SYSTEMS RESEARCH CENTER
CASE WESTERN RESERVE UNIVERSITY
UNIVERSITY CIRCLE
CLEVELAND, OHIO 44106
APRIL, 1970

APPROXIMATIONS IN OPTIMAL NONLINEAR FILTERING

BARRY WILLIAM LICHT

SRC 70-1

APPROXIMATIONS IN OPTIMAL

NONLINEAR FILTERING

by

Barry William Licht

ABSTRACT

The infinite-dimensional nature of the formal solution to the continuous optimal nonlinear filtering problem requires that appropriate approximations be introduced to the solution before a useful finite-dimensional filter can be arrived at. A method is presented for computing a nonlinear filter's upper performance-bound, which can be used to establish an absolute measure of suboptimal filter effectiveness. The method is based on a numerical evaluation of Bucy's representation theorem and is utilized in the studies of five relatively simple nonlinear filtering problems. For each of these problems the relinearized Kalman filter proves to be the most effective filter of those studied.

Improper modelling assumptions and inaccurate parameter estimates also contribute to the less than optimal performance of a nonlinear filter. The question of filter performance sensitivity to modelling errors is considered and two sensitivity computing techniques are presented and illustrated by examples.

A nonlinear filter is developed to estimate, on-line, the time-delay of a dynamic process.

ACKNOWLEDGEMENTS

The day of the totally self-reliant man having sadly passed us by, I wish to express my sincere appreciation to those people and institutions upon whom and which I have most heavily relied during the course of this research.

My thanks to Dr. Sanjoy K. Mitter under whose direction this study was conducted. His gentle prodding was invaluable to keeping the course of this effort on target.

To the Systems Research Center, I give thanks for the excellent facilities and the stimulating environment it provided during my years of graduate study. The financial assistance provided by Uncle Sam in the form of an N.D.E.A. fellowship is greatly appreciated.

To my parents, Mr. and Mrs. Henry A. Licht, I can only hope to express my appreciation for their many years of unquestioning trust and understanding. I am sure my prolonged stay in college must have at times appeared to them to be a career in itself.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

PRELIMINARIES

1.1 Basic Concepts

Today the state identification, state estimation, parameter estimation, and parameter tracking problems are each a familiar subject of study in the field of modern control theory. Many novel and apparently unrelated solutions to the above problems have been proposed since the concept of state-space was first introduced to control theory. Under certain restrictions when the solutions to these problems are to be carried out in a real-time/on-line environment, a unifying theory applicable to the solutions of all the above mentioned problems exists in the very sophisticated theory of optimal nonlinear filtering.

Applications of optimal nonlinear filtering theory are restricted to those in which the dynamic systems involved possess a known structure (i.e., a differential or difference equation description) and are acted upon by external influences (i.e., initial conditions and forcing functions) whose statistical descriptions are known or can be appropriately approximated. Clearly, an extensive modelling effort and random signals analysis are necessary prerequisites to any actual application of optimal nonlinear filtering, especially if optimality is indeed a necessity. In fact, probably the most significant benefit derived from the application of optimal filtering theory to an actual system is that it forces the system designer to truly understand the system and in addition to make use of every shred of information about that system available to him. Sophisticated estimation and control techniques which purport to yield optimal performance without requiring knowledge about the system's structure and/or disturbance statistics must be considered suspect with regards to not only the technique's optimality but also its practicality.

In an abstract sense, the general optimal filter may be defined as that which

P1 given a state-space representation of a dynamic system, all pertinent statistical information regarding the stochastic nature of that system, and a set of uncertain observations (Z) functionally related to the system's state vector (x) over some interval of time ending at the present, produces an (in some sense) optimal estimate of the system's state vector at the present.

Without any further restrictions than those imposed in P1 one may conclude from probability theory that the general optimal filter must in some way continuously compute the conditional probability density of the state vector given the observations (p(x/Z)) . No matter what particular criterion for optimality is chosen, the optimal estimate will be a direct function of the conditional probability density p(x/Z) . As examples, the maximum likelihood optimal estimator is given as $x_{mlh} \rightarrow \max_x p(x/Z)$, and the least-squares, or minimum variance, optimal estimator is given as the conditional mean

$$\hat{x} \triangleq \int_X x \; p(x/Z) \; dx$$

where X is the domain of definition of the state space.

In theoretically oriented disciplines, the transitions from abstraction to reality are rarely smooth ones. Many assumptions and simplifications must be employed before a workable solution to a real problem can be arrived at. In surveying the literature dealing with the optimal filtering problem, the pragmatical person would soon recognize that a sizeable gulf exists between the conceptually pleasing solutions to this problem proposed by theoreticians and those solutions deemed acceptable by the practicing engineer. But engineering is the true science of compromise and we introduce at this point the first of many such compromises, with the hope being that the final result of these compromises will still be applicable to a large class of physical problems.

For the sake of solubility the general optimal filtering problem as formulated in this thesis will be restricted to that of

<u>P2</u> determining the "best" estimate of the state vector x at time t given the observations

$$Z_t \triangleq \{z(s) : t_0 \leq s \leq t\}$$

where

$$z(t) = h(x(t),t) + v(t) \tag{1.1}$$

and $x(t)$ is an n-vector stochastic process satisfying

$$\frac{dx}{dt} = f(x(t),t) + g(x(t),t)w(t) \tag{1.2}$$

with the a'priori probability density $p(x(t_0))$ known. In P2, $z(t)$ is an m-vector of observation variables, $h(x,t)$ is an m-vector of observation functions, $v(t)$ is an m-vector of independent Gaussian white-noises with variance matrix R, and zero mean, $f(x,t)$ is a n-vector of state functions, and $w(t)$ is a q-vector of zero-mean Gaussian white-noises independent of $v(t)$ and possessing an identity variance matrix.

A choice of Gaussian white-noises for the $v(t)$ and $w(t)$ random process in P2 may seem objectionable and quite restricting to someone not familiar with the intricacies of random signal theory. But such a choice not only makes possible the theoretical solution of the optimal nonlinear filtering problem but also provides a realistic approximation to a large class of stochastic processes found in nature. Physicists, in fact, have for years been modelling physical stochastic processes as idealized white-noises.

The assumption of additive observation noise in the observation equation (1.1) could also be suspected of being not general enough for wide application of the theory. Again, however, no general solution is presently available to the more general filtering problem in which the observation equation satisfies

$$z(t) = h(x(t),v(t),t) \quad . \qquad (1.3)$$

Still, one could effectively argue that the majority of physical observation process noises would possess at least a partially additive component. Any remaining observation noise components in such processes could be suitably treated as additional state variables.

The optimal nonlinear filter problem specified by P2 defines in effect two separate but related problems. First is that of determining an effective model, or representation, of the system of concern (i.e., $f(x,t)$, $h(x,t)$, and $g(x,t)$) and all its appropriate statistics (i.e., $p(x(t_0))$, R) . Second is that of determining a practical solution to the filter problem itself. While the theoretician can in all conscience divorce these two problems from each other, an applications engineer interested primarily in applying filter theory to real problems must consider each problem equally and recognize the importance of both.

While basic concepts and abstract discussion are fine, examples last longer. Two such examples of potential applications of optimal nonlinear filters are mentioned here; the first dealing with the familiar missile tracking problem and the second having to do with the less familiar and perhaps more difficult basic oxygen steel (BOF) process dynamic control problem.

In attempting to intercept and destroy an incoming ballistic warhead, the intercept radar control system must be able to distinguish between the trajectories of harmless jamming bodies and those of actual warheads. For this reason not only must the position and velocity of a body be estimated but also the value of the ballistic coefficient (β) for that body, where β is defined as the ratio of the body's aerodynamic coefficient to its mass. Under the proper assumptions this problem can be formulated as a nonlinear filtering problem in the form of P2.

Assume that a ballistic body is being tracked by a single radar device which provides noisy estimates of the target's range (r), azimuth (α), and elevation (θ) and that the air density of the atmosphere can be modelled by the relation

$$\rho = \rho_0 \exp(-kz)$$

where z is the target's altitude in a right-handed cartesian coordinate system and ρ_0 and k are known constant parameters of the atmosphere. Then if the state vector is defined by the relation

$$x = (x_1, x_2, x_3, x_4, x_5, x_6, x_7)^T \equiv (x, y, z, \dot{x}, \dot{y}, \dot{z}, \beta)^T$$

the state space model may be given as

$$\begin{aligned}
\frac{dx_1}{dt} &= x_4 \\
\frac{dx_2}{dt} &= x_5 \\
\frac{dx_3}{dt} &= x_6 \\
\frac{dx_4}{dt} &= -x_7\rho_0\exp(-kx_3)(x_4^2 + x_5^2 + x_6^2)^{1/2} x_4 + w_1(t) \\
\frac{dx_5}{dt} &= -x_7\rho_0\exp(-kx_3)(x_4^2 + x_5^2 + x_6^2)^{1/2} x_5 + w_2(t) \\
\frac{dx_6}{dt} &= -x_7\rho_0\exp(-kx_3)(x_4^2 + x_5^2 + x_6^2)^{1/2} x_6 + w_3(t) - g \\
\frac{dx_7}{dt} &= w_4(t)
\end{aligned} \tag{1.4}$$

where the process disturbance noises w_1, w_2, w_3, and w_4 are included to compensate for any imperfections in the model structure and g

represents the acceleration of gravity. The observation equations for this problem would be assumed to satisfy the following:

$$z_1(t) = (x_1^2 + x_2^2 + x_3^2)^{1/2} + v_1(t)$$

$$z_2(t) = \tan^{-1}(x_3/(x_1^2 + x_2^2)^{1/2}) + v_2(t) \qquad (1.5)$$

$$z_3(t) = \tan^{-1}(x_2/x_1) + v_3(t)$$

where z_1, z_2, and z_3 represent the radar's estimates of respectively, the range, azimuth, and elevation of the target.

Note that the target tracking problem as represented by Equations (1.4) and (1.5) is in the general form of P2. Of course, the a'priori probability density for the state vector would have to be estimated before the optimal nonlinear filter theory could be applied to this problem's solution. Alternatively, a suitable number of statistics of the state vector at the moment of target detection could be substituted for the initial joint probability density of the state vector. This difficulty is inherent to all nonlinear filtering problems and is a nontrival one.

While aerospace applications of optimal filtering theory are well known, industrial applications of the theory have for the most part been lacking, especially in the area of process control. To stimulate interest in this area consider, briefly, how the theory of nonlinear filtering might be applied to the solution of the BOF process dynamic control problem.

The relatively short blow-time (approximately twenty-five minutes) of the BOF process makes it desirable to place the critical phases of the process under automatic control. An especially critical phase of the process occurs during the final five minutes of the blow when the steel temperature and carbon content are to be carefully regulated. Unfortunately, the instrumentation available makes it impossible to measure these variables directly. Instead, the carbon content and temperature of the steel during the final minutes of the melt must be inferred from available measurements of flue-gas chemical composition and temperature and from a single bomb-calorimeter temperature measurement of the steel batch five minutes before completion of the oxygen blow. A lack of quality instrumentation, the need for indirect measurement of certain process variables, and the requirement for precise control of these variables makes the BOF process dynamic control problem appear ripe for the application of optimal nonlinear filtering.

A nonlinear filter implemented as part of a BOF process dynamic control system would in effect integrate in real-time the separate measurements from the flue-gas instruments and the bomb-calorimeter and yield a statistically optimal estimate of the steel's carbon content and temperature during at least the final five minutes of the blow. The requirements for a successful application of the optimal nonlinear filter to the BOF process are many. For example, an accurate state-space representation of the dynamics of the BOF chemical and thermal reaction processes must be available. Also, the statistics of the state vector at the instant the filter is initialized must be determined. These two requirements alone make it necessary to carry out an extensive BOF process modelling effort before even considering an application of a non-linear filter to the process. Tribus and Kornblum [45] have considered this problem in developing an endpoint control system for the BOF process.

