(\mathbf{Q}^{-1})^T] \tag{C.40}$$

But

$$\mathbf{C}\mathbf{Q}^{-1}(\mathbf{Q}^{-1})^T = \mathbf{C}\mathbf{P}^{-1}(\mathbf{P}^{-1})^T = \mathbf{I} \tag{C.41}$$

also

$$\beta^T\beta = \alpha^T\mathbf{Q}\mathbf{Q}^T\alpha = \alpha^T\mathbf{C}\alpha \tag{C.42}$$

Thus,

$$\alpha^T\mathbf{W}^{-1}\alpha \leqq \alpha^T\mathbf{C}\alpha\,\mathrm{tr}[\mathbf{W}^{-1}\mathbf{C}^{-1}] \tag{C.43}$$

which proves the lemma. ∎

LEMMA C.4. Let $\mathbf{K}(i, j)$ be an $n \times n$ positive definite matrix and let $\mathbf{g}(i)$ be an $n \times 1$ vector. Then for any $N \geqq 1$,

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{g}^T(i)\mathbf{K}(i,j)\mathbf{g}(j) \leqq \sum_{i=1}^{N}\mathrm{tr}\,\mathbf{K}(i,i)\sum_{i=1}^{N}\mathbf{g}^T(i)\mathbf{g}(i) \tag{C.44}$$

Proof. By an extension of the Karhunen-Loeve expansion to discrete-time vector-valued processes (see Kelly and Root), we can show that there exists a decomposition of an $\mathbf{g}(i) \in \mathbf{R}^n$ such that

$$\mathbf{g}(i) = \sum_{k=1}^{N}\mathbf{g}_k\phi_k(i) \tag{C.45}$$

where $\phi(i) \in \mathbf{R}^n$ and they satisfy

$$\lambda\phi_\ell(i) = \sum_{j=1}^{N}\mathbf{K}(i,j)\phi_\ell(j) \tag{C.46}$$

Furthermore,

$$\mathbf{K}(i,j) = \sum_{\ell=1}^{N} \lambda_\ell \boldsymbol{\phi}_\ell(i)\boldsymbol{\phi}_\ell^T(j) \tag{C.47}$$

where with

$$\sum_{\ell=1}^{N} \boldsymbol{\phi}_\ell^T(i)\boldsymbol{\phi}_k(i) = \delta_{k\ell} \tag{C.48}$$

Using (C.47) and (C.45) in the left-hand side of (C.44), we obtain

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\Big(\sum_{\ell=1}^{N} g_\ell \boldsymbol{\phi}_\ell^T(i)\Big)\Big(\sum_{n=1}^{N}\lambda_n\boldsymbol{\phi}_n(i)\boldsymbol{\phi}_n^T(j)\Big)\Big(\sum_{m=1}^{N} g_m\boldsymbol{\phi}_n(j)\Big)$$
$$= \sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{g}^T(i)\mathbf{K}(i,j)\mathbf{g}(j) \tag{C.49}$$

Using (C.48), this can be shown to equal

$$\sum_{\ell=1}^{N} g_\ell^2\lambda_\ell \tag{C.50}$$

Now, since $\mathbf{K}(i,j)$ is positive definite, λ_i is greater than zero, and since g_i is real, g_i^2 is greater than zero. Thus,

$$\sum_{\ell=1}^{N} g_\ell^2\lambda_\ell \leqq \Big(\sum_{\ell=1}^{N} g_\ell^2\Big)\Big(\sum_{\ell=1}^{N}\lambda_\ell\Big) \tag{C.51}$$

But

$$\sum_{i=1}^{N}\mathbf{g}^T(i)\mathbf{g}(i) = \sum_{i=1}^{N}\Big(\sum_{j=1}^{N} g_j\boldsymbol{\phi}_j^T(i)\Big)\Big(\sum_{k=1}^{N} g_k\boldsymbol{\phi}_k(i)\Big)$$
$$= \sum_{j=1}^{N} g_j^2 \tag{C.52}$$

Likewise,

$$\mathrm{tr}\mathbf{K}(i,i) = \sum_{j=1}^{N}\lambda_j\,\mathrm{tr}\boldsymbol{\phi}_j(i)\boldsymbol{\phi}_j^T(i)$$
$$= \sum_{j=1}^{N}\lambda_j\boldsymbol{\phi}_j^T(i)\boldsymbol{\phi}_j(i) \tag{C.53}$$

Now sum over i to obtain

$$\sum_{i=1}^{N}\mathrm{tr}\mathbf{K}(i,i) = \sum_{i=1}^{N}\Big(\sum_{j=1}^{N}\lambda_j\boldsymbol{\phi}_j^T(i)\boldsymbol{\phi}_j(i)\Big) = \sum_{j=1}^{N}\lambda_j \tag{C.54}$$

Then, using (C.52) and (C.54) in (C.51), we obtain the desired bound. ∎

We can now use the previous lemmas to prove the following theorem.

THEOREM C.3

Let $\mathbf{P}(N)$ be the covariance matrix of the optimal MMSE filter at time N, given $n+1$ measurements. Then,

$$\mathbf{P}(N) \leqq \mathbf{M}^{-1}(N, N-n) + A\mathbf{W}(N, N-n) \tag{C.55}$$

where

$$A = n^2 \frac{\beta\delta}{\alpha\delta} \tag{C.56}$$

Proof. Let $\mathbf{K}(i, \ell)$ equal

$$\mathbf{K}(i, \ell) = \mathbf{R}^{-1/2}(i)\mathbf{C}(i)\boldsymbol{\Phi}(i, N)W(N, \max(N-i, \ell)) \\ \boldsymbol{\Phi}^T(\ell, N)\mathbf{C}^T(\ell)\mathbf{R}^{-1/2}(\ell) \tag{C.57}$$

and let $\mathbf{g}(i)$ be given by

$$\mathbf{g}(i) = \mathbf{R}^{-1/2}(i)C(i)\boldsymbol{\Phi}(i, N)\beta \tag{C.58}$$

where β is the $n \times 1$ vector given by

$$\beta = \mathbf{M}^{-1}(N, N-n)\mathbf{x} \tag{C.59}$$

where $\mathbf{x}$ is arbitrary. Then, using the previous lemma, we have

$$\sum_{i=N-n}^{N} \sum_{j=N-n}^{N} \mathbf{g}^T(i)\mathbf{K}(i, j)\mathbf{g}(j) \leqq \sum_{i=N-n}^{N} \mathrm{tr}\mathbf{K}(i, i) \sum_{i=N-n}^{N} \mathbf{g}^T(i)\mathbf{g}(i) \tag{C.60}$$

But, clearly, (C.60) is a bound for the second expression for the bound in (C.17). Now, using (C.58) and (C.59), we obtain

$$\sum_{i=N-n}^{N} \mathbf{g}^T(i)\mathbf{g}(i) = \mathbf{x}^T\mathbf{M}^{-1}(N, N-n)\mathbf{x} \tag{C.61}$$

Similarly,

$$\mathrm{tr}\mathbf{K}(i, i) = \mathrm{tr}[\boldsymbol{\Phi}^T(\ell, N)\mathbf{C}^T(\ell)\mathbf{R}^{-1}(i)\mathbf{C}(i)\boldsymbol{\Phi}(i, N)\mathbf{W}(N, i)] \tag{C.62}$$

But $\mathbf{W}(N, i) \leqq \delta\,\mathbf{I}$ for any i, so that

$$\mathrm{tr}\mathbf{K}(i, i) \leqq \mathrm{tr}\boldsymbol{\Phi}^T(i, N)\mathbf{C}^T(i)\mathbf{R}^{-1}(i)\mathbf{C}(i)\boldsymbol{\Phi}(i, N)\delta$$

But, since $\mathrm{tr}\,(\mathbf{A} + \mathbf{B}) = \mathrm{tr}\,\mathbf{A} + \mathrm{tr}\,\mathbf{B}$, we have

$$\sum_{i=N-n}^{N} \mathrm{tr}\mathbf{K}(i, i) \leqq \delta\,\mathrm{tr} \sum_{i=N-n}^{N} \boldsymbol{\Phi}^T(i, N)\mathbf{C}^T(i)\mathbf{R}^{-1}(i)\mathbf{C}(i)\boldsymbol{\Phi}(i, N) \tag{C.63}$$

And using the bound on $\mathbf{M}(N, N-n)$, we have

$$\sum_{i=N-n}^{N} \mathrm{tr}\mathbf{K}(i, i) \leqq n\beta\delta \tag{C.64}$$

Now from Lemma C.3 we have

$$\mathbf{x}^T\mathbf{M}^{-1}(N, N-n)\,\mathbf{x} \leqq \mathbf{x}^T\mathbf{W}(N, N-n)\mathbf{x}\,\mathrm{tr}\mathbf{W}^{-1}(N, N-n)\mathbf{M}^{-1}(N, N-n) \tag{C.65}$$

But $\mathbf{W}^{-1}(N, N-n) \leqq 1/\gamma\,\mathbf{I}$ and $\mathbf{M}^{-1}(N, N-n) \leqq 1/\alpha\,\mathbf{I}$, so that

$$\mathrm{tr}\mathbf{W}^{-1}(N, N-n)\mathbf{M}^{-1}(N, N-n) \leqq \frac{n}{\gamma\alpha} \tag{C.66}$$

Thus,

$$\sum_{j=N-n}^{N} \sum_{i=N-n}^{N} \mathbf{g}^T(i)\mathbf{K}(i, \ell)\mathbf{g}(\ell) \leqq n^2 \frac{\beta\delta}{\gamma\alpha} \mathbf{x}^T\mathbf{W}(N, N-n)\mathbf{x} \tag{C.67}$$

which implies as a result of Lemma C.1 that

$$\mathbf{P}(N) \leqq \mathbf{M}^{-1}(N, N-n) + A\mathbf{W}(N, N-n) \tag{C.68}$$

where A equals $n^2\,\beta\delta/\gamma\alpha$. ∎

This provides us with an upper bound for the covariance function $\mathbf{P}(N)$. We now proceed to obtain a lower bound by transforming the system of equations.