1.2 Historical Perspective[1]

There exist in the literature two fundamental approaches to the solution of nonlinear filtering problems. These two approaches will be referred to here as the statistical (also variational or least-squares) and probabilistic viewpoints of filtering theory. Each approach possesses advantages and disadvantages over that of the other but it is with the probabilistic approach that this thesis is primarily concerned.

Those adhering to the statistical viewpoint of nonlinear filtering thoery hold that it is a disadvantage and often even an impossibility to require probabilistic information regarding the observation and disturbance processes for the solution to any particular filtering problem. For this reason the statistical approach, as outlined by Lee [33], assumes the process to be specified by the equations

$$\frac{dx}{dt} = f(x,t) + g(x,t)u(t) \tag{1.6}$$

and

$$z(t) = h(x,t) + (\text{measurement error}) \tag{1.7}$$

where $u(t)$ represents any unknown function, as compared with the Gaussian white-noise process $w(t)$ assumed in Equation (1.2) of P2. Similarly, no particular knowledge about the observation error process in Equation (1.7) is presumed other than that it exists.

---

1 Bracketed numbers refer to references listed at the end of this thesis.

The statistical method of solution proceeds by finding that function u(t) which minimizes the integral[1]

$$\int_0^t [||z(t) - h(x,t)||_Q^2 + ||\frac{dx}{dt} - f(x,t)||_W^2]dt \qquad (1.8)$$

subject to the constraint Equation (1.6), where Q and W are positive semidefinite matrices representing weighting functions for the measurement errors and disturbance inputs. Thus the filtering problem is reduced to one in the calculus of variations and one which may be conveniently handled by the many existing techniques of optimal control theory. For example, Bellman, et al. [33] have applied the theory of invariant imbedding to the sequential solution of the two-point boundary value problem which results from the Euler-Lagrange equations for this variational problem without the disturbance process u(t) present. Detchmendy and Stridhar [14] followed a similar procedure in treating the more general disturbed case.

The probabilistic (sometimes referred to as Bayesian) approach to nonlinear filtering, while requiring more a'priori process information than that of the statistical approach, is able to provide an (in some statistical sense) optimal solution to a large class of problems for which the statistical approach can not. Assuming the process to be specified by the relations expressed in <u>P2</u>, the probabilistic approach utilizes all available a'priori information about the process prior to an output observation and the

---

1 $||x||_A^2$ denotes the quadratic form associated with the symmetric matrix A .

observation itself to determine in a sequential fashion an estimate of the a'posteriori joint probability density for the state vector; i.e., $p(x/Z)$. Differential (or in the discrete case, difference) equations satisfied by certain parameters of this probability density function are then derived and constitute an easily implemented filtering algorithm.

Over the past decade a very elegant theory of optimal non-linear filtering has been developed from the probabilistic viewpoint of the filtering problem. In 1958 Stratonovich presented his views on the theory of optimal nonlinear filtering at a seminar on probability theory at Moscow University. Approximately one year later the first major work [40] applicable to the solution of the optimal filter problem appeared and in it Stratonovich introduced the concept of conditional Markov processes; conditional with respect to the output observation data for the system. A dynamical equation for the conditional density function when the disturbance and measurement noises are both jointly Gaussian and white was derived by Stratonovich [41] in this work and while his results were considered in error by many [18,27] Stratonovich has denied this [42] , stating that the misunderstanding arises from his treatment of the problem utilizing physical (i.e., band-limited) white-noise rather than mathematical, or pure, white-noise.[1] A recent work of Stratonovich [43] contains all his original papers

1 Stratonovich's explanation is now generally accepted and no controversy presently exists.

and provides a comprehensive study of conditional Markov processes and their application to optimal control theory.

Stratonovich's pioneering efforts in the area of optimal nonlinear filtering theory were followed by those of Kushner [27-31], who derived for the Gaussian disturbance and measurement noise case a partial-differential-integral equation which is satisfied by the conditional density function of the state vector and which reduces to the Fokker-Planck equation as the variance of the observation noise becomes infinite. This equation was used by Kushner [27,28] to determine differential equations for the conditional expectation of functions of the state vector and these and other results of the theory were later given a more rigorous and formal justification by Kushner [29]. It was also pointed out by Kushner how the general optimal nonlinear filter, if such a device could be constructed, would consist of an infinite set of coupled differential equations each governing the evolution of a moment, or parameter, of the conditional density function and each containing an observation function driving term.

Bucy's [4,5] approach to the optimal nonlinear filtering problem differed from that of Kushner by being more mathematical and less intuitive, but the results of his efforts for the Gaussian disturbance and measurement noise case were identical to those of Kushner. An important intermediate result of Bucy's work was that of a representation theorem which demonstrates how the conditional, or a'posteriori, density function at some instant of time can be represented as a direct function of the a'priori density, $p(x(t_0))$, and the conditional expectation of an exponential functional of the observation data over the time interval of t_0 to t. As demonstrated in a following section of this thesis, a physical interpretation of this theorem's result suggests a convenient method for obtaining a numerical estimate of the conditional probability density $p(x(t)/Z)$. The relatively straightforward proof of Bucy's representation theorem can be found in Bucy and Joseph's [6] recent text as well as in the thesis of Schwartz [36].

Wonham's [47] work provides an excellent introduction to the theory of optimal filtering and includes a numerical comparison of the performance of an optimal nonlinear filter with that of a Wiener filter. Wonham utilized a representation theorem from Doob [15] to determine a differential equation for the conditional probability of a scalar Poisson process with linear state measurements obscured by additive Gaussian white-noise.

A unified approach to the optimal nonlinear filtering problem which subsumes the results of Stratonovich, Kushner, Bucy, and Wonham was presented by Fisher [16]. However, this particular work is a quite formidable one and would be of questionable value to the individual primarily concerned with applications of the theory.

No historical synopsis of the theory of optimal nonlinear filtering would be complete without properly crediting Ito [22], who was primarily responsible for the development of the stochastic calculus, the mathematical basis upon which the theory of optimal filtering was developed. In the same light, Doob's fundamental work [15] must be considered invaluable to the development of the theory.

Since it was first shown that the formal solution to the optimal nonlinear filtering problem as specified in P2 consists of a partial-differential-integral equation satisfied by the conditional probability density $p(x/Z)$ and an easily derived, associated infinite set of differential equations governing the evolution of the parameters (e.g., moments) of that density function, most research efforts in the field have been concerned with determining effective finite-dimensional approximations to this inherently infinite-dimensional solution. Many diverse methods of solution to this approximation problem have appeared in the literature [2,9,31,37] and the most prominent of them are presented in some detail in Chapter III of this thesis. A criticism common to most of these efforts is that the effectiveness of each technique when evaluated, if at all, was determined relative to some other approximately optimal filtering technique. The question which naturally arises is how approximate are the approximations? This question is both a formidable and relevant one and its answer is a major concern of this thesis.

1.3 Thesis Content

In Chapter II the basic elements of the fundamental theory of optimal nonlinear filtering are introduced. The level of rigor in this introduction is purposefully not high with the intent being to present in as clear a fashion as possible the highlights of the theory and their relation to each other. Beginning with the concept of white-noise and brownian motion, the presentation concludes with the development of the infinite set of differential equations satisfied by the statistical moment parameters of the a'posteriori probability density function, $p(x/Z)$. Consideration is given the nonlinear filtering problem for the continuous process, discrete observation case.

The major concern of Chapter III is with that of the many available finite-dimensional methods of approximation to the optimal filtering solution. Only the most prominent and/or potentially useful techniques are discussed including such familiar methods as the relinearized and second-order Kalman filters and such not so familiar methods as the assumed-form density function filter.

In the literature the principal method of determining the performance of an approximately optimal nonlinear filtering technique has been to compare for a specific example(s) the performance of the filter with that of some other approximate filters. In Chapter IV, an absolute, rather than relative, method of performance comparison is presented and is based upon Bucy's representation theorem, which is given an intuitively pleasing interpretation. The chapter concludes with the results of five computational experiments in nonlinear filtering which demonstrate the usefulness of the absolute performance comparison method.

The performance degradation of approximately optimal nonlinear filters caused by the finite-dimensional approximation to the optimal filtering solution is just one of many contributors to the suboptimal performance of nonlinear filters, and not necessarily the major one. Uncertainties in process model structure, a'priori statistics, and observation noise processes all produce a fall-off in performance from that predicted by theory. In Chapter V a technique is investigated for determining the performance sensitivity of nonlinear filters to these various error sources. A Monte Carlo simulation method is also used to study the error sensitivities of a nonlinear filter utilized as a linear parameter estimator.

A particular application of optimal nonlinear filtering theory is considered in Chapter VI; that of estimating the time-delay of a first-order linear process from noisy measurements of the step response of the process. A summary and extensions of the work presented in this thesis are given in the final chapter.

We claim that the idea in Chapter IV of computing upper performance-bounds for nonlinear filters utilizing the Bucy representation is unique and relevant and that the results of the computational experiments discussed in this chapter shed much new light on the practical aspects of optimal nonlinear filtering. In particular, it is indicated by examples that the relinearized Kalman filter yields essentially optimal performance for a large class of common nonlinear filtering problems.[1] Counter examples are presented to the hypothesis that the relinearized Kalman filter provides the best general solution to the nonlinear filter problem and emphasis is given for the first time to the influence of the higher than second-order conditional moment parameters on filter performance. The method of approximate filtering based on assumed-form density functions, as introduced in Chapter III and applied in Chapter VI , is felt to be especially relevant to the solution of the filtering problem, and the use of uniform-density functions is unique.

1 These conclusions apply only to the continuous measurement nonlinear filtering problems.

1.4 Notational Conventions

In this thesis we have made no attempt to distinguish between vectors and matrices by the use of special notation. Only in those cases where the intent is not clear is it stated whether a variable is a vector or a matrix. Any subscripted variables denote scalar elements of a vector or matrix; e.g.,

$$x_i = i^{\underline{th}} \text{ element of a vector } x$$

$$x_{ij} = i - j^{\underline{th}} \text{ element of a matrix } x \ .$$

In addition, a matrix consisting of a certain set of elements may be denoted by square brackets in the following fashion:

$$[x_{ij}] = \text{the matrix with elements } x_{ij} \ .$$

Two standard symbolic notations are used interchangeably throughout this thesis to signify the expectation of a random variable. For example, if $\phi(x)$ is a function of the random variable x having a probability density $p(x)$, then $E\{\phi(x)\}$ and $\widehat{\phi(x)}$ are both used to denote the expected value of the random function $\phi(x)$, which is given formally as

$$\int \phi(x)p(x)dx \overset{\Delta}{=} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(x)p(x)dx_1 dx_2 \ldots dx_n \ .$$

Conditional expectations and probability densities are indicated in the usual manner by a slash between the random variable and the conditioning information; e.g.,

$E\{\phi(x)/Z\}$ = expected value of $\phi(x)$ given the information Z related to x

$= \int\phi(x)p(x/Z)dx$.

The following list of symbols and their definitions are used throughout this thesis:

A^T	Transpose of a Matrix A
A^{-1}	Inverse of a Matrix A
$A^{1/2}$	"Square-Root" of a Positive Definite Matrix A
$w(t)$, $v(t)$	Gaussian White-Noise Process Vectors
$\nabla f(x)$	Matrix of First Partial-Derivatives of $f(x)$ w.r.t. the vector x
$\nabla^2 f(x)$	Matrix of Second Partial-Derivatives of $f(x)$ w.r.t. x
$\delta(\tau)$	Dirac Delta Function = $\begin{cases} 1, & \tau = 0 \\ 0, & \tau \neq 0 \end{cases}$
$\exp(\cdot)$	Exponential Function
$N(m,\Lambda)$	The Multivariate Gaussian (Normal) Distribution of Mean m and Covariance Λ
I	The Identity Matrix

CHAPTER II

OPTIMAL NONLINEAR FILTERING: BASIC THEORY

2.1 Introduction

The assumption of ideal white-noise sources in the basic optimal nonlinear filtering problem makes it possible to determine a mathematical solution to this problem but introduces certain mathematical difficulties to the theory not amenable to the rules of the ordinary calculus. Indiscriminate use of the dirac delta function, so common in engineering mathematics, is unacceptable to the theoretical solution of the optimal nonlinear filtering problem. We introduce in this chapter the basic results of the theoretical solution to this problem and precede this introduction with that of the stochastic calculus, a calculus developed initially by Ito [22] and the basis of the theory of nonlinear stochastic processes. In the interests of clarity, the level of presentation here is partly intuitive, but the reader should not let this detract from the relevance and fundamental validity of the results of this theory. Where possible, those individuals originally responsible for what follows are duly credited.

2.2 The Stochastic Calculus

Ito's [22] development of the stochastic calculus proceeded in a totally rigorous fashion and is highly recommended to those whose interests are primarily theoretical. For those whose interests are more applied than theoretical we present briefly the fundamental notions of the stochastic calculus and give, where possible, their relations to the theory of optimal nonlinear filtering.

2.2.1 White-Noise and Brownian Motion

Recall from P2 in Chapter I that the nonlinear filtering problem is concerned with processes modelled by the differential equations

$$\frac{dx}{dt} = f(x,t) + g(x,t)w(t) \tag{1.2}$$

and

$$z(t) = h(x,t) + v(t) \tag{1.1}$$

where $w(t)$ and $v(t)$ are ideal Gaussian white-noise process vectors of zero means and identity correlation matrices; i.e., in the $w(t)$ case,

$$E\{w(t)\} = 0 \tag{2.1}$$

and

$$E\{w(t)w(s)\} = I\ \delta(t-s) \quad . \tag{2.2}$$

The concept of ideal white-noise, while not a physically realistic one, has proven to be a mathematically useful one to the physicist and engineer. In the stochastic calculus, however, mathematicians have with good reason chosen to deal formally with the stochastic process defined as the integral of the ideal Gaussian white-noise process.

The brownian motion process is given as

$$b(t) = \int_0^t w(\gamma)d\gamma \tag{2.3}$$

where the initial value $b(0)$ is arbitrarily chosen to be zero, and $w(t)$ is ideal Gaussian white-noise. Consider the following:

From Eq's. (2.2) and (2.3),

$$E\{(b(t) - b(s))^2\} = E\{\int_s^t w(\zeta)d\zeta \cdot \int_s^t w(\gamma)d\gamma\}$$

$$= \int_s^t \int_s^t E\{w(\zeta)w(\zeta)\} \, d\gamma d\zeta$$

$$= \int_s^t \int_s^t \delta(\zeta - \gamma)d\zeta d\gamma = \int_s^t d\gamma$$

$$= |t - s| \quad . \tag{2.4}$$

In the limit, as times t and s are allowed to become arbitrarily close, Eq. (2.4) yields the intuitively troublesome relation

$$E\{(db(t))^2\} = dt \ . \tag{2.5}$$

One might conclude from Eq. (2.5) that it makes no sense to speak of an ideal white-noise process, $w(t)$, since

$$w(t) = \frac{db}{dt} \approx \frac{(dt)^{1/2}}{dt} = \frac{1}{(dt)^{1/2}}$$

implies the non-existence of such a process in the continuous case. For this reason, the stochastic calculus was developed strictly in terms of the brownian motion process, $b(t)$.

2.2.2 Stochastic Integrals

Surveying the literature dealing with the theory of optimal nonlinear filtering, one would generally find the process Eq. (1.2) written in its alternate form

$$dx(t) = f(x,t)dt + g(x,t)db(t) \quad (2.6)$$

Since b(t) is an ideal brownian motion process, Eq. (2.6) is a stochastic differential equation which must be considered as only a symbolic representation of the stochastic integral equation

$$x(t) - x(s) = \int_s^t f(x(\gamma),\gamma)d\gamma + \int_s^t g(x(\gamma),\gamma)db(\gamma) \quad (2.7)$$

where the stochastic integral as defined by Ito [22] is given as[1]

$$\underset{\max \Delta\gamma \to 0}{\text{l.i.m.}} \sum_{k=1}^{N-1} g(x(\gamma_k),\gamma_k)[b(\gamma_{k+1}) - b(\gamma_k)] \quad (2.8)$$

and has been shown [23] to have the following properties:

(i) $E\{\int_0^t g(\gamma)db(\gamma)\} = 0$

(ii) $E\{\int_0^t g_1(\gamma)db(\gamma) \cdot \int_0^t g_2(\xi)db(\xi)\} = \int_0^t E\{g_1(\gamma)g_2(\gamma)\}d\gamma$

(iii) $E\{\int_0^t g(\gamma)db(\gamma)/b(\xi)\ ,\ \xi \leq s < t\} = \int_0^s g(\gamma)db(\gamma)$

Stratonovich [42] defines a symmetric version of the stochastic integral as

---

1 If x_n is a sequence of random variables, then to say that x is the limit-in-the-mean (l.i.m.) of the sequence x_n implies that $\lim_{n \to \infty} E\{|x_n - x|^2\} = 0$.

$$\underset{\max \Delta\gamma \to 0}{\text{l.i.m.}} \sum_{k=1}^{N-1} g\left(\frac{\gamma_{k+1} + \gamma_k}{2}\right) [b(\gamma_{k+1}) - b(\gamma_k)] \tag{2.9}$$

which does not possess the useful properties (i), (ii), and (iii) but which does yield a stochastic calculus the rules of which are essentially identical to those of the ordinary calculus. In this thesis the stochastic integral will be defined in the Ito sense and as a result certain rules of the ordinary calculus will not apply.[1]

2.2.3 Ito's Differential Rule

One difficulty in applying the stochastic calculus is that the differential rule of the ordinary calculus does not hold true in its usual form. An intuitive derivation of Ito's differential rule can be easily achieved by a simple Taylor series expansion and an application of the brownian motion's second-order incremental property (i.e., $E\{(db)^2\} = dt$) . In particular, let $x(t)$ satisfy the stochastic differential equation (2.6) and let $\phi(x,t)$ be some regular scalar function of x and t . Then a truncated Taylor series expansion of $\phi(x,t)$ about x and t yields the relation

$$d\phi(x,t) = \frac{\partial\phi}{\partial t} dt + dx^T \nabla\phi(x) + \frac{1}{2} dx^T \nabla^2 \phi(x) dx \tag{2.10}$$

where $\nabla\phi(x)$ and $\nabla^2\phi(x)$ are respectively the vector of first partials and matrix of second partials of $\phi(x,t)$ w.r.t. x .

---

1 The Ito calculus is used in most of the non-Russian literature dealing with nonlinear filtering.

Then from Eq's. (2.6) and (2.10) we get by substitution

$$d\phi(x,t) = \frac{\partial\phi}{\partial t}\,dt + [f^T(x,t)dt + db^T g^T(x,t)]\nabla\phi(x) + \frac{1}{2} f^T(x,t)\nabla^2\phi(x)f(x,t)(dt)^2 + db^T g^T(x,t)\nabla^2\phi(x)f(x)dt + \frac{1}{2}db^T g^T(x,t)\nabla^2\phi(x)g(x,t)db \quad . \tag{2.11}$$

From the forementioned properties of brownian motion,

$$E\{dbdt\} = 0$$

and

$$E\{db_i db_j\} = \begin{cases} dt \ , & i = j \\ 0 \ , & i \neq j \end{cases} ,$$

Eq. (2.11) yields Ito's differential rule,

$$d\phi(x,t) = \frac{\partial\phi}{\partial t}\,dt + [f^T(x)dt + db^T g^T(x)]\nabla\phi(x,t) + \frac{1}{2}\,tr[g^T(x)\nabla^2\phi(x)g(x)]dt \tag{2.12}$$

The differential rule is commonly stated in its more familiar form

$$d\phi(x,t) = \{\frac{\partial\phi}{\partial t} + \mathcal{L}[\phi(x,t)]\}dt + db^T g^T(x)\nabla\phi(x) \tag{2.13}$$

where the differential operator $\mathcal{L}[\cdot]$ satisfies the defining relation

$$\mathcal{L}[\cdot] \stackrel{\Delta}{=} \sum_i f_i(x,t)\,\frac{\partial\cdot}{\partial x_i} + \frac{1}{2}\sum_{i,j}[g^T g]_{i,j}\,\frac{\partial^2\cdot}{\partial x_i \partial x_j} \; . \tag{2.14}$$

The $\mathcal{L}[\cdot]$ operator and its formal adjoint

$$\mathcal{L}^*[\cdot] = -\sum_i \frac{\partial[f_i(x,t)\ \cdot]}{\partial x_i} + \frac{1}{2}\sum_{i,j} \frac{\partial^2([g^Tg]_{i,j}\ \cdot)}{\partial x_i \partial x_j} \qquad (2.15)$$

are of fundamental importance to the study of stochastic processes satisfying stochastic differential equations (2.6) . Integrating by parts one may easily show that

$$\int \mathcal{L}[\phi(x)]\ p(x)dx = \int \phi(x)\ \mathcal{L}^*[p(x)]dx \qquad (2.16)$$

where $p(x)$ is a well behaved probability density function defined over the state space.

2.2.4 The Fokker-Planck Equation[1]

The Fokker-Planck equation has been a familiar one to the field of physics for a good many years and describes, in effect, the dynamic behavior of the joint probability density of the process $x(t)$ satisfying the stochastic differential equation (2.6) . Kailath and Frost [23] have shown recently how a simple application of Ito's differential rule can be utilized to derive the Fokker-Planck equation. The procedure follows.

Consider again the stochastic process $x(t)$ satisfying the equation

$$dx(t) = f(x,t)dt + g(x,t)db(t) \qquad (2.6)$$

and let[2]

$$\lambda(t) = \exp\ i\xi^T(x(t) - x(s))\ .$$

1 In mathematics the F-P equation is usually referred to as Kolmogorov's equation.

2 The letter i denotes the imaginary unit; $\sqrt{-1}$.

By Ito's differential rule

$$d\lambda(t) = \mathcal{L}[\lambda(t)]dt + db^T(t)g^T(x,t)i\xi\lambda(t)$$

or, equivalently,

$$\lambda(t) = 1.0 + \int_s^t \mathcal{L}[\lambda(\gamma)]d\gamma + \int_s^t db^T(\gamma)g^T(x,\gamma)i\xi\lambda(\gamma) \ . \tag{2.17}$$

Taking the expectations of both sides of Eq. (2.17) conditioned on the variable $x(s)$ yields the relation[1]

$$E\{\lambda(t)/x(s)\} = 1.0 + \int_s^t E\{\mathcal{L}[\lambda(\gamma)]/x(s)\}d\gamma$$

which upon differentiation gives

$$\frac{\partial E\{\lambda(t)/x(s)\}}{\partial t} = E\{\mathcal{L}[\lambda(t)]/x(s)\}$$

$$= \int \mathcal{L}[\lambda(t)]p(x(t)/x(s))dx(t) \ . \tag{2.18}$$

From the $\mathcal{L}[\cdot]$ operators' adjoint property expressed in Eq. (2.16), we get

$$\frac{\partial E\{\lambda(t)/x(s)\}}{\partial t} = \int\lambda(t) \ \mathcal{L}^*[p(x(t)/x(s))]dx(t) \ . \tag{2.19}$$