From Chapter 4 (see Problem 4.11) we know that the discrete-time filter is given by the following three equations:

$$\begin{aligned}\hat{\mathbf{x}}(k+1) &= \mathbf{P}(k+1)\mathbf{M}^{-1}(k+1)\hat{\mathbf{x}}(k) \\ &\quad + \mathbf{P}(k+1)\mathbf{C}^T(k+1)\mathbf{R}^{-1}(k+1)\mathbf{z}(k+1)\end{aligned} \tag{C.69}$$

$$\mathbf{P}(k+1) = [\mathbf{M}^{-1}(k+1) + \mathbf{C}^T(k+1)\mathbf{R}^{-1}(k+1)\mathbf{C}(k+1)]^{-1} \tag{C.70}$$

$$\mathbf{M}(k+1) = \boldsymbol{\Phi}(k+1, k)\mathbf{P}(k)\boldsymbol{\Phi}^T(k+1, k) + \mathbf{Q}(k) \tag{C.71}$$

This is for the system

$$\mathbf{x}(k+1) = \boldsymbol{\Phi}(k+1, k)\mathbf{x}(k) + \mu(k) \tag{C.72}$$

$$\mathbf{z}(k+1) = \mathbf{C}(k+1)\mathbf{x}(k+1) + \mathbf{w}(k+1) \tag{C.73}$$

Now let $\tilde{\mathbf{P}}(k+1)$ equal $\mathbf{P}^{-1}(k+1)$ and $\tilde{\mathbf{M}}(k+1)$ equal $\mathbf{M}^{-1}(k+1)$. Then we immediately have

$$\tilde{\mathbf{P}}(k+1) = \tilde{\mathbf{M}}(k+1) + \mathbf{C}^T(k+1)\mathbf{R}^{-1}(k+1)\mathbf{C}(k+1) \tag{C.74}$$

Define the matrix $\mathbf{M}_1(k+1)$ as

$$\mathbf{M}_1(k+1) = (\boldsymbol{\Phi}^T(k+1, k))^{-1}\mathbf{P}(k)\boldsymbol{\Phi}^{-1}(k+1, k) \tag{C.75}$$

Then $\tilde{\mathbf{M}}(k+1)$ satisfies

$$\tilde{\mathbf{M}}(k+1) = [\mathbf{M}_1^{-1}(k+1) + \mathbf{Q}(k)]^{-1} \tag{C.76}$$

and $\mathbf{M}_1(k+1)$ satisfies, using $\tilde{\mathbf{P}}(k)$.

$$\begin{aligned}\mathbf{M}_1(k+1) &= (\boldsymbol{\Phi}^T(k+1, k))^{-1}\mathbf{M}(k)\boldsymbol{\Phi}^{-1}(k+1, k) \\ &\quad + (\boldsymbol{\Phi}^T(k+1, k))^{-1}\mathbf{C}^T(k)\mathbf{R}^{-1}(k)\mathbf{C}(k)\boldsymbol{\Phi}^{-1}(k+1, k)\end{aligned} \tag{C.77}$$

Now the equations for $\tilde{\mathbf{M}}(k+1)$ and $\mathbf{M}_1(k+1)$ correspond to those for $\mathbf{P}(k+1)$ and $\mathbf{M}(k+1)$, respectively, except for a system given by

$$\mathbf{x}_*(k+1) = [\boldsymbol{\Phi}^T(k+1, k)]^{-1}\mathbf{x}_*(k) + [\boldsymbol{\Phi}^T(k+1, k)]^{-1}\mathbf{C}^T(k)\mathbf{s}(k) \tag{C.78}$$

where $E[\mathbf{s}(k)] = \mathbf{0}$ and $E[\mathbf{s}(k)\mathbf{s}^T(k)] = \mathbf{R}^{-1}(k)$. The measurement equation corresponds to

$$\mathbf{z}_*(k+1) = \mathbf{x}_*(k+1) + \mathbf{m}(k+1) \tag{C.79}$$

with $E[\mathbf{m}(k)] = \mathbf{0}$; $E[\mathbf{m}(k)\mathbf{m}^T(k)] = \mathbf{Q}^{-1}(k)$.

But from our previous theorem we know that for any discrete-time system the covariance of the state at N, given $n+1$ past measurements, is bounded by

$$\mathbf{P}_*(N) \leqq \mathbf{M}_*^{-1}(N, N-n) + A_{1*}\mathbf{W}_*(N, N-n)$$

were the asterisk indicates that these are evaluated for this system. In this example we have the fact that $\mathbf{x}_*(N) = \Phi^T(k, N)\mathbf{x}(k)$, so that by following Theorem C. 1 we obtain,

$$\mathbf{M}_*(N, N-n) = \sum_{k=N-n}^{N} \Phi(N, k)\mathbf{Q}(k-1)(\Phi^T(N, k)) \tag{C.80}$$

which implies that

$$\mathbf{M}_*(N, N-n) = \mathbf{W}(N, N-n) \tag{C.81}$$

Similarly,

$$\begin{aligned}\mathbf{W}_*(N, N-n) &= \sum_{k=N-n}^{N-1} \Phi^T(k+1, N)\Phi^T(k, k+1)\mathbf{C}^T(k)\mathbf{R}^{-1}(k) \\ &\quad \mathbf{C}(k)\Phi(k, k+1)\Phi(k+1, N) \\ &= \sum_{k=N-n}^{N-1} \Phi^T(k, N)\mathbf{C}^T(k)\mathbf{R}^{-1}(k)\mathbf{C}(k)\Phi(k, N)\end{aligned} \tag{C.82}$$

Now this implies that

$$\begin{aligned}\mathbf{P}_*(N) = \tilde{\mathbf{M}}(N) = \tilde{\mathbf{P}}(N) - \mathbf{C}^T(N)\mathbf{R}^{-1}(N)\mathbf{C}(N) \leqq \mathbf{M}_*^{-1}(N, N-n) \\ + A_{1*}\mathbf{W}_*(N, N-n)\end{aligned} \tag{C.83}$$

But since $A_1 \geqq 1$, we have

$$\tilde{\mathbf{P}}(N) \leqq \mathbf{M}_*^{-1}(N, N-n) + A_{1*}[\mathbf{W}_*(N, N-n) + \mathbf{C}^T(N)\mathbf{R}^{-1}(N)\mathbf{C}(N)] \tag{C.84}$$

Now, clearly,

$$\mathbf{W}_*(N, N-n) + \mathbf{C}^T(N)\mathbf{R}^{-1}(N)\mathbf{C}(N) = \mathbf{M}(N, N-n) \tag{C.85}$$

Thus,

$$\tilde{\mathbf{P}}(N) \leqq \mathbf{W}^{-1}(N, N-n) + A_{1*}\mathbf{M}(N, N-n) \tag{C.86}$$

or

$$\mathbf{P}(N) \leqq [\mathbf{W}^{-1}(N, N-n) + A_{1*}\mathbf{M}(N, N-n)]^{-1} \blacksquare \tag{C.87}$$

We can now combine this result in the following corollary.

COROLLARY C.1. The covariance matrix $\mathbf{P}(N)$ is bounded by

$$\begin{aligned}[\mathbf{W}^{-1}(N, N-n) + A\mathbf{M}(N, N-n)]^{-1} \leqq \mathbf{P}(N) \leqq \mathbf{M}^{-1}(N, N-n) \\ + A\mathbf{W}(N, N-n)\end{aligned} \tag{C.88}$$

where

$$A = \frac{\beta\delta}{\alpha\gamma}\, n^2 \tag{C.89}$$

We have now upper- and lower-bounded the covariance matrix. We now want to introduce a function $V_p(\mathbf{x}(k), k)$ and finally show that it is a Lyapanov function for the estimate equation.

DEFINITION C.2. Let $V_p(\mathbf{x}(k), k)$ be given by

$$V_p(\mathbf{x}(k), k) = \mathbf{x}^T(k)\mathbf{P}^{-1}(k)\mathbf{x}(k) \tag{C.90}$$

where $\mathbf{P}(k)$ is the covariance matrix of the optimal filter.

As an immediate result from Corollary C.1 we note that there exist positive constants C_1 and C_2 such that for some N

$$C_1\|\mathbf{x}(N)\|^2 \leqq V_p(\mathbf{x}(N), N) \leqq C_2\|\mathbf{x}(N)\|^2 \tag{C.91}$$

This is the first requirement for this to be a Lyapanov function. The second is demonstrated in the following lemma.

LEMMA C.5. Let $\mathbf{P}(k)$ be the covariance function of the filter in equation (C.5). Let $V_p(\mathbf{x}(k), k)$ be given by

$$\mathbf{x}^T(k)\mathbf{P}^{-1}(k)\mathbf{x}(k)$$

Then

$$V_p(\mathbf{x}, k) - V_p(\mathbf{x}, k - N) \leqq -\gamma_3\|\mathbf{x}\| \tag{C.92}$$

The result of this lemma and Corollary C.9 imply that $\mathbf{x}^T(k)\mathbf{P}^{-1}(k)\mathbf{x}(k)$ is a Lyapanov function for the given system.

Proof. Consider the system given by the following equations:

$$\mathbf{x}(k) = \Phi(k, k-1)\mathbf{x}(k-1) + \mathbf{u}(k) \tag{C.93}$$

where we let $\mathbf{u}(k)$ be a control input. We establish a cost functional J as

$$J = \sum_{i=k-N}^{k} [\mathbf{x}^T(i)\mathbf{C}^T(i)\mathbf{R}^{-1}(i)\mathbf{C}(i)\mathbf{x}(i) + \mathbf{u}^T(i)\mathbf{M}^{-1}(i)\mathbf{u}(i)] \tag{C.94}$$

We now seek to determine the control sequence $\{\mathbf{u}^*(i)\}$ that will minimize this given J. Define the vectors $\mathbf{X}$ and $\mathbf{U}$ as

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}(k) \\ \mathbf{x}(k-1) \\ \vdots \\ \mathbf{x}(k-N) \end{bmatrix} \tag{C.95}$$

$$\mathbf{U} = \begin{bmatrix} \mathbf{u}(k) \\ \mathbf{u}(k-1) \\ \vdots \\ \mathbf{u}(k-N) \end{bmatrix} \tag{C.96}$$

and the matrices

$$\mathbf{M} = \begin{bmatrix} \mathbf{R}(k) & 0 & \cdots & 0 \\ 0 & \mathbf{R}(k-1) & \cdots & 0 \\ 0 & & \cdots & \mathbf{R}(k-N) \end{bmatrix} \tag{C.97}$$

$$\mathbf{B} = \begin{bmatrix} \mathbf{M}(k) & 0 & \cdots & 0 \\ 0 & \mathbf{M}(k-1) & \cdots & 0 \\ 0 & & \cdots & \mathbf{M}(k-N) \end{bmatrix} \tag{C.98}$$

and

$$\mathbf{L} = \begin{bmatrix} \mathbf{C}(k) & 0 & \cdots & 0 \\ 0 & \mathbf{C}(k-1) & \cdots & 0 \\ 0 & & \cdots & \mathbf{C}(k-N) \end{bmatrix} \tag{C.99}$$

Then the cost functional J is given by

$$J = \mathbf{X}^T\mathbf{L}^T\mathbf{M}^{-1}\mathbf{L}\mathbf{X} + \mathbf{U}^T\mathbf{B}^{-1}\mathbf{U} \tag{C.100}$$

Now we can make use of the system dynamics if we define two more matrices $\mathbf{C}$ and $\mathbf{D}$:

$$\mathbf{C} = \begin{bmatrix} \Phi(k, k-N-1) \\ \Phi(k-1, k-N-1) \\ \vdots \\ \Phi(k-N, k-N-1) \end{bmatrix} \tag{C.101}$$

and

$$\mathbf{D} = \begin{bmatrix} \mathbf{I} & \Phi(k, k-1) & \Phi(k, k-2) & \cdots & \Phi(k, k-N) \\ 0 & \mathbf{I} & \Phi(k-1, k-2) & & \vdots \\ & & \mathbf{I} & & \vdots \\ & 0 & & & \\ & & & & \mathbf{I} \end{bmatrix} \tag{C.102}$$

and define $\mathbf{x}_0$ as an initial condition given by

$$\mathbf{x}_0 = \mathbf{x}(k-N-1) \tag{C.103}$$

Therefore,

$$\mathbf{X} = \mathbf{C}\mathbf{x}_0 + \mathbf{D}\mathbf{U} \tag{C.104}$$

This follows directly from our discussion of Section 2.2 on discrete systems. Using this in (C.100) gives:

$$J = [\mathbf{C}\mathbf{x}_0 + \mathbf{D}\mathbf{U}]^T\mathbf{L}^T\mathbf{M}^{-1}\mathbf{L}[\mathbf{C}\mathbf{x}_0 + \mathbf{D}\mathbf{U}] + \mathbf{U}^T\mathbf{B}^{-1}\mathbf{U} \tag{C.105}$$

Following the results outlined in Chapter 6, we can now formally take derivatives and solve for the optimum $\mathbf{U}$. This yields

$$\mathbf{U}^* = -[\mathbf{D}^T\mathbf{L}^T\mathbf{M}^{-1}\mathbf{L}\mathbf{D} + \mathbf{B}^{-1}]^{-1} \cdot \mathbf{D}^T\mathbf{L}^T\mathbf{M}^{-1}\mathbf{L}\mathbf{C}\mathbf{x}_0 \tag{C.106}$$

Substituting (C.106) into (C.100), we obtain the minimum cost as

$$J^* = \mathbf{x}_0^T\mathbf{C}^T\mathbf{L}[\mathbf{M} + \mathbf{L}\mathbf{D}\mathbf{B}\mathbf{D}^T\mathbf{L}^T]^{-1}\mathbf{L}\mathbf{C}\mathbf{x}_0 \tag{C.107}$$

By the definition of **L**, **C**, **M**, we know that

$$\beta_1 \mathbf{I} \leqq \boldsymbol{\Phi}^T(k-N-1,k)\mathbf{C}^T\mathbf{L}^T\mathbf{M}^{-1}\mathbf{L}\mathbf{C}\boldsymbol{\Phi}(k-N-1,k) \leqq \beta_2 \mathbf{I} \quad \text{(C.108)}$$

Since $\mathbf{R}(k)$ is bounded below, it is obvious that

$$\mathbf{x}_0^T\mathbf{C}^T\mathbf{L}^T\mathbf{L}\mathbf{C}\mathbf{x}_0 \geqq \beta_3\|\mathbf{x}_0\| \quad \text{(C.109)}$$

where β_3 is a finite positive nonzero constant. This follows immediately from the fact that

$$\mathbf{x}_0 = \boldsymbol{\Phi}(k-N-1,k)\mathbf{x}(k) \quad \text{(C.110)}$$

and multiplying both sides of (C.108) by $\mathbf{x}(k)$ and using (C.110) with the realization that $\mathbf{M}^{-1}$ is related to the positive $\mathbf{R}(k)$ matrices. Also, now since $\mathbf{R}(k)$, $\boldsymbol{\Phi}(k, k-1)$, $\mathbf{M}(k)$ and $\mathbf{C}(k)$ are all bounded above and $\mathbf{M}$ is positive we have

$$\beta_4 \mathbf{I} \leqq \mathbf{M} + \mathbf{L}\mathbf{D}\mathbf{B}\mathbf{D}\mathbf{L}^T \leqq \beta_5 \mathbf{I}; \quad \text{(C.111)}$$

$$0 < \beta_4, \beta_5 < 0 \quad \text{(C.112)}$$

which gives

$$\beta_4^{-1}\mathbf{I} \geqq [\mathbf{M} + \mathbf{L}\mathbf{D}\mathbf{B}\mathbf{D}\mathbf{L}^T]^{-1} \geqq \beta_5^{-1}\mathbf{I} \quad \text{(C.113)}$$

Therefore, we find that the optimum value is lower-bounded by

$$\begin{aligned} J^* &= \mathbf{x}_0\,\mathbf{C}^T\mathbf{L}^T[\mathbf{M} + \mathbf{L}\mathbf{D}\mathbf{B}\mathbf{D}\mathbf{L}^T]^{-1}\mathbf{L}\mathbf{C}\mathbf{x}_0 \geqq \beta_5^{-1}\mathbf{C}^T\mathbf{L}^T\mathbf{L}\mathbf{C}\mathbf{x}_0 \\ &\geqq \beta_3^2\beta_5^{-1}\|\mathbf{x}_0\| = \beta_3^2\beta_5^{-1}\|\mathbf{x}(k-N-1)\| \end{aligned} \quad \text{(C.114)}$$

We now plan to use the boundedness of J^* to insure that the rate of change of the Lyapanov function is always negative. Let us now take equation (C.69) and use the undriven portion

$$\mathbf{x}(k+1) = \mathbf{P}(k+1)\mathbf{M}^{-1}(k+1)\mathbf{x}(k) \quad \text{(C.115)}$$

and rewrite it as

$$\mathbf{x}(k+1) = \mathbf{x}'(k+1) + \mathbf{u}(k+1) \quad \text{(C.116)}$$

where

$$\mathbf{x}'(k+1) = \boldsymbol{\Phi}(k+1,k)\mathbf{x}(k) \quad \text{(C.117)}$$

and

$$\mathbf{u}(k+1) = [\mathbf{P}(k+1)\mathbf{M}^{-1}(k+1) - \mathbf{I}]\mathbf{x}'(k+1) \quad \text{(C.118)}$$

The Lyapanov function thus becomes

$$\begin{aligned} V_p(\mathbf{x}(k),k) &= \mathbf{x}^T(k)\mathbf{P}^{-1}(k)\mathbf{x}(k) \\ &= \mathbf{x}^T(k)[\mathbf{M}^{-1}(k) + \mathbf{C}^T(k)\mathbf{R}^{-1}(k)\mathbf{C}(k)]\mathbf{x}(k) \end{aligned} \quad \text{(C.119)}$$

Now recall that

$$\mathbf{x}(k) = \mathbf{x}'(k) + [\mathbf{P}(k)\mathbf{M}^{-1}(k) - \mathbf{I}]\mathbf{x}'(k) \quad \text{(C.120)}$$

Therefore, (C.119) becomes

$$V_p(\mathbf{x}(k), k) = [\mathbf{x}'(k) + [\mathbf{P}(k)\mathbf{M}^{-1}(k) - \mathbf{I}]\mathbf{x}'(k)]^T$$
$$[\mathbf{M}^{-1}(k) + \mathbf{C}^T(k)\mathbf{R}^{-1}(k)\mathbf{C}(k)][\mathbf{x}'(k) + [\mathbf{P}(k)\mathbf{M}^{-1}(k) - I]\mathbf{x}'(k)] \quad \text{(C.121)}$$

which is equal to

$$V_p(\mathbf{x}(k), k) = \mathbf{x}'^T(k)\mathbf{M}^{-1}(k)\mathbf{x}'(k) - \mathbf{x}^T(k)\mathbf{C}^T(k)\mathbf{R}^{-1}(k)\mathbf{C}(k)\mathbf{x}(k)$$
$$+ 2\mathbf{x}^T(k)[\mathbf{P}^{-1}(k) - \mathbf{M}^{-1}(k)]\mathbf{x}(k)$$
$$+ \mathbf{x}^T(k)\mathbf{M}^{-1}(k)\mathbf{x}(k) - \mathbf{x}'^T(k)\mathbf{M}^{-1}(k)\mathbf{x}'(k) \quad \text{(C.122)}$$

and yields

$$V_p(\mathbf{x}(k), k) = \mathbf{x}'^T(k)\mathbf{M}^{-1}(k)\mathbf{x}'(k) - \mathbf{x}^T(k)\mathbf{C}^T(k)\mathbf{R}^{-1}(k)\mathbf{C}(k)\mathbf{x}(k)$$
$$- [\mathbf{x}(k) - \mathbf{x}'(k)]\mathbf{M}^{-1}(k)[\mathbf{x}(k) - \mathbf{x}'(k)] \quad \text{(C.123)}$$

But recall that

$$\mathbf{M}^{-1}(k) = [\Phi(k, k-1)\mathbf{P}(k-1)\Phi^T(k, k-1) + \mathbf{Q}(k)]^{-1} \quad \text{(C.124)}$$