But the expression on the right side of Eq. (2.19) is simply the Fourier transform of $\mathcal{L}^*[p(x(t)/x(s))]$ w.r.t. $x(t)$, and the expectation $E\{\lambda(t)/x(s)\}$ is the Fourier transform $(\mathcal{F}\{\cdot\})$ of $p(x(t)/x(s))$, again w.r.t. $x(t)$. Hence,

$$\frac{\partial}{\partial t} \mathcal{F}\{p(x(t)/x(s))\} = \mathcal{F}\{ \ \mathcal{L}^*[p(x(t)/x(s)]\}$$

or, equivalently,

---

1 Note that $E\{db\} = 0$, accounting for the disappearance of the second integral term of Eq. (2.17) when the conditional expectations are taken.

$$\mathcal{F}\{\frac{\partial p(x(t)/x(s))}{\partial t}\} = \mathcal{F}\{ \mathcal{L}^*[p(x(t)/x(s))]\} , \qquad (2.20)$$

Inverse transformation of Eq. (2.20) yields the Fokker-Planck equation,

$$\frac{\partial p(x(t)/x(s))}{\partial t} = \mathcal{L}^*[p(x(t)/x(s))] . \qquad (2.21)$$

While solutions to the Fokker-Planck equation are almost nonexistent except for the simplest cases, useful differential equations for the expected values of functions of $x(t)$ can be easily derived. In particular, if $\phi(x)$ is some regular function of $x(t)$, then

$$E\{\phi(x)\} = \int \phi(x)\ p(x,t)dx$$

and

$$\frac{d\ E\{\phi(x)\}}{dt} = \int \phi(x)\ \frac{\partial p(x,t)}{\partial t}\ dx$$

$$= \int \phi(x)\ \mathcal{L}^*[p(x,t)]\ dx$$

which from the adjoint property of Eq. (2.16) becomes

$$\frac{d\ E\{\phi(x)\}}{dt} = \int \mathcal{L}[\phi(x)]\ p(x,t)\ dx$$

$$= E\{\mathcal{L}[\phi(x)]\} . \qquad (2.22)$$

Eq. (2.22) can be used to determine ordinary differential equations that are satisfied by the statistical moment functions of $p(x,t)$.

2.2.5 Modelling Considerations

Up to this point in this chapter, it has been tacitly assumed that the physical stochastic processes of concern can be satisfactorily modelled by a set of first-order nonlinear differential equations driven by ideal Gaussian white-noise. There are, however, inherent theoretical difficulties associated with the practice of representing a physical white-noise driven process as an ideal, or mathematical, white-noise driven process. One such difficulty has been identified by Wong and Zakai [46] and others [20,36].

Consider the physical process described by the vector differential equations

$$\frac{dx_i}{dt} = f_i(x,t) + \sum_{j=1}^{n} g_{ij}(x,t)\, u_j(t) \tag{2.23}$$

where the $u_j(t)$ are independent physical white-noise processes. It has been shown[1] that such a process should be modelled as a mathematical process satisfying the stochastic differential equations

$$dx_i = f_i(x,t)dt + \sum_{j=1}^{n} g_{ij}(x,t)db_j(t) + \frac{1}{2} \sum_{j,k}^{n} \frac{\partial g_{ij}(x,t)}{\partial x_k} g_{kj}(x,t)dt \tag{2.24}$$

where $w_j(t) = db_j/dt$ are independent, mathematical white-noises and the differentials in Eq. (2.24) are taken in the Ito sense[2]. In addition it may be shown that the variance of the ideal white-noises $w_j(t)$ should be chosen such that the areas under the

---

1 See, for example, Reference 45.

2 Note that if the $g_{ij}(x,t)$ are not functions of x, then the physical and mathematical model equations are identical.

autocorrelation curves of $w_j(t)$ and $u_j(t)$ are equal. In this thesis the terms in Eq. (2.24)

$$\frac{1}{2} \sum_{j,k}^{n} \frac{\partial g_{ij}(x,t)}{\partial x_k} g_{kj}(x,t)dt$$

will be referred to as the Ito correction terms.

In nonlinear filtering the inverse modelling problem also exists in that if a stochastic differential equation description of a nonlinear filter has been derived, then a set of equivalent ordinary differential equations must be determined before the filter can be correctly implemented. The Ito correction terms in the filter's stochastic differential representation must be accounted for by subtracting them from the filter's corresponding ordinary differential equation representation.

2.3 Theoretical Solution to the Optimal Nonlinear Filtering Problem

The stochastic calculus discussed briefly in the previous section can be applied to the solution of the optimal filtering problem as stated in P2 of Chapter I . It will be shown that this solution takes the form of a stochastic partial-differential-integral equation which is satisfied by the probability density of the system's state vector conditioned upon the observation data for the system (i.e., $p(x/Z_t)$). A representation theorem from Bucy [4-6] provides a formal means for deriving this equation and also suggests a numerical method for computing the conditional density $p(x/Z_t)$. The moment parameters of this density function are easily shown to satisfy an infinite set of coupled, nonlinear stochastic differential equations driven by the observation functions, $z(t)$.

2.3.1 Bucy's Representation Theorem

Suppose that the stochastic process $x(t)$ satisfies the stochastic differential equation

$$dx(t) = f(x,t)dt + g(x,t)db(t) \tag{2.6}$$

and that the observations $\zeta(t)$[1] are available with

$$d\zeta(t) = h(x,t)dt + R^{1/2}(t)da(t) \tag{2.25}$$

where the a'priori probability density function $p(x(t_o))$ is known and $w(t) = db/dt$ and $v(t) = da/dt$ are independent unity-variance ideal Gaussian white-noise processes. Then if the observation information is represented as

$$\mathcal{Z}_t = \{\zeta(s) : t_o \leq s \leq t\} .$$

Bucy [6] has proved the following representation

Theorem (Bucy):

$$p(x(t)=X/\mathcal{Z}_t) = \frac{E\{\exp\phi/x(t)=X\}\ p(x(t)=X)}{E\{\exp\phi\}} \tag{2.26}$$

1 Note that $z(t) = d\zeta/dt$, so that $\zeta(t)$ is formally the integral of the $z(t)$ process.

where

$$\phi = \int_{t_o}^{t} h^T(x(s),s)R^{-1}(s)d\zeta(s) - \frac{1}{2} \int_{t_o}^{t} h^T(x(s),s)R^{-1}(s)h(x,s)ds \tag{2.27}$$

$$dx(s) = f(x(s),s)ds + g(x(s),s)db(s) \; ; \; x(t) = X$$

and

$$E\{\exp\phi/X\} = \lim_{m \to \infty} \int . . . \int \exp\{\phi\}p(x(s_1), \ldots ,$$

$$x(s_{m-1}),x(t))dx(s_1) \ldots dx(s_{m-1})$$

$$t_o = s_1 < s_2 . . . < s_m = t$$

with the observations $\mathcal{Z}_t$ held fixed in Eq. (2.27).

E.O.T.

While Mortenson [35] has proven Bucy's representation theorem using function space concepts, Bucy's proof presented in Bucy [6] and in Schwartz [36] is more straightforward then Mortenson's and requires only Baye's rule and Ito's definition of the stochastic integral. The probability density function $p(x(t))$ in Eq. (2.26) represents the a'priori, or unconditional, density of the state vector and, as demonstrated in section 2.2.4, satisfies the Fokker-Planck equation. Note also that the denominator of the right side of Eq. (2.26) is given as the integral of the numerator over the state space; i.e.,

$$E\{\exp\phi\} = \int E\{\exp\phi/X\} \; p(X) \; dX \; .$$

The Bucy representation is given a physical interpretation in Chapter IV of this thesis where it is demonstrated how the representation theorem may be utilized to obtain numerical estimates of $p(x/Z_t)$.

2.3.2 Conditional Version of the Fokker-Planck Equation

For pragmatical reasons, a nonlinear filter must be a sequential device and for this reason Bucy's representation for $p(x/Z_t)$ can not be directly implemented as an optimal nonlinear filter. Instead, a differential expression for this conditional density function is required.

By multiple applications of Ito's differential rule to the Bucy representation (Eq. (2.26)) Bucy has shown that the conditional probability density of the state vector, $p(x/Z_t)$, satisfies the stochastic partial-differential-integral equation

$$dp(x/Z) = \overset{*}{\mathcal{L}}[p(x/Z)]dt + p(x/Z)\ (d\zeta(t) - \widehat{h(x,t)}dt)^T \cdot R^{-1}(h(x,t) - \widehat{h(x,t)}) \qquad (2.28)$$

where

$$\widehat{h(x,t)} = \int h(x,t)p(x/Z)dx \ .$$

Comparing Eq's. (2.21) and (2.28) one may note the similarity between the Fokker-Planck equation and its conditional counterpart, Eq. (2.28). In fact, Eq. (2.28) reduces to the Fokker-Planck equation as the observation noise variance becomes infinite. Like the Fokker-Planck equation, its conditional version is difficult, if not impossible, to solve directly but can be utilized to obtain stochastic differential equations for parameters of the conditional probability density $p(x/Z)$.

2.3.3 Conditional Moment Equations

If $\phi(x)$ is some regular function of x where

$$dx = f(x,t)dt + g(x,t)db(t) \qquad (2.6)$$

and

$$d\zeta = h(x,t)dt + R^{1/2}da(t) \ , \qquad (2.25)$$

then Kushner [27-29] has shown how the differential equation for the conditional expectation of $\phi(x)$ may be easily obtained from Eq. (2.28) and the defining relation

$$\widehat{\phi(x)} \equiv E\{\phi(x)/\mathcal{Z}_t\} \overset{\Delta}{=} \int\phi(x)p(x/\mathcal{Z}_t)dx \ . \qquad (2.29)$$

Taking the differential of $\widehat{\phi(x)}$ yields

$$d(\widehat{\phi(x)}) = \int\phi(x)\ (dp(x/\mathcal{Z}_t))dx$$

which from Eq. (2.28) becomes

$$d(\widehat{\phi(x)}) = \int\phi(x)\ [\mathcal{L}^*[p]dt + p(h - \hat{h})R^{-1}(d\zeta - \hat{h}\,dt)]dx \ . \qquad (2.30)$$

Applying the adjoint property of the $\mathcal{L}[\cdot]$ operators to Eq. (2.30) we get

$$d(\widehat{\phi(x)}) = dt\int\mathcal{L}[\phi(x)]p(x)dx + \int p(x)(\phi(x)h(x) - \phi(x)\,\widehat{h(x)})^T R^{-1}(d\zeta - \hat{h}(x)dt)dx \qquad (2.31)$$

or, equivalently,

$$d(\widehat{\phi(x)}) = \widehat{\mathcal{L}[\phi]}dt + (\widehat{\phi h} - \hat{\phi}\hat{h})^T R^{-1}(d\zeta - \hat{h}\,dt) \ . \qquad (2.32)$$

Letting $\phi(x) = x_i$ we can obtain from Eq. (2.32) the stochastic differential equation for the $i^{\underline{th}}$ component of the state vector's conditional mean; i.e.,

$$dm_i = \widehat{f_i(x,t)}\,dt + (\widehat{x_i h(x)} - m_i \widehat{h(x)})^T R^{-1}(d\zeta - \hat{h}\,dt) \quad (2.33)$$

where

$$m_i = \hat{x}_i = \int x_i p(x/Z_t)dx \ .$$

While Eq. (2.32) could be utilized to find the stochastic differential equations satisfied by the higher-order moment parameters of $p(x/Z_t)$, the central moment parameters of this density (e.g., the variance = $E\{(x-m)(x-m)^T/Z_t\}$) are of primary concern to the nonlinear filter problem, and Eq. (2.32) can not be used to obtain the differential relations for the central moments, which are by definition functions of both the state vector and its conditional mean.