Therefore, (C.117) becomes

$$V_p(\mathbf{x}(k), k) = \mathbf{x}'^T(k)[\Phi(k, k-1)\mathbf{P}(k-1)\Phi^T(k, k-1) + \mathbf{Q}(k)]^{-1}$$
$$\cdot\mathbf{x}'(k) - \mathbf{x}^T(k)\mathbf{C}^T(k)\mathbf{R}^{-1}(k)\mathbf{C}(k)\mathbf{x}(k) - \mathbf{u}^T(k)\mathbf{M}^{-1}(k)\mathbf{u}(k) \quad \text{(C.125)}$$

and now, using the transition matrix properties, we have with the use of (C.117)

$$V_p(\mathbf{x}(k), k) = \mathbf{x}^T(k-1)[\mathbf{P}(k-1) + \Phi(k-1, k)\mathbf{Q}(k)\Phi^T(k-1, k)]^{-1}\mathbf{x}(k-1)$$
$$- \mathbf{x}^T(k)\mathbf{C}^T(k)\mathbf{R}^{-1}(k)\mathbf{C}(k)\mathbf{x}(k) - \mathbf{u}^T(k)\mathbf{M}^{-1}(k)\mathbf{u}(k) \quad \text{(C.126)}$$

but since the matrix

$$\mathbf{P}(k-1) + \Phi(k-1, k)\mathbf{Q}(k)\Phi^T(k-1, k)$$

is greater than $\mathbf{P}(k-1)$ alone since $\mathbf{P}(k)$ and $\mathbf{Q}(k)$ are positive definite, we have

$$V_p(\mathbf{x}(k), k) \leqq \mathbf{x}^T(k-1)\mathbf{P}^{-1}(k-1)\mathbf{x}(k-1) - \mathbf{x}^T(k)\mathbf{C}^T(k)\mathbf{R}^{-1}(k)\mathbf{C}(k)\mathbf{x}(k)$$
$$- \mathbf{u}^T(k)\mathbf{M}^{-1}(k)\mathbf{u}(k) \quad \text{(C.127)}$$

Therefore, we have for any $\mathbf{x}(k-N)$ by induction:

$$V_p(\mathbf{x}(k), k) - V_p(\mathbf{x}(k-N), k-N) \leqq$$
$$- \sum_{i=k-N}^{k} [\mathbf{x}^T(i)\mathbf{C}^T(i)\ \mathbf{R}^{-1}(i)\mathbf{C}(i)\mathbf{x}(i) + \mathbf{u}^T(i)\mathbf{M}^{-1}(i)\mathbf{u}(i)] \quad \text{(C.128)}$$

But from definition on J we have

$$V_p(\mathbf{x}(k), k) - V_p(\mathbf{x}(k-N), k-N)) \leqq -J^* \quad \text{(C.129)}$$

and since J^* is the minimal cost and by (C.114), we have a bound on J^* that becomes

$$V_p(\mathbf{x}(k), k) - V_p(\mathbf{x}(k-N), k-N) \leqq -\beta_3^2\beta_5^{-1}\|\mathbf{x}(k-N-1)\|^2 \quad \text{(C.130)}$$

But recall that

$$\mathbf{x}(k) = \mathbf{P}(k)\mathbf{M}^{-1}(k)\mathbf{x}(k-1) \quad \text{(C.131)}$$

so that

$$\mathbf{x}(k) = \mathbf{P}(k)\mathbf{M}^{-1}(k)\mathbf{P}(k-1)\mathbf{M}^{-1}(k-1)\cdots \mathbf{P}(k-N)\mathbf{M}^{-1}(k-N)\mathbf{x}(k-N-1) \quad \text{(C.132)}$$

Therefore,

$$\mathbf{x}(k-N-1) = \theta(k, k-N-1)^{-1}\mathbf{x}(k) \quad \text{(C.133)}$$

Where

$$\theta(k, k-N-1) = \mathbf{P}(k)\mathbf{M}^{-1}(k)\cdots\mathbf{P}(k-N)\mathbf{M}^{-1}(k-N) \quad \text{(C.134)}$$

This then yeilds

$$\|\mathbf{x}(k-N-1)\| = \|\theta(k, k-N-1)^{-1}\mathbf{x}(k)\| \geqq \beta_6\|\mathbf{x}(k)\| \quad \text{(C.135)}$$

Therefore, using (C.135) in (C.130)yields

$$V_p(\mathbf{x}(k), k) - V_p(\mathbf{x}(k-N), k-N) \leqq -\beta_3^2\beta_5^{-1}\beta_6\|\mathbf{x}(k)\|^2 \equiv \gamma_3(\|\mathbf{x}(k)\|) \quad \text{(C.136)}$$

which proves the existence of a Lyapanov function and also proves the theorem. ∎

As a result of this lemma we can now say that $V_p(\mathbf{x}(k),k)$ is a Lyapanov function and as a result of Theorem 4.1 in Chapter 2 in Theorem C.1 is u.a.s.i.l. This implies bounded-input–bounded-output stability of the system also. Similarly, it allows us to say that the divergence problem is also stable under stochastic observability and controllability conditions.

The extension of these results to continuous-time systems is contained in Bucy [3]. Also a different approach to discrete-time stability is in Bucy [2]. The general issue of the stability of the Riccati equation is discussed elsewhere but is beyond the scope of the present problem.

GLOSSARY OF SYMBOLS

$\mathbf{A}(t)$	State matrix
$\mathbf{B}(t)$	Forcing function matrix
$\mathbf{C}(t)$	Measurement matrix
$C(\lambda; \tau)$	Lipschitz cylinder
$E[\ ]$	Expectation
$\mathbf{f}(\ )$	Nonlinear portion of dynamic system
~~$\mathbf{h}$~~	Nonlinear function of state in the measurement system
$\mathbf{M}_c(t_0, t_1)$	Continuous-time observability matrix
$\mathbf{M}_d(0, N)$	Discrete-time observability matrix
$\mathbf{W}_c(t_0, t_1)$	Continuous-time controllability matrix
$\mathbf{W}_d(0, N)$	Discrete-time controllability matrix
$V(\mathbf{x}(t), t)$	Continuous-time Lyapanov function
$V(\mathbf{x}(k), k)$	Discrete-time Lyapanov function
p, $p(\mathbf{x})$, $p_{\mathbf{x}(t)}(\mathbf{u})$	Probability density function
$p(\mathbf{x}/\mathbf{y})$, $p_{\mathbf{x}(t)/\mathbf{y}(t)}(\mathbf{u}/\mathbf{v})$	Conditional probability density function
$P, p[\ ]$	Probability or provability distribution function
H	Hilbert space
T_f	Total variation of a function f on an interval I
$\mathbf{I}$	A closed and bounded interval
$\mathbf{u}(t)$, $\mathbf{v}(t)$, $\mathbf{s}(t)$, $\mathbf{r}(t)$, $\mathbf{w}(t)$	Continuous-time noise processes; may be Wiener or white noise
$\mathbf{u}(k)$, $\mathbf{v}(k)$, $\mathbf{w}(k)$	Discrete-time Gaussian random variable noises
$\mathbf{u}(t)$	External deterministic control
$\mathbf{K}(k)$	Optimum gain of Kalman-Bucy filter
$\mathbf{x}(t, \omega)$, $\mathbf{x}(t)$	Random process
$\mathbf{x}(t)$	Continuous-time state variable
$\mathbf{x}(k)$	Discrete-time state variable
$\hat{\mathbf{x}}(t)$, $\hat{\mathbf{x}}(k)$	Optimum estimates (MMSE)
$\tilde{\mathbf{x}}(t)$, $\tilde{\mathbf{x}}(k)$	Errors of optimum estimators
$\mathbf{x}^*(t)$, $\mathbf{x}^*(k)$	Approximate optimum estimates

$\bar{\mathbf{x}}(t)$	Expansion point
$\mathbf{z}(t)$, $\mathbf{y}(t)$	Continuous-time measurement
$\mathbf{z}(k)$, $\mathbf{y}(k)$	Discrete-time measurement
$\mathbf{Q}(t)$, $\mathbf{R}(t)$, $\mathbf{U}(t)$, $\mathbf{S}(t)$	Continuous-time noise covariance matrices
$\mathbf{Q}(k)$, $\mathbf{R}(k)$, $\mathbf{U}(k)$, $\mathbf{S}(k)$	Discrete-time covariance matrices
t,s	Time variable
T	Sampling time
$M(ju)$	Characteristic function
$\mathbf{A}$	Gradient matrix associated with $f(\ )$ nonlinearity
$\mathbf{C}$	Gradient matrix associated with $h(\ )$ nonlinearity
$\mathbf{n}_g(\ )$	Gaussian portion of continuous-time Noise Process
$\mathbf{n}_p(\ )$	Poisson portion of continuous-time noise process
$O_{t_0 t}$	Minimum σ-field associated with Measurements
$\Phi(t, t_0)$	Continuous-time transition Matrix
$\Phi(k, j)$	Discrete-time transition matrix
$\psi(t, t_0)$	Adjoint system transition matrix
$\Phi(t, \mathbf{x}(0), t_0)$	Solution of continuous-time nonlinear state equation
$\Phi(k, \mathbf{x}(0), t_0)$	Solution of discrete-time nonlinear state equation
Λ	Covariance matrix
λ, $\lambda(t)$	Arrival rate of Poisson process
σ^2	Variance of a Gaussian random variable
∇	Gradient operator
sup	Supremum
inf	Infimum
max	Maximum
min	Minimum
$\int_I$	Ito integral
$\int_S$	Stratonovich integrla
$\int$, $\int_R$	Riemann integral
l.i.m.	Limit in the mean

BIBLIOGRAPHY

Anderson, B. D. O.

[1] "Stability Properties of Kalman-Bucy Filters," *J. Frank. Inst.*, **291**, 137–144 (1971).

Anderson, B. D. O. and J. B. Moore

[1] "The Kalman-Bucy Filter as a True Time Varying Wiener Filter," *IEEE Trans. Sys. Man, Cy.*, **SMC-1**, No. 2 119–128 (April 1971).

Anderson, T. W.

[1] "The Integral of a Symmetric Unimodal Function Over a Symmetric Convex Set and Some Probability Inequalities," *Proc. Am. Math. Soc.*, **6**, 170–196 (1955).

Anstrom K. J.

[1] "On a First-Order Stochastic Differential Equation," *Int. J. Control*, **1**, 301–326 (1965).

[2] *Introduction to Stochastic Control Theory*, Academic, New York, 1970.

Aoki, M.

[1] *Optimization of Stochastic Systems*, Academic, New York, 1967.

Aronszajn, N.

[1] "Theory of Reproducing Kernels," *Trans. Am. Math. Soc.*, **63**, 337–404 May 1950.

Athans, M.

[1] "On the Equivalence of Linearized Kalman Filters," *MIT Lincoln Lab. Rpt. 1961-65* (1967).

Athans, M. and P. L. Falb

[1] *Optimal Control*, McGraw-Hill, New York, 1966.

Athans, M. and E. Tse

[1] "A Direct Derivation of the Optimal Linear Filtering Using the Maximum Principle," *IEEE*, **AC-12**, No. 6, 690–698 (December 1967).

Athans, M., R. P. Wishner, and A. Bertolini

[1] "Suboptimal State Estimation for Continuous-Time Nonlinear Systems from Discrete Noisy Measurements," *IEEE*, **AC-13**, No. 5, 504–514 (October 1968).

Baggeroer, A. B.

[1] "A State Variable Approach to the Solution of Fredholm Integral Equations." *IEEE*, **IT-15**, No. 5, 557–569 (September 1969).
[2] *State Variables and Communication Theory*, MIT Press, Cambridge, Mass., 1970.

Bailey, N. T. J.
[1] *The Elements of Stochastic Processes*, Wiley, New York, 1964.

Balakrishnan, A. V.
[1] "A General Theory of Nonlinear Estimation Problems in Control Systems," *J. Math. Anal. App.*, **8**, 4–30 (1964).
[2] "State Estimation for Infinite Dimensional Systems," *Computer and System Sci.*, **1**, 391–403 (1967).

Bar-David, I.
[1] "Communication Under the Poisson Regime," *IEEE*, **IT-15**, No. 1, 31–37 (January 1969).

Barucha-Reid, A. T.
[1] *Elements of the Theory of Stochastic Processes and Their Applications*, McGraw-Hill, New York, 1960, pp. 334–356

Bass, R. W., V. D. Norum, and L. Schwartz
[1] "Optimal Multichannel Nonlinear Filtering," *Hughes Aircraft Report SSD 50064R* (1965)
[2] "Optimum Multichannel Nonlinear Filtering," *J. Math. Anal. Appl.*, **16**, 152–164 (1966).

Battin, R. H.
[1] *Astronautical Guidance*, McGraw-Hill, New York, 1964.

Bellman, R. and R. Kalaba
[1] *Dynamic Programming and Modern Control Theory*, Academic, New York, 1965.

Ben-Israel, A. and A. Charnes
[1] "Contributions to the Theory of Generalized Inverses," *J. Soc. Ind. Appl. Math.* **11**, No. 3, 667–669 (September 1963).

Bensoussan, A.
[1] "Sur L'Identification et le Filtrage de Systems Gouvernes par des Equations Aux Derivees Partielles," Ph. D. diss. University of Paris, 1969.

Bertsekas, D. P. and I. B. Rhodes
[1] "On the Minimax Reachability of Target Sets and Target Tubes," *Automatica*, 7, 233–247 (1971).
[2] "Recursive State Estimation for a Set Membership Description of Uncertainty," *IEEE Trans. A. C.*, **AC-16**, No. 2, 117–128 (April 1971).

Billingsley, P.
[1] *Convergence of Probability Measures*, Wiley, New York, 1968.

Bochner, S.
[1] "Stochastic Processes," *Ann. Math.* **48**, No. 4, 1014–1061 (October 1947).

Breiman, L.

[1] *Probability Theory*, Addison-Wesley, Reading, Mass,. 1968.

Brockett, R. W.
[1] *Finite Dimensional Dynamical Systems*, Wiley, New York, 1970.

Bryson, A. E. and M. Frazier
[1] "Smoothing for Nonlinear Dynamic Systems," *Proc. Optimum Systems Synthesis Conference*, Wright Patterson AFB, Ohio TDR ASD-TRD-63-119, 354–364 February 1963.

Bryson, A. E. and D. E. Johansen
[1] "Linear Filtering for Time-Varying Systems Using Measurements Containing Colored Noise," *IEEE*, **AC-10**, No. 1, 4–10 (January 1965).

Bucy, R. S.
[1] "Nonlinear Filtering Theory," *IEEE*, **AC-1**, No. 2, 198 (April 1965).
[2] "A Priori Bounds for the Riccati Equation," *Proc. Sixth Berkeley Smyp.*, University of California Press, Berkeley, 1972, pp. 645–656
[3] "The Riccati Equation and Its Bound" (in press).

Bucy, R. S., C. Hecht, and K. D. Senne
[1] "An Engineers Guide to Building Nonlinear Filters," F. J. Seiler Research Lab., SRL-TR-72-0004

Bucy, R. S. and P. D. Joseph
[1] *Filtering for Stochastic Processes with Applications to Guidance*, Wiley, New York, 1968.

Bucy, R. S. and K. D. Senne
[1] "Digital Synthesis of Non-Linear Filters," *Automatica*, **7**, 287–299 (1971).

Cameron, R. H. and W. T. Martin
[1] "The Transformation of Wiener Integrals by Nonlinear Transformations," *Trans. Am. Math. Soc.*, **66**, 253–283 (1949).

Clark, J. M. C.
[1] "The Representation of Functionals of Brownian Motion by Stochastic Integrals," *Ann. Math. Stat.*, **41**, No. 4, 1282–1295 (1970).

Clark, J. R.
[1] "Estimation for Poisson Processes with Applications in Optical Communication," Ph. D. Thesis, MIT, 1971.

Coles, W. J
[1] "Matrix Riccati Differential Equations," *Jour. SIAM*, **13**, No. 3, 627–634 (September 1965).

Cox, H.
[1] "On the Estimation of State Variables and Parameters for Noisy Dynamic Systems," *IEEE*, **AC-9**, No. 1, 5–13 (January 1964).

Cramer, H. and M. R. Leadbetter
[1] *Stationary and Related Stochastic Processes*, Wiley New York, 1967.

Culver, C. O.
[1] "Optimal Estimation for Nonlinear Stochastic Systems," Sc. D. Thesis, MIT, 1969.

[2] "Optimal Estimation for Nonlinear Systems," AIAA Conference, Princeton, N.J., 1969, AIAA Paper No. 69–852

Davenport, W. B. and W. L. Root
[1] *Random Signals and Noise*, McGraw-Hill, New York, 1958.

De Russo, P. M., R. J. Roy, and C. M. Close
[1] *State Variables for Engineers*, Wiley, New York, 1965.

Desoer, C. A.
[1] *Notes for a Second Course on Linear Systems*, Van Nostrand, New York, 1970.

Desoer, C. A. and R. H. Whalen
[1] "A Note on Pseudo Inverses," *Jour. SIAM*, **11**, No. 2, 442–447 (June 1962).

Detchmendy, D. M. and R. Sridhar
[1] "Sequential Estimation of States and Parameters in Noisy Nonlinear Dynamical Systems," *Trans. ASME*, 362–368 (June 1966).

Deutsch, R.
[1] *Estimation Theory*, Prentice-Hall, Englewood Cliffs, N. J., 1965.

Deyst, J. J. and C. F. Price
[1] "Conditions for Asymptotic Stability of the Discrete Minimum-Variance Linear Estimator," *IEEE*, **AC-13**, No. 6, 702–705 (December 1968).

Dolph, C. L. and M. A. Woodbury
[1] "On the Relation Between Green's Functions and Covariances of Certain Stochastic Processes and its Application to Unbiased Linear Prediction," *Trans. Amer. Math. Soc.*, **72**, 519–550 (1952).

Doob, J. L.
[1] "The Brownian Movement and Stochastic Equations," *Ann. Math.* **43**, No. 2, 351–369 (April 1942).
[2] *Stochastic Processes*, Wiley, New York, 1953.

Dudley, R. M.
[1] "Gaussian Processes on Several Dimensions," *Ann. Math. Stat.*, **36**, No. 3, 771–788 (June 1965).
[2] "On Prediction Theory for Nonstationary Sequences," *Proc. Fifth Berkeley Symp. on Math. Stat. and Prob.*, University of California Press, 1966.

Duncan, T. E.
[1] "Probability Densities for Diffusion Processes with Applications to Nonlinear Filtering Theory and Detection Theory," Ph. D. Thesis, Stanford University, 1967.
[2] "Evaluation of Likelihood Ratios," *Info. Control*, **13**, 62–74 (1968).
[3] "On the Nonlinear Filtering Problem," University of Michigan Report AD-687157

Duttweiler, D. L.
[1] "Reproducing Kernel Hilbert Space Techniques for Detection and Estimation," Ph.D. Thesis, Stanford University, 1970.

Dym, H. and H. P. McKean

[1] "Application of DeBranges Spaces of Integral Functions to the Prediction of Stationary Gaussian Processes," *Ill. Jour. Math.*, **14**, No. 12, 299–343 (1970).

Dynkin, E. B.
[1] *Markov Processes*, Vol. I, Academic, New York, 1965.
[2] *Markov Processes*, Vol. II, Academic, New York, 1965.

Einstein, A.
[1] *Investigations on the Theory of the Brownian Movement*, Dover, New York, 1956.

Elliot, D.
[1] "Controllable Nonlinear Systems Driven by White Noise," Ph.D. Thesis, UCLA, 1969.

Evans, J. E.
[1] "Preliminary Results on Performance Bounds for the Detection of Stochastic Signals in Additive White Gaussian Noise," MIT *RLE*, **QPR**, 113–124 (April 1970).
[2] "Chernoff Bounds on the Error Probability for the Detection of Non-Gaussian Signals," Sc. D. Thesis, MIT, 1971.

Falb, P. L.
[1] "Infinite Dimensional Filtering: The Kalman-Bucy Filter in Hilbert Space," *Info. Control.*, **11**, 102–127 (1967).

Feller, W.
[1] *An Introduction to Probability Theory and Its Applications*, Vol. I, Wiley, New York, 1950.
[2] *An Introduction to Probability Theory and Its Applications*, Vol. II, Wiley, New York, 1966.

Fisher, J. R.
[1] "Optimal Nonlinear Filtering," C. T. Leondes (Ed.) *Advances in Control System*, Academic, New York, 1967.

Fisher, J. R. and E. B. Stear
[1] "Optimal Nonlinear Filtering for Independent Increment Processes—Part I," *IEEE*, **IT-3**, No. 4, 558–578 (October 1967).

Fisk, D. L.
[1] "Quasi-martingales and Stochastic Integrals," Ph.D. Thesis, Michigan State University, 1963.
[2] "Quasi-Martingales," *Trans. Am. Math. Soc.*, **120**, 369–389 (1965).

Fitzgerald, R. J.
[1] "Divergence of the Kalman Filter," *IEEE*, **AC-16**, No. 6, 736–747 (December 1971).

Fleming, W. H.
[1] "Optimal Continuous-Parameter Stochastic Control," *SIAM Rev.*, **11**, No. 4, 470–509 (October 1969).

Foster, M.