The stochastic differential equations for the conditional central moment parameters may best be determined by taking the Ito differential of the defining moment relations. For example, if Eq. (2.6) and (2.25) hold true then the differential equation for the second central moment

$$m_{ij} \overset{\Delta}{=} E\{(x_i - m_i)(x_j - m_j)/Z_t\}$$

may be obtained by computing the Ito differential

$$dm_{ij} = d(\int (x_i - m_i)(x_j - m_j)p(x/Z_t)dx)$$

where

$$m_i \overset{\Delta}{=} E\{x_i/Z_t\}$$

satisfies its own differential equation; i.e., Eq. (2.33).

In Appendix I (see also Kushner [31]) it is shown that

$$dm_{ij} = \left[-(\widehat{hx_i} - \widehat{hm_i})^T R^{-1} (\widehat{hx_j} - \widehat{m_j h}) + \widehat{\mathcal{L}[(x_i - m_i)(x_j - m_j)]}\right] dt +$$

$$(d\zeta - \hat{h}\,dt)^T R^{-1}\, \overline{(h - \hat{h})(x_i - m_i)(x_j - m_j)}\,. \qquad (2.34)$$

In a similar manner it can be demonstrated that if

$$m_{iij} \overset{\Delta}{=} \int (x_i - m_i)^2 (x_j - m_j)\, p(x/Z_t)\, dx$$

then

$$dm_{iij} = [-2\hat{f}_i m_{ij} - \hat{f}_j m_{ii} + \widehat{\mathcal{L}[(x_i - m_i)^2 (x_j - m_j)]} -$$

$$2(\widehat{hx_i} - \widehat{hm_i})^T R^{-1}\, \overline{(h - \hat{h})\,(x_i - m_i)(x_j - m_j)} - \qquad (2.35)$$

$$(\widehat{hx_j} - \widehat{hm_j})^T R^{-1}\, \overline{(h - \hat{h})\,(x_i - m_i)^2}]dt + (d\zeta - \hat{h}dt)^T R^{-1}\,\cdot$$

$$\cdot\left[-2m_{ij}(\widehat{hx_i} - \widehat{hm_i}) - m_{ii}(\widehat{hx_j} - \widehat{hm_j}) + \overline{(h - \hat{h})(x_i - m_i)^2 \cdot (x_j - m_j)}\right]$$

where the $\mathcal{L}$ operator in Eq's. (2.34) and (2.35) operate w.r.t. the $x(t)$ process and not the $m(t)$ process.

In theory, at least, Eqs. (2.33)-(2.35) and their higher-order central moment counterparts constitute an optimal nonlinear filter. The conditional mean vector, m, provides a minimum variance, or least-squares, present best estimate of the state vector while the variance matrix, $P = [m_{ij}]$, provides a measure of confidence for this estimate. Note, however, from Eq's. (2.33) - (2.35) that the lower-order moment equations contain terms which are functions of higher-order moment parameters (assume for simplicity that the observation functions, $h(x,t)$, are linear in x). The chain is endless and the optimal filter equations are inherently of infinite dimension, with one exception being that of the linear Gaussian case.

If both the process and observation equations are linear and if the initial state vector probability density is Gaussian, then it may be shown that the conditional density function $p(x(t)/Z_t)$ is always Gaussian, and hence is parameterized by a finite number of parameters--the mean m and variance parameters P. Since the odd-order central moments of a Gaussian density function are zero, Eq's. (2.33) and (2.34) contain in the linear, Gaussian case terms which are functions of only the mean and variance parameters. The mean and variance equations (2.33) and (2.34) constitute in effect a finite-dimensional, optimal filter; the Kalman filter [24-26]. Unfortunately the general case is not so simple since little can be said about the general form the conditional density takes and as a result approximations must be made to the optimal filter equations.

2.4 The Continuous Process, Discrete Observation Problem

We digress slightly from the main theme of this thesis and consider the solution to the optimal nonlinear filtering problem for the case when observations are made only at discrete instants of time. While the continuous observation problem is the primary interest of this thesis, the discrete observation problem is a highly relevant one and its solution can be neatly formalized by a simple extension of the theory presented in the preceding sections of this chapter.

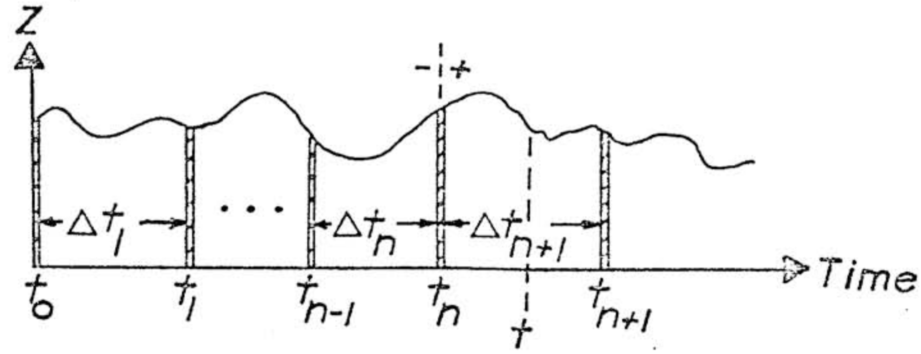


FIGURE 2.1 Schematic of Continuous Process With Discrete Measurements

Consider the problem of determining

$$p(x(t)/Z_n) = p(x(t)/z_0, z_1, \ldots, z_n); \quad \sum_0^n \Delta t_n \leq t$$

where

$$dx(t) = f(x,t)dt + g(x,t)db(t) \tag{2.6}$$

and

$$z_n = h(x(t_n),n) + R_n^{1/2} v_n , \tag{2.37}$$

$$p(v_0, v_1, \ldots, v_n) = \prod_{i=0}^{n} N(0,1) .$$

By intuitive reasoning, we can expect that between the n-1st and nth observation sampling instants the conditional probability density $p(x(t)/Z_{n-1}) \equiv P(x,t)$ satisfies the Fokker-Planck equation,

$$\frac{\partial P(x,t)}{\partial t} = \overset{*}{\mathcal{L}}[P(x,t)] \tag{2.21}$$

from which, in the usual manner (Sec. 2.2.4), differential equations can be derived for the moment parameters of $p(x/Z_n)$ between samples. At the $n^{\underline{th}}$ sampling instant these moments and $p(x(t)/Z)$ must be updated using the following procedure:

From Bayes Rule,

$$p(x(t_n^+)/z_n) = \frac{p(z_n/x(t_n^+))\ p(x(t_n^+))}{p(z_n)}$$

$$= \frac{p(z_n/x(t_n^+))\ p(x(t_n^+))}{\int(\text{numerator})dx(t_n^+)} \ . \qquad (2.38)$$

From Eq. (2.37)

$$p(z_n = Z/x = X) = p(Z = h(X,n) + R_n^{1/2}v_n) = p(v_n = R_n^{-1/2}(Z - h(X,n))$$

$$= N(h(X,n),R_n) \ . \qquad (2.39)$$

Then the exact solution for $p(x(t_n^+)/z_n)$ is given by Eq. (2.38) with $p(z_n/x(t_n^+))$ satisfying relation (2.39) and the a'priori density $p(x(t_n^+))$ being the solution of Eq. (2.21) at $t = t_n^-$ (i.e., $P(x(t_n^-)))$. The calculation of the updated moments at $t = t_n$ requires that functions $p(x(t_n^+)/z_n)$ be integrated over the state space. Such integrations are impractical to perform for on-line solutions and some sort of approximations must be made in order to derive a useful, totally recursive filter algorithm.

If we make the following assumptions:

<u>A1</u>: That $P(x(t_n^-))$, satisfying Eq. (2.21) at $t = t_n^-$, and $p(x(t_n^+)/z_n)$ are Gaussian; i.e.,

$$P(x(t_n^-)) = N(m_-,\Lambda_-)\ ,\ p(x(t_n^+)/z_n) = N(m_+,\Lambda_+)\ .$$

<u>A2</u>: That $h(x) = h(m) + \nabla h(m)(x-m)$ where

$$\nabla h = \begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \frac{\partial h_1}{\partial x_2} & \cdots & \frac{\partial h_1}{\partial x_m} \\ \vdots & \vdots & & \vdots \\ \frac{\partial h_k}{\partial x_1} & \frac{\partial h_k}{\partial x_2} & & \frac{\partial h_k}{\partial x_m} \end{bmatrix} ,$$

then Eq. (2.39) yields

$$p(z_n/X) = C_1 \exp\{(X - m_-)^T \nabla h^T(m_-,n) R^{-1}(z_n - h(m_-,n)) - \frac{1}{2}(X-m_-)^T \nabla h^T(m_-,n) R_n^{-1} \nabla h(m_-,n)(X-m_-)\}$$

where C_1 is a constant independent of the value of X. Eq. (2.38) becomes

$$p(x(t_n^+) = X/z_n) = C_1 \exp\{(X-m_-)^T \nabla h^T R_n^{-1}(z_n - h(m_-,n)) - \frac{1}{2}(X-m_-)^T \nabla h^T R_n^{-1} \nabla h(X-m_-) - \frac{1}{2}(X-m_-)^T \Lambda_-^{-1}(X-m_-)\}\ . \tag{2.40}$$

By completing squares in Eq. (2.40), we can obtain the updated statistics at t_n, m_+ and Λ_+, in terms of m_-, Λ_-, and z_n; we find that

$$\Lambda_+^{-1} = \Lambda_-^{-1} + \nabla h^T(m_-,n) R_n^{-1} \nabla h(m_-,n) \tag{2.41}$$

and

$$m_+ = m_- + \Lambda_+ \nabla h^T(m_-,n) R_n^{-1}(z_n - h(m_-,n))\ . \tag{2.42}$$

The following useful relation for Λ_+ can be obtained from Eq. (2.41) by using the Matrix Inversion Lemma (see Ho [21]):

$$\Lambda_+ = \Lambda_- - \Lambda_- \nabla h^T (\nabla h \Lambda_- \nabla h^T + R)^{-1} \nabla h \Lambda_- \ . \qquad (2.43)$$

Note that Eq. (2.43) does not require the inversion of Λ_- that Eq. (2.41) does.

The algebraic Eq's. (2.42) and (2.43) together with the mean and variance differential equations, derived from the Fokker-Planck equation (2.6), constitute a discrete-measurement, continuous-process approximately minimum-variance nonlinear filter. A similar procedure can be carried out utilizing a second-order Taylor series approximation to $h(x)$ in <u>A2</u> .

2.5 Summary

It was shown how the independent increment property of ideal white-noise leads to the troublesome relation $E\{(db)^2\} = dt$ for the associated brownian motion process $b(t)$. We saw how this relation led to the development by Ito of the stochastic calculus, the rules of which differ from those of the ordinary calculus and the foundation upon which the theory of optimal nonlinear filtering is based. A simple application of Ito's stochastic differential rule resulted in a rederivation of the Fokker-Planck equation, an ordinary partial differential equation satisfied by the probability density of the state vector conditioned upon any a'priori information regarding the state.

Consideration was given to Bucy's rederivation of a stochastic partial differential-integral equation similar to the Fokker-Planck equation and satisfied by the state vector probability density conditioned upon a'priori information and the a'posteriori information provided by the observations, $z(t)$. From the conditional version of the Fokker-Planck equation, stochastic differential equations for the moment parameters of the conditional density function, $p(x/Z)$, were derived and these, in effect, constitute a general infinite-dimensional, optimal filter. The finite dimensional optimal filter which results in the linear Gaussian case was shown to be the exception to this generally infinite-dimensional problem.

The theoretical solution to the continuous process, discrete observation nonlinear filtering problem was formalized by way of applications of the Fokker-Planck equation and Bayes rule. An approximate method of minimum variance filtering was proposed for the discrete measurement case.