[1] "An Application of the Wiener-Kolmogorov Smoothing Theory to Matrix Inversion," *J. SIAM*, **9**, No. 3, 387–397 (September 1961).

Freedman, D.
[1] *Brownian Motion and Diffusion*, Holden-Day, San Francisco, 1971.

Frost, P. A.
[1] "Nonlinear Estimation in Continuous Time Systems," Ph.D. Thesis, Stanford University, 1968.

Frost, P. A. and T. Kailath
[1] "An Innovations Approach to Least-Squares Estimation—Part III: Nonlinear Estimation in White Gaussian Noise," *IEEE*, **AC-16**, No. 3, 217–226 (June 1971).

Fujisaki, M., G. Kallianpur, and H. Kunita
[1] "Stochastic Differential Equations for the Nonlinear Filtering Problem" (in press).

Genin, Y.
[1] "A Note on Linear Minimum Variance Estimation Problems," *IEEE*, **AC-13**, No. 1, 103 (February 1968).

Geesey, R.
[1] "Canonical Modelling of Gaussian Random Processes with Application to Estimation and Detection of Continuous Time Processes," Ph.D. Thesis, Stanford University, 1968.

Gikhman, I. I. and A. V. Skorokhod
[1] *Introduction to the Theory of Random Processes*, Saunders, Philadelphia, 1969.

Gilman, A. S.
[1] "Mean Square Performance Bounds for Almost-Linear Systems," Ph.D. Thesis, Washington University, 1972

Gnedenko, B. V.
[1] *The Theory of Probability*, Chelesa, New York, 1962.

Gould, L. A.
[1] *Chemical Process Control*, Addison-Wesley, Reading, Mass., 1969.

Greville, T. N. E.
[1] "The Pseudo-Inverse of a Rectangular or Singular Matrix and Its Application to the Solution of Systems of Linear Equations," *SIAM Rev.*, **1**, No. 1, 38–43 (January 1959).
[2] "Some Applications of the Pseudo-Inverse Matrix," *SIAM Rev.*, **2**, No. 1, 15–22 (January 1960).

Halmos, P.
[1] "Theory of Unbiased Estimation," *Ann. Math. Stat.*, **17**, 34–43 (1946).
[2] *Measure Theory*, Van Nostrand, Princeton, New Jersey, 1950.
[3] *Introduction to Hilbert Space and the Theory of Spectral Multiplicity*, Chelsea, New York, 1951.
[4] *Finite Dimensional Vector Spaces*, Van Nostrand, New York, 1958.

Heffes, H.
[1] "The Effect of Erroneous Models on the Kalman Filter Response," *IEEE*, **AC-11**, No. 3, 541–543 (July 1966).

Hida, T.
[1] *Stationary Stochastic Processes*, Princeton University Press, Princeton, New Jersey, 1970.

Ho, Y, C,
[1] "On the Stochastic Approximation Method and Optimal Filtering Theory," *J. Math. Anal. Appl.* **6**, 152–154 (1962).

Ho, Y, C, and A. K. Agrawala
[1] "On Pattern Classification Algorithms—Introduction and Survey," *IEEE*, **AC-13**, No. 6, 676–680 (December 1968).

Hoffman, K. and R. Kunze
[1] *Linear Algebra*, Prentice-Hall, Englewood Cliffs, New Jersey, 1961.

Horowitz, E.
[1] "A Characterization of Measures for a Class of Continuous-Time Partially Observable Markov Processes," Ph.D. Thesis, University of California, Los Angeles, 1970.

Huddle, J. R. and D. A. Wismer
[1] "Degradation of Linear Filter Performance Due to Modeling Error," *IEEE*, **AC-13**, No. 4, 421–423 (August 1968).

Ince, E. L.
[1] *Ordinary Differential Equations*, Dover, New York, 1956.

Ito, K.
[1] "Isotropic Random Current," *Proc. Third Berkeley Symposium on Mathematical Statistics and Probability, Berkeley and Los Angeles*, University of California Press, **2**, 125–132 (1956).
[2] *Lectures on Stochastic Processes*, Tata Institute of Fundamental Research, Bombay, India, 1961.
[3] "Stochastic Integral," *Proc. Imp. Acad.* (Tokyo), **20**, 519–529 (1944).
[4] "On Stochastic Differential Equations," *Mem. Am. Math. Soc.*, **4**, 1–51 (1951).
[5] "Multiple Wiener Integral," *Math. Soc. Japan*, **3**, No. 1, 157–169 (May 1951).

Ito, K. and H. P. McKean
[1] *Diffusion Processes and Their Sample Paths*, Springer, Berlin, 1965

Jaswinski, A. H.
[1] "Filtering for Nonlinear Systems," *IEEE*, **AC-11**, 765–766 (October 1966).
[2] *Stochastic Processes and Filtering Theory*, Academic, New York, 1970.

Kac, M.
[1] "Random Walk and the Theory of Brownian Motion," Address given at Annual Meeting of Association at Swarthmore, Pennsylvania, December 26–27, 1946.

[2] *Probability and Related Topics in Physical Sciences, Wiley* New York, 1959.

Kailath, T. and M. Zakai

[1] "Absolute Continuity and Radon-Nikodym Derivatives for Certain Measures Relative to Wiener Measure," *Ann. Math. Stat.*, **42**, No. 1, 130–140 (1971).

Kailath, T.

[1] "Correlation Detection of Signals Perturbed by a Random Channel," *IRE*, **IT-6**, No. 3, 361–366 July(1960).

[2] "An Innovations Approach to Least Squares Estimation—Part I: Linear Filtering in Additive White Noise," *IEEE*, **AC-13**, No. 6, 646–654 (December 1968).

[3] "A General Likelihood Ratio Formula for Random Signals in Gaussian Noise," *IEEE*, **IT-15**, No. 3, 350–362 (May 1969).

[4] "The Innovations Approach to Detection and Estimation Theory," *Proc. IEEE*, **58**, No. 5, 680–695 (May 1970).

[5] "A Further Note on a General Likelihood Formula for Random Signals in Gaussian Noise," *IEEE*, **IT-16**, No. 4, 393–396 (July 1970).

[6] "Some Extensions of the Innovations Theorem," *B.S.T.J.*, **50**, No. 4, 1487–1495 (1971).

[7] "RKHS Approach to Detection and Estimation Problems—Part I: Deterministic Signals in Gaussian Noise," *IEEE*, **IT-17**, No. 5, 530–549 (September 1971).

[8] "Likelihhod Ratios for Gaussian Processes," *IEEE*, **IT-16**, No. 3, 276–288 (May 1970).

[9] "The Structure of Radon-Nikodym Derivatives with Respect to Wiener and Related Measures," *Ann. Math. Stat.*, **42**, No. 3, 1054–1067 (1971).

Kailath, T. and P. A. Frost

[1] "An Innovations Approach to Least Squares Estimation Part II," *IEEE*, **AC-13**, No. 6, 655–661 (December 1968).

Kailath T. and Geesey R.

[1] "An Innovations Approach to Least-Squares Estimation—Part IV: Recursive Estimation Given the Covariance Functions" (in press).

Kalaba, R.

[1] "On Nonlinear Differential Equations, The Maximum Operation, and Monotone Convergence," *J. Math. Mech.* **8**, No. 4 519–574 (1959).

Kaillianpur, G. and C. Striebel

[1] "Estimation of Stochastic Systems: Arbitrary System Process with Additive White Noise Observation Errors," *Ann. Math. Stat.*, **39**, No. 3, 785–801 (1968).

[2] "Stochastic Differential Equations Occurring in the Estimation of Continuous Parameter Stochastic Processes," *Theo. Prob. App.*, **14**, No. 4, 567–594 (1969).

[3] "Stochastic Differential Equations in Statistical Estimation Problems," in P. R. Krishnaiah (Ed.), *Multivariate Analysis*, Vol. II, Academic, New York, 1969.

Kalman, R. E.
[1] "A New Approach to Linear Filtering and Prediction Problems," *J. Basic Engineering*, 35–45 (March 1960).
[2] "On the General Theory of Control Systems," *Proc. First International Congress on Aut. Cont., Moscow 1960*, Butterworth's, London, 1961, Vol. I pp. 481–492.
[3] "New Methods in Wiener Filtering Theory," in J. L. Bogdanoff and F. Kozin (Eds.), *Proc. First Symp. on Engineering Applications of Random Function Theory and Probability*, Wiely, New York, 1963.
[4] "Mathematical Description of Linear Dynamical Systems," *SIAM Control*, 1 No. 2, 152–192 (1963).

Kalman, R. E. and J. E. Bertram
[1] "Control Systems Analysis and Design Via the "Second Method" of Lyapanov," *J. Basic Engineering*, 371–400 (June 1960).

Kalman, R. E. and R. S. Bucy
[1] "New Results in Linear Filtering and Prediction Theory," *J.* Basic Engineering, 95–108 (March 1961).

Kalman, R. E., P. L. Falb, and M. A. Arbib
[1] *Topics in Mathematical System Theory*, McGraw-Hill New York, 1969.

Kashyap, R. L.
[1] "Maximum Likelihood Identification of Stochastic Linear System," *IEEE*, **AC-15**, No. 1, 25–34 (February 1970).

Kelly, E. J. and W. L. Root
[1] "A Representation of Vector-Valued Random Processes," *J. Math. Phys.*, **39**, No. 9, 211–216 (October 1960).

Kestelman, H.
[1] *Modern Theories of Integration*, Dover, New York, 1960.

Klinger, A.
[1] "Prior Information and Bias in Sequential Estimation," *IEEE*, **AC-13**, No. 1, 102–103 (February 1968).

Kolmogorov, A. N.
[1] "Stationary Sequences in Hilbert Space," *Byulleten Moskouskogo Gosudarstvennogo Universiteta Matematika* **2**(6), 1–40 (1941). [Engl. trans.: *Bull. Moscow State Univ.*, **2**, No. 6.]

Kolmogorov, A. N. and S. V. Fomin
[1] *Introductory Real Analysis*, Prentice-Hall, Englewood Cliffs, New Jersery, 1970.

Kou, S. R., D. L. Elliot, and T. J. Tarn
[1] "Observability of Nonlinear Systems," *Info. Control*, **22**, 89–99 (1973).

Kunita, H. and S. Watanabe
[1] "On Square Integrable Martingales," *Nagoga Math. J.*, **30**, 209–245 (1967).