CHAPTER III

APPROXIMATE METHODS OF OPTIMAL NONLINEAR FILTERING

3.1 Introduction

It was shown in the previous chapter that with the exception of the linear, Gaussian case, the general solution to the optimal nonlinear filtering problem consists of an infinite set of stochastic differential equations with observation function driving terms and being satisfied by certain parameters of the conditional probability density $p(x(t)/Z_t)$. In particular, a portion of the minimum variance filter equations were presented and consisted of the differential equations for the conditional mean and the conditional central moment parameters. Because of their fundamental importance, the mean and covariance differential equations for the minimum variance filter are rewritten here in their more commonly occurring form, with

$$z(t) = \frac{d\zeta}{dt} = h(x,t) + v(t)^{1} . \tag{1.1}$$

They are

$$\frac{dm_i}{dt} = \widehat{f_i(x,t)} + (\widehat{x_i h(x)} - m_i \widehat{h(x)})^T R^{-1}(z(t) - \widehat{h(x)}) \tag{3.1}$$

and

$$\frac{dm_{ij}}{dt} = \widehat{\mathcal{L}[(x_i - m_i)(x_j - m_j)]} - (\widehat{hx_i} - \hat{h} m_i)^T R^{-1} (\widehat{hx_j} - m_j \hat{h})$$
$$+ (z(t) - \hat{h})^T R^{-1} \widehat{(h - \hat{h})(x_i - m_i)(x_j - m_j)} \tag{3.2}$$

---

1 In the remainder of this thesis stochastic differential equations will be written for convenience as ordinary differential equations. The treatment of stochastic differential equations as ordinary differential equations is absolutely correct only if the Ito correction terms are accounted for (see Section 2.2.5).

where if $\phi(x)$ is any function of x, then

$$\widehat{\phi(x)} = \int \phi(x) p(x/Z_t) dx \ .$$

The presently available methods for replacing the infinite dimensional optimal filter solution, as represented by Eq's. (3.1) and (3.2) and their higher-order conditional moment counterparts, by an approximating infinite dimensional solution are the primary concern of this chapter. Some of the more common and potentially useful of these methods are presented and briefly discussed. Upon reading this chapter, one should not become discouraged by the large number and variety of these filtering techniques; most being very similar and easily mastered if a thorough understanding of the conditional moment equations is achieved. The fundamental importance of the conditional moment equations (e.g., Eq's. (3.1) and (3.2)) to the optimal filter approximation problem can not be over emphasized.

3.2 Nominal Trajectory Kalman Filtering

The most common and easiest to implement method of nonlinear filtering is that of Kalman filtering about a nominal trajectory. The procedure is that of linearizing all nonlinear functions about some a'priori determined nominal state trajectory $(x^n(t))$ and utilizing a Kalman filter to estimate the perturbations of the actual trajectory from the nominal trajectory. It may be easily demonstrated that the linearized Kalman filter equations are given as

$$\frac{dm}{dt} = \nabla f(x^n,t)\tilde{m} + P\nabla h^T(x^n,t)R^{-1}(z - z^n - \nabla h(x^n,t)\tilde{m}) \qquad (3.3)$$

and

$$\frac{dP}{dt} = \nabla f(x^n,t)P + P\,\nabla f^T(x^n,t) + g(x^n,t)g^T(x^n,t) -$$

$$P\nabla h^T(x^n,t)R^{-1}\nabla h(x^n,t)P \tag{3.4}$$

where

$$\tilde{m} = E\{(x - x^n)/Z_t\}$$

$$P = E\{(x - x^n - \tilde{m})(x - x^n - \tilde{m})^T/Z_t\} = [m_{ij}]$$

$$\frac{dx^n}{dt} = f(x^n,t) \ ; \ z^n = h(x^n,t) \tag{3.5}$$

$$x^n(0) = E\{x(0)\} \ ; \ \tilde{m}(0) = 0$$

$$P(0) = E\{(x(0) - x^n(0))(x(0) - x^n(0))^T\}$$

$$\nabla f = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix} \ ; \ \nabla h = \begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial h_m}{\partial x_1} & \cdots & \frac{\partial h_m}{\partial x_n} \end{bmatrix}.$$

One primary advantage of the linearized Kalman filter is that since the covariance matrix (P) is independent of the observations and the conditional mean it can be precomputed off-line and stored in memory to be recalled during the actual filtering operation. A less obvious but equally important feature of this filter concerns its stability requirements for in general, the "optimality" of a filter does not in itself guarantee its "stability". Kalman [26] has shown that if a system of the form

$$\frac{d\tilde{x}}{dt} = \nabla f(t)\tilde{x} + g(t)w(t)$$

$$z(t) = \nabla h(t)\tilde{x} \qquad (3.6)$$

is both observable[1] and controllable[2] then the Kalman filter for that system is stable. At this time, the specific requirements for the stability of nonlinear filters are not known and the influence of nonlinear observability and controllability on filter stability has not been clearly demonstrated.

The primary disadvantage of the linearized Kalman filter is its lack of efficiency for significant disturbance levels and initial state vector variances. This disadvantage can be lessened by replacing in Eq. (3.3) the linearized error function $z - z^n - \nabla h(x^n)\tilde{m}$ by the intuitively more logical error function $z - h(m)$. A second obvious disadvantage is that the partial derivatives in ∇h and ∇f must all exist.

3.3 Relinearized Kalman Filtering

If the nominal trajectory of the linearized Kalman filter is allowed to vary with time and chosen to be that trajectory which passes through the present best state-estimate, then the relinearized (also modified or extended) Kalman nonlinear filter results. From Eq's. (3.3) - (3.6) we have

---

1 The system (3.6) is called observable if it is possible to determine the value of $x(t_0)$ from the values of $z(t)$ over a finite interval of time $[t_0, t_f]$.

2 The system (3.6) is called controllable if at any time t the state $x(t)$ can be transformed to any other desired state in a finite interval of time by a suitable choice of the disturbance function $w(t)$ over that time interval.

$$\frac{d\tilde{m}}{dt} = \frac{d\hat{x}}{dt} - \frac{dx^n}{dt} = \nabla f(x^n)\tilde{m} + P\nabla h^T(x^n)R^{-1}(z - z^n - \nabla h(x^n)\tilde{m})$$

or, letting $m \equiv \hat{x}$,

$$\frac{dm}{dt} = f(x^n) + \nabla f(x^n)(m - x^n) + P\nabla h^T(x^n)R^{-1}(z - h(x^n) - \nabla h(x^n) \cdot (m - x^n)) \ . \tag{3.7}$$

Eq's. (3.7) and (3.4) yield the following nonlinear filter equations if we choose as the nominal state that of the present best estimate (i.e., $x^n \equiv m$):

$$\frac{dm}{dt} = f(m) + P\nabla h^T(m)R^{-1}(z - h(m)) \tag{3.8}$$

$$\frac{dp}{dt} = \nabla f(m)P + P\nabla f^T(m) + g(m)g^T(m) - P\nabla h^T(m)R^{-1}\nabla h(m)P \tag{3.4}$$

where

$$m = E\{x/Z_t\}$$

$$P = E\{x - m)(x - m)^T/Z_t\} = [m_{ij}]$$

$$m(0) = E\{x(0)\}$$

$$P(0) = E\{(x(0) - m(0))\ (x(0) - m(0))^T\} \ .$$

Computational experiments conducted during this study (see Chapter IV) have proven the relinearized Kalman filter to be a remarkably effective filtering algorithm. Its primary disadvantages are that, first, the variance equation must be computed on-line during the filtering operation and, second, that the partial derivatives in $\nabla h(x)$ and $\nabla f(x)$ must exist at all points in the finitely probable region of the state space.

3.4 Second-Order Optimal Filtering

Bass, Norum, and Schwartz [2] have suggested that the functions $f_i(x)$, $h(x)$, and $g(x)g(x)^T$ in the optimal filter equations (3.1) and (3.2) be approximated by the first three terms in the Taylor series expansions about the state's present best estimate; e.g.,

$$f_i(x) \approx f_i(m) + \nabla f_i(m)(x - m) + \frac{1}{2}(x - m)^T \nabla^2 f_i(m)(x - m) .$$

Under the further assumptions that m_{ijk} and m_{ijkl} are negligible, Bass, et. al. showed that the following nonlinear filter equations can be derived from the conditional moment equation for $p(x/Z_t)$:

$$\frac{dm}{dt} = f(m) + \frac{1}{2}\nabla^2\{f(m)P\} + P\nabla h^T(m)R^{-1}(z - h(m) - \frac{1}{2}\nabla^2\{h(m)P\}) \qquad (3.10)$$

$$\frac{dP}{dt} = P\nabla f^T(m) + \nabla f(m)P - P(\nabla h^T(m)R^{-1}\nabla h(m))P + g(m)g^T(m) + \frac{1}{2}\nabla\{g(m)g^T(m)P\} - \frac{1}{2}\nabla^2\{h(m)P\}^T R^{-1}(z - h(m) - \frac{1}{2}\nabla^2\{h(m)P\})P \qquad (3.11)$$

where

$$\nabla^2\{fP\} \triangleq \begin{bmatrix} tr[\nabla^2 f_1(x)P] \\ tr[\nabla^2 f_2(x)P] \\ \vdots \\ tr[\nabla^2 f_n(x)P] \end{bmatrix}$$

$$tr[\nabla^2 f_i(x)P] = \sum_{j,k} m_{jk} \frac{\partial^2 f_i}{\partial x_j \partial x_k}$$

$$m = E\{x/Z_t\} \; ; \; P = E\{(x - m)(x - m)^T/Z_t\}$$

$$m(0) = E\{x(0)\} \; , P(0)=E\{(x(0)-m(0))(x(0)-m(0))^T\}$$

This nonlinear filtering technique possesses the obvious disadvantages of being more complex than that of the relinearized filtering technique and requiring the existence of not only the first partial derivatives of $f(x)$ and $h(x)$, but also the second partial derivatives in $\nabla^2\{fP\}$, $\nabla^2\{gg^TP\}$, and $\nabla^2\{hP\}$. It remains to be proven whether or not the performance of the second-order filter is significantly, if at all, greater than that of the less complex and restrictive relinearized Kalman filter.

A general comment at this point seems appropriate. While one can if he wishes accept at face value a nonlinear filtering algorithm (such as that represented by Eq's. (3.10) and (3.11)) and all its inherent assumptions, it is our preference to work directly with the conditional moment equations (e.g., Eq's. (3.1) and (3.2)) and introduce our own simplifying assumptions to these equations. Taylor series expansions, higher-order moment assumptions, and other fundamental assumptions can be utilized in coming up with an effective nonlinear filtering algorithm for a particular application problem.

3.5 Wide-Sense Kalman Filtering

Doob's [15] concept of wide-sense properties of stochastic processes was applied by Bucy to the nonlinear filtering problem which results from a process having linear dynamics and a non-Gaussian initial state probability density. One method of solution for this particular filtering problem consists of the Kalman filter equations whose initial values are determined from the mean and variance parameters of the a'priori probability density. Although the resulting filter is referred to as a wide-sense Kalman filter, it should not be misconstrued as being a general wide-sense optimal nonlinear filter; it isn't.

Lo [34] has considered the use of functional transformations to convert a general nonlinear filtering problem into an equivalent wide-sense Kalman filtering problem (i.e., one with linear dynamics and measurements and a non-Gaussian a'priori density function). The method is best explained by way of an example.