Kushner, H. J.
[1] "On the Differential Equations Satisfied by Conditional Probability Densities

of Markov Processes, with Applications," *J. SIAM Control*, Ser, A., **2**, No. 1, (1964).

[2] "On the Dynamical Equations of Conditional Probability Density Functions, with Applications to Optimal Stochastic Control Theory," *J. Math. Anal. Appl.* 8, 332–344 (1964).

[3] "Dynamical Equations for Optimal Nonlinear Filtering," *J. Differential Equations*, 3, 179–190 (1967).

[4] "Approximations to Optimal Nonlinear Filters," *IEEE*, **AC-12**, No. 5, 546–556 (October 1967).

[5] *Stochastic Stability and Control*, Academic, New York, 1967.

[6] "Filtering for Linear Distributed Parameter Systems," *J. SIAM Control*, **8**, No. 3, 346–359 (August 1970).

[7] *Introduction to Stochastic Control*, Holt, Rinehart New York, 1971.

Kuznetsov, P. I., Stratonovich, R. L. and V. I. Tikhonov

[1] "Quasi-Moment Functions in the Theory of Random Processes," *Theo. Prob. Appl.*, **5**, No. 1, 80–97 (1960).

[2] "Some Problems with Conditional Probability and Quasi-Moment Functions, "*Theo. Prob. Appl.* **6**, 422–427 (1961).

Lainiotis, D. G.

[1] "A Nonlinear Adaptive Estimation Recursive Algorithm, "*IEEE*, **13**, No. 2, 197–198 (April 1968).

Lainiotis, D. G., J. G. Deshpande, and T. N. Upadhyay

[1] "Optimal Adaptive Control: A Nonlinear-Separation Theorem," *Int. J. Control*, **15**, No. 5, 787–888 (1972).

Larson, R. E. and J. Peschon

[1] "A Dynamic Programming Approach to Trajectory Estimation," *IEEE*, **AC-11**, No. 3, 537–540 (July 1966).

Lasalle, J. and S. Lefschetz

[1] *Stability by Lyapanov's Second Method*, Academic, New York, (1961).

Levin, J. J.

[1] "On the Matrix Riccati Equation," *Proc. Am. Math. Soc.*, **10**, 519–524 (1959).

Levy, P.

[1] *Processus Stochastiques et Mouvement Brownien*, Gauthier-Villars, Paris, 1948.

[2] "A Special Problem of Brownian Motion, and a General Theory of Gaussian Random Functions," *Proc. Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. II,U. of Calif. Press, Berkeley, 1956, pp.133–175

Lighthill, M. J.

[1] *Introduction to Fourier Analysis and Generalized Functions*, Cambridge University Press, Cambridge, England, 1958.

Lipschutz, S.

[1] *Theory and Problems of General Topology*, McGraw-Hill, New York, 1965.

Loeve, M.

[1] *Probability Theory*, Van Nostrand, Princeton, New Jersey, 1963.

Luenberger, D. G.
[1] *Optimization by Vector Space Methods*, Wiely, New York, 1969.
[2] "Observing the State of a Linear System," *IEEE*, Vol. **ME-8**, 74–80 (April 1964).

Masani, P.
[1] "Wiener's Contributions to Generalized Harmonic Analysis, Prediction Theory and Filter Theory," *Bull. Am. Math. Soc.*, **72**, No. 1, pt. 2, 73–125 (January 1966),

Masani, P. and N. Wiener
[1] "Nonlinear Prediction," in U. Grenander, *Probability and Statistics*, Wiley, New York, 1959, pp. 190–212.

McGarty, T. P.
[1] "The Estimation of the "Center of Gravity" of a Photon Density Profile in Noise," *IEEE*, **AES-5**, No. 6, 974–980 (November 1969).
[2] "The Estimation of the Constituent Densities of the Upper Atmosphere by Means of a Recursive Filtering Algorithm," *IEEE*, **AC-16**, No. 6, (December 1971).
[3] "An Inversion Procedure for a Photon Limited Environment-Star Occultation," MIT Draper Lab. Report No. AER-8-4 (September 1970).

McKean, H. P.
[1] "Brownian Motion with a Several-Dimensional Time," *Theory Prob. Appl.* **8**, 335–354 No. 4, (1963).
[2] *Stochastic Integrals*, Academic, New York, 1969.

McShane, E. J.
[1] *A Riemann-Type Integral that Includes Lebesgue-Stieltjes, Bochner and Stochastic Integrals*, Am. Math. Soc. No. 88, Providence, R.I., 1969.
[2] "Stochastic Differential Equations and Models of Random Processes," in *Proc. Sixth Berkeley Symp.*, University of California Press., Berkeley, 1972, pp. 263–294.

Meditch, J. S.
[1] "Orthogonal Projection and Discrete Linear Smoothing," *J. SIAM Control*, **J**, No. 1, 74–89 (1967).
[2] *Optimal Estimation and Control*, McGraw-Hill, New York, 1969.
[3] "On State Estimation for Distributed Parameter Systems," *J. Franklin Inst.*, **290**, No. 1, 49–59 July (1970).

Mehr, C. B. and J, A. McFadden
[1] "Certain Properties of Gaussian Processes and Their First-Passage Times," *J. Royal Stat. Soc.*, Ser. B. **27**, No. 3, 505–522 (1965).

Mehra, R.
[1] "On the Identification of Variances and Adaptive Kalman Filtering," *IEEE*, **AC-15**, No. 2, 175–184 (April 1970).
[2] "A Comparison of Several Nonlinear Filters for Re-entry Vehicle Tracking," *IEEE*, **AC-16**, No. 4, 307–319 (August 1971).

Meyer, P.
[1] "A Decomposition Theorem for Supermartingales," Ill. *J. Math.*, **6**, 193–205 (1962).
[2] "Decomposition of Supermartingales; The Uniqueness Theorem," *Ill. J. Math.*, **7**, 1–17 (1963).

Millar, P. W.
[1] "Stochastic Integrals and Processes with Stationary Independent Increments," in *Proc. Sixth Berkeley Symp.*, University of Calif. Press, Berkeley, 1972, pp. 307–331

Moore, J. T.
[1] *Elements of Linear Algebra and Matrix Theory*, McGraw-Hill, New York, 1968.

Moore, R. L.
[1] "Adaptive Estimation and Control for Nuclear Power Plant Laod Changes," Ph.D. Thesis, MIT, 1971.

Mortensen, R. E.
[1] "Optimal Control of Continuous-Time Stochastic Systems," Ph.D. Thesis, University of California, Berkeley, 1966.
[2] "The Representation Theorem," Symp. on Nonlinear Est., 1970, Western Periodicals Co., No. Hollywood, California

Mowery, V. O.
[1] "Least Squares Recursive Differential-Correction Estimation in Nonlinear Problems," *IEEE*, **AC-10**, No. 4, 399–407 (October 1965).

Moyal, J. E.
[1] "Stochastic Processes and Statistical Physics," *J. Royal Stat. Soc.*, Ser. B. **11**, No. 2, 150–210 (1949).

Murphy, W. J.
[1] "Optimal Stochastic Control of Discrete Linear Systems with Unknown Gain," *IEEE*, **AC-13**, No. 4, 338–344 (August 1968).

Nash, R. A. and F. B. Tuteur
[1] "The Effect of Uncertainties in the Noise Coveriance Matrices on the Maximum Likelihood Estimate of a Vector," *IEEE*, **AC-13**, No. 1, 86–88 (February 1968).

Neal, S. R.
[1] "Linear Estimation in the Presence of Errors in Assumed Plant Dynamics," *IEEE*, **AC-12**, No. 5, 592–594 (October 1967).
[2] "Nonlinear Estimation Techniques," *IEEE*, **AC-13**, 705–708 (December 1968).

Nelson, E.
[1] *Dynamical Theories of Brownian Motion*, Princeton University Press, Princeton, New Jersey, 1967.

Neveu, J.
[1] *The Calculus of Probability*, Holden-Day, San Francisco, 1965.

Newton, G. C., L. A. Gould, and J. F. Kaiser
[1] *Analytical Design of Linear Feedback Controls*, Wiley, New Yrok, 1957.

Ogata, K.
[1] *State Space Analysis of Control Systems*, Prentice-Hall, Englewood Cliffs, New Jersey 1967.

Ohap, R, F, and A. R. Stubberud
[1] "A Technique for Estimating the State of a Nonlinear System," *IEEE*, AC-10, No. 2, 150–155 (April 1965).

Papoulis, A.
[1] *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 1965.

Parzen, E.
[1] *Stochastic Processes*, Holden-Day (San Francisco, 1962.
[2] "An Approach to Time Series Analysis," *Ann. Math. Stat.*, **32**, 951–989 (1961).
[3] "Extraction and Detection Problems and Reproducing Kernel Hiblert Spaces," *J. SIAM Control*, Ser. A., **1**, No. 1, 35–62 (1962).
[4] "Probability Density Functionals and Reproducing Kernel Hilbert Spaces," in M. Rosenblatt (Ed.), *Time Series Analysis*, Wiley, New York, 1963, pp. 155–169.

Park, S. K. and D. G. Lainiotis
[1] "Monte Carlo Study of the Optimal Non-Linear Estimator *Int. J. Control*, **16**, No. 6, 1029–1040 (1972).

Pawula, R. F.
[1] "Generalizations and Extensions of the Fokker-Planck-Kolmogorov Equations," *IEEE*, **IT-13**, No. 1, 33–41 (January 1967).

Pearson, J. B.
[1] "A Note on Nonlinear Filtering," *IEEE*, **AC-13**, No. 1 103–105 (February 1968).

Penrose, R.
[1] "A Generalized Inverse for Matrices," *Proc. Camb. Phil. Soc.*, **51**, pt. 3, 406–413 (1955).
[2] "On Best Approximate Solutions of Linear Matrix Equations," *Proc. Camb. Phil. Soc.*, **52**, pt. 3, 17–19 (1956).

Pindyck, R. S.
[1] "An Application of the Linear Quadratic Tracking Problem to Economic Stabilization Policy," *IEEE*, **AC-17**, No. 3, 287–300 (June 1973).

Polak, E. and E. Wong
[1] *Notes for a First Course on Linear Systems*, Van Nostrand New York. 1970.

Price, C. F.
[1] "An Analysis of the Divergence Problem in the Kalman Filter," *IEEE*, **AC-13**, No. 6, 699–702 December (1968).

Prohorov, Y, V. and Yu, A, Rozanov
[1] *Probability Theory*, Springer, New York, 1969.

Reid, W. T.
[1] "A Matrix Differential Equation of Riccati Type," *Am. J. Math.*, **68**, 237–246 (1946).
[2] "Solutions of a Riccati Matrix Differential Equation as Functions of Initial Values," *J. Math. Mech.*, **8**, No. 2, 221–230 (1959).

Reif, F.
[1] *Statistical and Thermal Physics*, McGraw-Hill, New York, 1965.

Rudin, W.
[1] *Principles of Mathematical Analysis*, McGraw-Hill, New York, 1964.
[2] *Real and Complex Analysis*, McGraw-Hill, New York, 1966.

Sage, A. P. and J. L. Melsa
[1] *Estimation Theory*, McGraw-Hill, New York, 1971.

Schlapfer, F. M.
[1] "Set-Theoretic Estimation of Distributed Parameter Systems," Ph.D. Thesis, MIT, 1970.

Schmeidler, W.
[1] *Linear Operators in Hilbert Space*, Academic, New York, 1965.

Schwartz, L. and E. B. Stear
[1] "A Computational Comparison of Several Nonlinear Filters," *IEEE*, **AC-13**, 83–86 (February 1968).

Schweppe, F. C.
[1] "Recursive State Estimation: Unknown but Bounded Errors and System Inputs," *IEEE*, **AC-13**, No. 1, 22–28 (February 1968).
[2] "Evaluation of Likelihood Functions for Gaussian Signals," *IEEE*, **IT-11** 61–70, (January 1965).

Sherman, S.
[1] "Non-Mean-Square Error Criteria," *IRE*, **IT-4**, No. 3, 125–126 (1958).
[2] "A Theorem on Convex Sets with Applications," *Ann. Am. Math. Soc.*, **26**, 763–767 (1955).

Skorokhod, A. V.
[1] *Studies in the Theory of Random Processes*, Addison-Wesley, Reading, Mass., 1965.

Smiriga, N.
[1] "An Integral Representation of a Continuous Linear Stochastic Process with Independent Pieces," *Theo. Prob. Appl.*, **14**, No. 1, 24–34 (1969).

Sneddon, I. N.
[1] *Elements of Partial Differential Equations*, McGraw-Hill New York, 1957.

Snyder, D. L.
[1] "The State Variable Approach to Continuous Estimations," Ph.D. Thesis, MIT, Cambridge, Mass., 1966.
[2] "The State Variable Approach to Analog Communication Theory," *IEEE*, **IT-14**, No. 1, 94–104 (January 1968).
[3] *The State Variable Approach to Continuous Estimation with Application to Analog Communication Theory, MIT* Press, Cambridge, Mass., 1969.

[4] "Estimation of Stochastic Intensity Functions of Conditional Poisson Processes," Monograph No. 128, Biomedical Computer Lab., Washington University, St. Louis, Mo., April 1970.
[5] "Detection of Nonhomogeneous Poisson Processes Having Stochastic Intensity Functions," Monograph No. 129, Biomedical Computer Lab., Washington University, St. Louis, Mo., April 1970.
[6] "Filtering and Detection for Doubly Stochastic Poisson Processes," *IEEE*, **IT-18**, 91–102 (1971).
[7] "Information Processing for Observed Jump Processes," *Info. Control*, **22**, 69–78 (1973).

Sorenson, H. W.
[1] "Kalman Filtering Techniques," C. Leondes (Ed.), *Advances in Control Systems*, Vol. III, Academic, 1966, pp. 219–292.
[2] "On the Error Behavior in Linear Minimum Variance Estimation Problems," *IEEE*, **AC-12**, No. 5, 557–562 (October 1967).
[3] "Controllability and Observability of Linear Stochastic Time-Discrete Control Systems," C. (Ed.), *Advances in Control Systems*, Vol. VI, Academic, New York, 1969, pp. 95–158

Sosulin, Yu. G.
[1] "Optimum Reception of Pulsed Signals in the Presence of Noise," *Radio Engr. Elec. Phys.*, **12**, 745–754 (1967).

Spiegel, M. R.
[1] *Real Variables*, McGraw-Hill, New York, 1969.

Srinivasan, K.
[1] "State Estimation by Orthogonal Expansion of Probability Distribution," *IEEE*, **AC-15**, No. 1, 3–10 (February 1970).

Stratonovich, R. L.
[1] "Conditional Markov Processes," Theory *Prob.* Appl., **5**, No. 2, 156–178 (1960).
[2] "On the Differential Equations Satisfied by Conditional Probability Densities of Markov Processes, with Applications," *J. SIAM Control*, Ser. A., **2**, No. 1, (1964).
[3] "A New Representation for Stochastic Integrals and Equations," *J. SIAM Control*, **4**, No. 2 (1966).
[4] "Detection and Estimation of Signals in Noise When One or Both are Non-Gaussian," *Proc. IEEE*, **58**, No. 5, 670–679 (May 1970).

Stratonovich, R. L. and Yu. G. Sosulin
[1] "Optimal Detection of a Markov Process in Noise," *Engr. Cyber.*, **2**, 7–19 (November-December 1964).
[2] "Optimum Detection of Diffusion Processes in White Noise," *Radio Engr. Elec. Phys.*, **10**, 704 (1965).
[3] "Optimum Reception of Signals in Nongaussian Noise," *Radio Engr. Elec. Phys.*, **11**, 497–507 (1966).

Striebel, C. T.
[1] "Partial Differential Equations for the Conditional Distribution of a Markov Process Given Noisy Observations," *J. Math. Anal. Appl.*, **11**, 151–159 (1965).

Swerling, P.
[1] "Topics in Generalized Least Squares Signal Estimation," *J. SIAM Appl. Math.*, **14**, No. 5, 998–1031 (September 1966).
[2] "Classes of Signal Processing Procedures Suggested by Exact Minimum Mean Square Error Procedures," *J. SIAM Appl. Math.*, **14**, No. 6, 1199–1224 (November 1966).
[3] "Modern State Estimation Methods from the Viewpoint of the Method of Least Squares," *IEEE*, **AC-16**, No. 6, 707–719 (December 1971).

Taylor, A. E.
[1] *Functional Analysis*, Wiley, New York, 1958.

Tzafestas, S. G. and J. M. Nightingale
[1] "Optimal Filtering, Smoothing and Prediction in Linear Distributed-Parameter Systems," *Proc. IEEE*, **115**, No. 8, 1207–1212 (August 1968).
[2] "Concerning Optimal Filtering Theory of Linear Distributed-Parameter Systems," *Proc. IEEE*, **115**, No. 11, 1737–1742 (November 1968).
[3] "Maximum-Likelihood Approach to the Optimal Filtering of Distributed-Parameter Systems," *Proc. IEEE*, **116**, No. 6, 1085–1093 (June 1969).

Uhlenbeck, G. E. and L. S. Ornstein
[1] "On the Theory of the Brownian Motion," *Phys. Rev.* **36** 823–841, (September 1, 1930).

Van Trees, H. L.
[1] *Detection, Estimation, and Modulation Theory, Vol.* I, Wiley, New York, 1968.
[2] *Detection, Estimation, and Modulation Theory*, Vol. II, Wiley New York, 1971.
[3] *Detection, Estimation, and Modulation Theory*, Vol. III, Wiley, New York, 1971.
[4] "Applications of State-Variable Techniques in Detection Theory," *Proc. IEEE*, **58**, No. 5, 653–669 (May 1970).

Ventzel, A. D.
[1] "On Equations of the Theory of Conditional Markov Processes," *Theor. Prob. Appl.*, **10**, 357–361 (1965).

Viterbi, A. J.
[1] "Phase-Locked Loop Dynamics in the Presence of Noise by Fokker-Planck Techniques," *Proc. IEEE*, **51**, 1737–1753 (December 1963).

Vulikh, B. Z.
[1] *Introduction to Functional Analysis*, Addison-Wesley, Reading, Mass., 1963.

Wang, M. C. and G. E. Uhlenbeck
[1] "On the Theory of the Brownian Motion II," *Rev. Modern Phys.*, **17**, Nos. 2 and 3, 323–342 (April-July 1945).

Wax, N.
[1] *Selected Papers on Noise and Stochastic Processes*, Dover, New York, 1954.

Wiener, N.
[1] *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, MIT Press, Cambridge, Mass., 1949.
[2] *Nonlinear Problems in Random Theory*, MIT Press, Cambridge, Mass., 1958.

Willems, J. C.
[1] *The Analysis of Feedback Systems*, MIT Press, Cambridge, Mass., 1970.

Wong, E.
[1] "Two-Dimensional Random Fields and Represnetation of Images," *SIAM J. Appl. Math.*, 756–770 (July 1968).
[2] *Stochastic Processes in Information and Dynamical Systems*, McGraw-Hill, New York, 1971.

Wong, E. and M. Zakai
[1] "On the Convergence of Ordinary Integrals to Stochastic Integrals," *Ann. Math. Stat.*, **36**, 1560–1564 (1965).

Wonham, W. M.
[1] "Some Applications of Stochastic Differential Equations to Optimal Nonlinear Filtering," *J. SIAM Control*, **2**, 347–369 (1965).
[2] "On the Separation Theorem of Stochastic Control," *J. SIAM Control*, **6**, No. 2, 312–326 (1968).
[3] "Random Differential Equations in Control Theory," in A. T. Bharucha-Reid (Ed.), *Probabilistic Methods in Applied Mathematics*, Academic, New York, 131–212 (1970).
[4] "On a Matrix Riccati Equation of Stochastic Contorl," *J. SIAM Control*, **6**, No. 4, 681–697 (1968).

Wozencraft, J. M. and I. M. Jacobs
[1] *Principles of Communication Engineering*, Wiley, New York, 1965.

Yaglom, A. M.
[1] "Second-Order Homogeneous Random Fields," *Proc. Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. II, University of California Press, Berkeley, 1961, pp. 593–620.
[2] *An Introduction to Stationary Random Functions*, Prentice-Hall, Englewood Cliffs, N.J., 1962.
[3] "Some Classes of Random Fields in n-Dimensional Space, Related to Stationary Random Processes," *Theory. Prob. Appl.* **2**, No. (1957).

Zadeh, L. A. and C. A. Desoer
[1] *Linear System Theory*, McGraw-Hill, New York, 1963.

Zemanian, A. H.
[1] *Distribution Theory and Transform Analysis*, McGraw-Hill, New York, 1964.