Consider the nonlinear filtering problem specified by the process and observation relations

$$\frac{dx_1}{dt} = -x_1 \quad (3.12)$$

and

$$z(t) = x_1 + x_1^3 + R^{1/2}v(t) \quad (3.13)$$

where

$$p(x_1(0)) = N(m_{10},\sigma_{10}^2) \;. \quad (3.14)$$

This problem may be transformed into a linear problem with a non-Gaussian initial probability density by letting

$$y_1 \equiv x_1 \qquad \text{and} \qquad y_2 \equiv x_1^3 \;.$$

Then from Eq's. (3.12) and (3.13),

$$\frac{dy_1}{dt} = -y_1 \qquad (3.15)$$

$$\frac{dy_2}{dt} = -3y_2$$

$$z(t) = y_1 + y_2 + R^{1/2}v(t) \qquad (3.16)$$

with

$$y_1(0) = x_1(0) \text{ and } y_2(0) = x_1^3(0) \; .$$

If we assume a Gaussian joint probability density for $y_1(0)$ and $y_2(0)$, the wide-sense Kalman filter for the transformed problem specified by Eq's. (3.15) and (3.16) would consist of the following relations:

$$\frac{dm_1}{dt} = - m_1 + (m_{11} + m_{12})R^{-1}(z - m_1 - m_2)$$

$$\frac{dm_2}{dt} = -3m_2 + (m_{12} + m_{22})R^{-1}(z - m_1 - m_2)$$

$$\frac{dm_{11}}{dt} = - (m_{11} + m_{12})^2 R^{-1} - 2m_{11} \qquad (3.17)$$

$$\frac{dm_{22}}{dt} = - (m_{12} + m_{22})^2 R^{-1} - 6m_{22}$$

$$\frac{dm_{12}}{dt} = - (m_{11} + m_{12})(m_{22} + m_{12})R^{-1} - 4m_{12}$$

with

$$m_1(0) = m_{10}$$

$$m_2(0) = m_{10}^3 + 3m_{10}\,\sigma_{10}^2$$

$$m_{11}(0) = \sigma_{10}^2$$

$$m_{22}(0) = 15\,\sigma_{10}^{6} + 36m_{10}^{2}\,\sigma_{10}^{4} + 9m_{10}^{4}\sigma_{10}^{2}$$

$$m_{12}(0) = 3\sigma_{10}^{4} + 3m_{10}^{2}\sigma_{10}^{2} \;.$$

Of course the task of determining the correct functional transformations to convert a general nonlinear filter problem into one with wide-sense properties is not an easy one. Lo [34] has developed a technique for determining those transformations for a particular class of nonlinear filtering problems and this technique is based on the concept of finite-dimensional sensor orbits[1].

The general merits of the wide-sense Kalman filter have yet to be determined. However, computational experiments were conducted in this study for the example problem specified by Eq's. (3.15) and (3.16) and the results of this study are presented in Chapter IV.

3.6 Assumed-Form Density Filtering

Perhaps the least restrictive method (i.e., that requiring the least number of preliminary assumptions) of approximately optimal nonlinear filtering is that which assumes a particular mathematical form for the conditional density function $p(x/Z_t)$ and then determines the filter's structures from the conditional moment equations. Kushner [31] first suggested this technique and applied it to a van der pol oscillator filtering problem, making the assumption of a Gaussian conditional density function.

One difficulty of this method is that the number of useful probability density function types is rather limited. The Gaussian

---

1 The sequence $\{\mathcal{L}^{i}[h(x)]\}_{i=0,1,\ldots}$ is called the sensor orbit.

density function first comes to mind and is extremely useful for problems having polynomial state and observation functions (i.e., f(x), g(x), and h(x)). Under the Gaussian assumption, however, it is difficult to evaluate analytically such functional expectations as

$$\widehat{f(x)} \overset{\Delta}{=} \int f(x)p(x/Z)dx \tag{3.18}$$

when the function $f(x)$ is not of a polynomial form. In these cases the uniform density function can be very useful since the integrals such as that in Eq. (3.18) are readily evaluated analytically, even the troublesome piece-wise linear function case.

A primary advantage to the assumed-form density filtering method is that it does not require differentiability of the system functions for its application. While the forementioned filtering techniques require the existence of the partial derivatives in $\nabla h(x)$ and $\nabla f(x)$, the assumed-form technique does not make this requirement, being based upon the integration operation rather than differentiation. As an example, consider the problem specified by the system equations

$$\frac{dx_1}{dt} = - x_1 + Q^{1/2}w(t) , \tag{3.19}$$

and

$$z(t) = h(x_1) + R^{1/2}v(t) , \tag{3.20}$$

where

$$h(x_1) = \begin{cases} +1, & x_1 \geq d \\ 0, & -d < x_1 < d \\ -1, & x_1 \leq -d \end{cases}$$

$$p(x_1(0)) = N(m_{10}, \sigma_{10}^2) .$$

From Eq's. (3.1) and (3.2) we find the conditional moments m_1 and m_{11} satisfying the two equations

$$\frac{dm_1}{dt} = - m_1 + (\widehat{x_1 h} - m_1 \hat{h})R^{-1}(z - \hat{h})$$

and (3.21)

$$\frac{dm_{11}}{dt} = - (\widehat{x_1 h} - m_1 \hat{h})^2 R^{-1} - 2m_{11} + Q + (z - \hat{h})R^{-1}(\widehat{x_1^2 h} - 2m_1 \widehat{xh} - \hat{h} m_{11} + m_1^2 \hat{h}) .$$

where the functional expectations $\hat{h}$, $\widehat{x_1 h}$, and $\widehat{x_1^2 h}$ will be determined by letting

$$p(x_1/Z_t) \simeq \tilde{p}(x_1) = \begin{cases} 1/2\sigma , & m_1 - \sigma \leq x_1 \leq m_1 + \sigma , \\ 0 , & \text{otherwise} \end{cases}$$

with $\sigma = \sqrt{3m_{11}}$ to make the variances of $p(x_1/Z_t)$ and $\tilde{p}(x_1)$ identical. Then the functional expectations

$$\hat{h} \approx \int h(x)\tilde{p}(x)dx ,$$

$$\widehat{xh} \simeq \int xh(x)\tilde{p}(x)dx ,$$

and

$$\widehat{x^2h} \approx \int x^2 h(x)\ddot{p}(x)dx$$

may be easily shown by direct integration to be specified by the relations in Tables 3.1 and 3.2[1].

3.7 Miscellaneous Methods

Sunahara [44] has developed an approximate method of nonlinear state estimation which is based on the method of stochastic linearization. The procedure is to replace the system nonlinear functions by linear functions whose coefficients are chosen in such a way as to minimize the expectation of the squared value of the error between the nonlinear and linear functions, conditioned upon the observations Z_t. This particular filtering technique, while intuitively pleasing in principle, appears to be quite unwieldy in practice and does not consider disturbance functions $g(x,t)$ which are functions of x.

Since Stratonovich's original contributions to the basic theory of optimal nonlinear filter theory, the Russian efforts in this field appear to have been limited to those of Dashevskii [9,10,11,12] who has considered a method of synthesizing nonlinear filters which is based on the equations for the conditional semi-invariants[2] of $p(x/Z_t)$. Utilizing the conditional version of the

1 The mode of the filter refers to one of a number of possible regions of the filter's state space and is specified by the relations in Table 3.1.

2 The k^{th} semi-invariant of a random variable x having the characteristic function $\theta(y)$ is

$$\mu_k = (-i)^k \partial^k \gamma(y)/\partial y^k \Big|_{y=0} ,$$

where $\gamma(y) = \ln \theta(y) = \ln \int_{-\infty}^{\infty} \exp(iyx)p(x)dx$ [19].

Table 3.1 - Uniform-Density Filter Mode Numbers; $\sigma = \sqrt{3m_{11}}$

$m_1 - \sigma$ \ $m_1 + \sigma$	$\cdot \leq -d$	$-d < \cdot < d$	$\cdot \geq d$
$\cdot \leq -d$	1	2	3
$-d < \cdot < d$		4	5
$\cdot \geq d$			6

Table 3.2 - Functional Expectations $\widehat{h}$, $\widehat{xh}$, and $\widehat{x^2h}$ for Each Mode of the Uniform-Density Filter

Mode No.	$\widehat{h(x)}$	$\widehat{xh(x)}$	$\widehat{x^2h(x)}$
1	-1	$-m_1$	$-m_1^2 - \sigma^2/3$
2	$\frac{(m_1 - \sigma + d)}{2\sigma}$	$\frac{((m_1 - \sigma)^2 - d^2)}{4\sigma}$	$\frac{((m_1 - \sigma)^3 + d^3)}{6\sigma}$
3	$\frac{m_1}{\sigma}$	$\frac{(m_1^2 + \sigma^2 - d^2)}{2\sigma}$	$\frac{(m_1^3 + 3\sigma^2 m_1)}{3\sigma}$
4	0	0	0
5	$\frac{(m_1 + \sigma - d)}{2\sigma}$	$\frac{((m_1 + \sigma)^2 - d^2)}{4\sigma}$	$\frac{((m_1 + \sigma)^3 - d^3)}{6\sigma}$
6	1	m_1	$m_1^2 + \frac{\sigma^2}{3}$

Fokker-Planck equation (Eq. (2.28)), Dashevskii [12] derives the differential equations for the semi-invariants of $p(x/Z_t)$, and in so doing approximates $p(x/Z_t)$ by a truncated version of the Edgeworth series [8];

$$p(x/Z) \simeq \frac{1}{\sqrt{2\pi\mu_2(t)}} \exp\{-\frac{(x-\mu_1(t))^2}{2\mu_2(t)}\}[1 + B_3H_3(\frac{x-\mu_1}{\sqrt{\mu_2}}) +$$

$$B_4H_4\frac{(x-\mu_1)}{\sqrt{\mu_2}} + B_5H_5\frac{(x-\mu_1)}{\sqrt{\mu_2}} + B_6H_6\frac{(x-\mu_1)}{\sqrt{\mu_2}} + \ldots]$$

where the $H_n(y) = (-1)^n \exp(y^2/2) \frac{d^n}{dy^n} \exp(-y^2/2)$ are Hermite polynomials and the coefficients $B_k(k=3,4,\ldots)$ have the form ($\sigma = \sqrt{\mu_2(t)}$)

$$B_3 = \frac{1}{3!}\frac{\mu_3}{\sigma 3}, \; B_4 = \frac{1}{4!}\frac{\mu_4}{\sigma 4}, \; B_5 = \frac{1}{5!}\frac{\mu_5}{\sigma^5}, \; B_6 = \frac{10}{6!}\frac{\mu_3^2}{\sigma^6} .$$

In as much as the first three semi-invariants of a probability density correspond to the first three moments of that density function, Dashevskii's introduction of the semi-invariants to the nonlinear filtering problem appears to have only theoretical value. It is doubtful that any practical filter applications would require the inclusion of any moments of higher than third order to the filtering algorithms.

For various reasons, certain individuals prefer to work with difference equation descriptions of dynamical systems. In this light, we recommend to such persons the work of Sorenson and Stubberud [39], who considered the problem of estimating the state of a system described by nonlinear difference equations from noisy nonlinear measurement data. A Bayes approach is taken. It is our personal preference to work with a differential equation description of the system and discretize, instead, the resulting continuous filter equations before implementing the filter on a digital computer. This latter method allows us to apply the conceptually pleasing results of the very complete theory of continuous optimal nonlinear filtering--the discrete optimal filtering theory is not nearly so complete or easily applied.

3.8 Summary

For application purposes the inherently infinite-dimensional solution to the optimal nonlinear filtering problem forces us to consider ways of approximating these solutions by those which are of finite dimension. In this chapter, we considered a number of the more common approximate filtering methods which have been proposed. These methods make use of such approximation techniques as truncated Taylor series expansions, nonlinear functional transformations, and assumptions regarding the conditional probability density $p(x/Z_t)$.

In surveying the literature dealing with these approximate filtering methods, one will notice that there exists a total lack of information regarding the absolute effectiveness of these methods. While certain investigators [1,7,38] have made somewhat half-hearted attempts at comparing the performance of their filtering technique with that of some other approximate technique, no one has made an effort to compare their filter performance with that of the optimal filter. The reason for this situation is clear when one considers the infinite-dimensional nature of the optimal filter. As a result no one can say with certainty which approximate filtering method is most effective or possesses the greatest potential for achieving near optimal performance with the smallest degree of complexity.

To help remedy this situation, we present and demonstrate in the following chapter a numerical method for computing upper performance-bounds of low-dimension optimal nonlinear filtering problems. It is hoped that the results of experiments conducted using this method will shed new light on the approximate filtering problem.

CHAPTER IV

COMPUTATIONAL EXPERIMENTS IN NONLINEAR FILTERING

4.1 Introduction

Although the development of the basic theory of optimal nonlinear filtering was completed almost five years ago, since that time almost no practical applications of the theory have appeared (an exception is the ballistic missile reentry problem discussed by Athans [1]). One reason for this situation may possibly be that so few technical people are aware of the existence of this theory, while many of those who have been introduced to the theory lose interest quite quickly when confronted with such alien terminology as infinite-dimensionality, stochastic calculus, semi-invariants, quasi-moments, and other examples of academic jargon which so liberally dot the landscape of the literature dealing with this subject. Much of this literature which has appeared since the basic theory was introduced has done much to confuse the picture and little to clarify it.

Some pressing questions related to the practical aspects of optimal nonlinear filtering theory remain to be answered;

- How much less than optimal is the performance of approximately optimal nonlinear filters?
- How significant are the effects of neglecting the higher-order moments of the optimal nonlinear filter?
- Do the more complex approximate filtering methods necessarily yield the more optimal performances?

- What filtering techniques show the most promise for application to the broadest class of practical problems?

To assist in answering the above questions, we have conducted computational experiments with five nonlinear filter problems. For each problem an upper performance-bound (i.e., the optimal performance) was computed and the approximate optimal filter which most nearly achieved optimal performance was determined. The results of these experiments are presented in this chapter along with a description of the method used to compute upper performance bounds. This method was first suggested by Bucy [5] but we believe this study to be the first actual implementation of the technique.

4.2 Computing Upper Performance-Bounds

The task of computing the optimal performance for a particular nonlinear filtering problem is by no means a simple one. Since the optimal estimate is a direct function of the conditional probability density $p(x/Z)$, a practical method for numerically computing this density function is required before the optimal performance can be estimated. Such a method has been developed for this study and proven to be useful and practical for low-dimension nonlinear filtering problems.

While one could attempt to compute $p(x/Z)$ by obtaining the numerical solution to the partial-differential-integral equation (2.28), the numerical difficulties in obtaining such solutions being what they are, this approach to the problem was deemed impractical. Kroy and Stubberud's [32] approach to the problem was based on the eigenfunction expansion schemes used in the solution of noise-detection problems and it was demonstrated on a scalar identification problem. This latter approach was also considered impractical in that the basic theory involved was both cumbersome and unclear and the methodology of the approach was not readily applicable to a general enough class of problems.

Bucy has suggested that the conditional probability density $p(x/Z)$ be computed by numerically approximating the Bucy representation, as given by Eq. (2.26). This approach was found to be intuitively pleasing, based on a straightforward and easily understood theory, and highly amenable to computer implementation. As such, the Bucy representation was utilized in this study to determine the conditional probability density $p(x/Z)$, which in turn was used to determine optimal state estimates and, after multiple Monte Carlo simulations, optimal performance measures. Two optimality criteria were investigated; that of, first, minimum variance and, second, maximum likelihood.

Recalling Bucy's representation theorem (sec. 2.3.1), we can write

$$p(x(t) = X/Z_t) = \frac{E\{\exp \phi/x(t) = X\}\ p(x(t) = X)}{E\{\exp \phi\}} \tag{4.1}$$

where

$$\phi = \int_{t_0}^{t} h^T(x(s),s)R^{-1}(s)[z(s) - \tfrac{1}{2}h(x(s),s)]ds \tag{4.2}$$

with

$$\frac{dx}{ds} = f(x,s) + g(x,s)w(s) \; ; \; x(t) = X \tag{4.3}$$

and the observations $Z_t = \{z(s): t_0 \leq s \leq t\}$ kept fixed in computing the functional ϕ . In Eq. (4.1) the probability density function $p(x = X)$ is the a priori probability density for the state vector at time t and the term $E\{\exp \phi\}$ is a normalizing factor given as

$$\int E\{\exp \phi/X\} \, p(x) \, dX \, .$$

In computing the functional ϕ by Eq. (4.2) it is necessary to solve the differential equation (4.3) backward in time since the final value $x(t) = X$. The expectations above must be taken over all possible disturbance functions $w(t)$, while

$$E\{\exp \phi/X\} \doteq \exp \phi/X$$

when no process disturbances, $w(t)$, are acting.

To develop a numerical procedure for computing $p(x/Z)$ it becomes necessary to discretize the representation functions in time, t, and state, x . For the sake of discussion, we will consider the scalar state and measurement problem; the extension to the multi-dimensional case being obvious. Then the discretized version of the Bucy representation may be written as (see Figure 4.1)

$$\tilde{p}(x(t) = X/Z\,) = \frac{\tilde{E}\{\exp \phi/X\} \; \tilde{p}(X)}{\tilde{E}\{\exp \phi\}} \tag{4.4}$$

where

$$p(x(t) = X/Z) \doteq \tilde{p}(X/Z) = p(X_k/Z) \; ; \; X_k \leq X < X_{k+1}$$

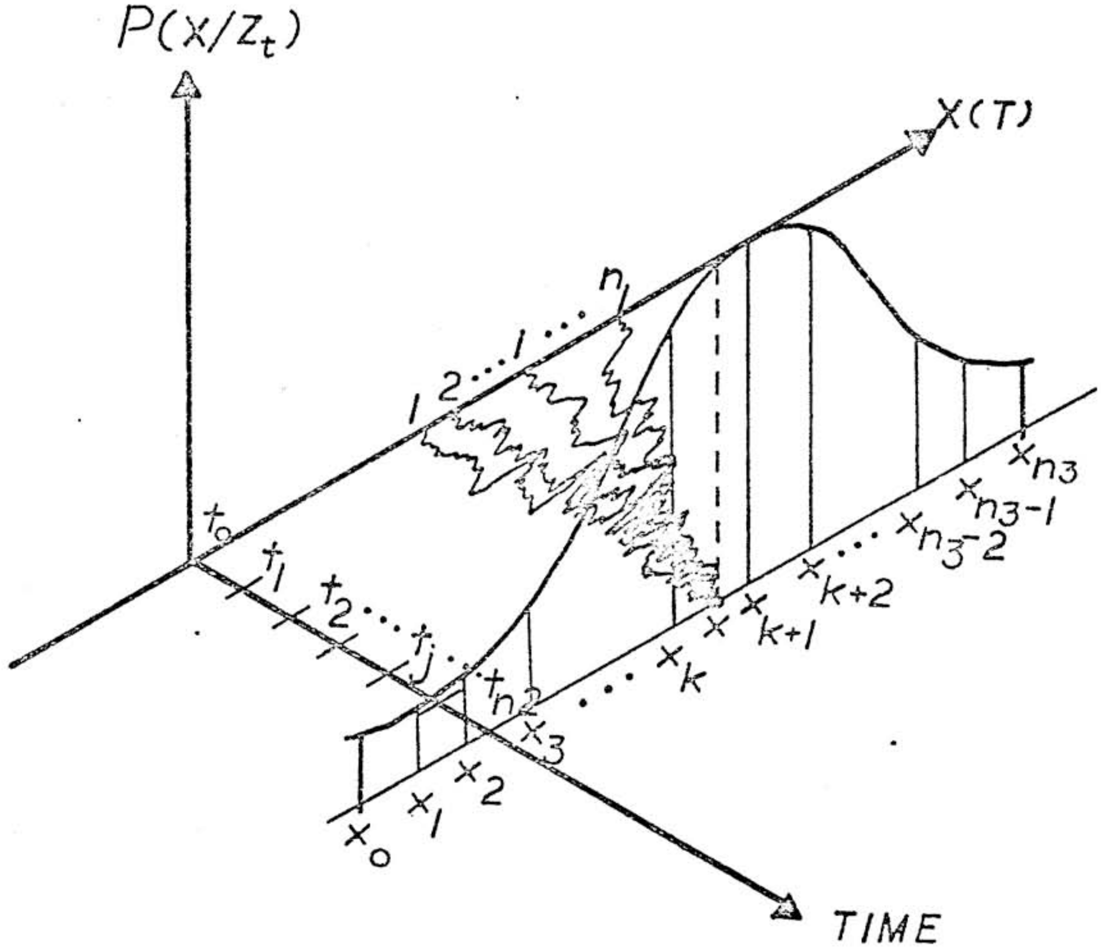


FIGURE 4.1 Notational Schematic for Discrete Representation Procedure

$$p(x(t) = X) \approx \tilde{p}(X) = p(X_k) \; ; \; X_k \leq X < X_{k+1}$$

$$\tilde{E}\{\exp \phi/X\} = \frac{1}{n_1} \sum_{i=1}^{n_1} \exp \phi_i/X \quad (4.5)$$

$$\phi_i = \sum_{j=0}^{n_2-1} h(x^i(t_j),t_j)R^{-1}(t_j)[z(t_j) - \tfrac{1}{2}h(x^i(t_j),t_j)][t_{j+1}-t_j] \quad (4.6)$$

$$x^i(t_j) = x^i(t_{j+1}) - [f(x^i(t_{j+1}),t_{j+1}) + g(x^i(t_{j+1}),t_{j+1}) \cdot$$

$$w(t_{j+1})][t_{j+1} - t_j] \quad (4.7)$$

$$x^i(t_{n2}) = X_k \; ; \; X_k \leq X < X_{k+1}$$

$$\tilde{E}\{\exp \phi\} = \sum_{k=0}^{n_3-1} \tilde{E}\{\exp \phi/X_k\} \, \tilde{p}(X_k)[X_{k+1} - X_k] \quad (4.8)$$

$$p(w(t_j)) = N(0,1/(t_{j+1} - t_j)) \, . \quad (4.9)$$

Eq's. (4.4)-(4.9) were used to numerically compute the conditional probability density $p(x(t)/Z_t)$ in the following manner. The state space at time t was divided into n_3 intervals over the total interval from x_0 to x_{n3} (see Fig. 4.1), while the functions $x(s)$ and $z(s)$ were divided into n_2 intervals over the time interval t_0 to t_{n2}. At each of the n_3 discrete points of X the numerical value of the term $E\{\exp \phi/X\}$ was determined by computing ϕ_i using Eq. (4.6) n_1 times and averaging the sum of the n_1 terms $\exp \phi_i/X$ (note if $w(t) = 0$ then $n_1 = 1$).

Eq. (4.8) was then used to compute the normalizing factor $\tilde{E}\{\exp \phi\}$ and the conditional density at each point x was determined from Eq. (4.4).

Two optimality criteria were investigated during this study; that of minimum variance and maximum likelihood. To determine the minimum variance state estimate at time t the conditional mean m was computed from the expression

$$m \overset{\Delta}{=} \int X\, p(X/Z)dX \approx \frac{\sum_{k=0}^{n_3-1} X_k\, \tilde{E}\{\exp\phi/X_k\}\, \tilde{p}(X_k)[X_{k+1}-X_k]}{\tilde{E}\{\exp \phi\}} \qquad (4.10)$$

while the maximum likelihood estimate was taken as the mode of $p(X/Z)$; i.e., that X_k which gives

$$\max_{X_k} \frac{\tilde{E}\{\exp\phi/X_k\}\, \tilde{p}(X_k)}{\tilde{E}\{\exp \phi\}} \, . \qquad (4.11)$$

Upper performance-bounds were computed for a particular filter problem by conducting a number of Monte Carlo simulations and computing for each the estimation errors $e(t) = x(t) - m(t)$, where $m(t)$ was computed from Eq. (4.10) or (4.11). The average squared-error statistic for these errors was then computed at a number of instants of time and used as an upper performance bound. Approximate filters whose performances were to be compared with the upper performance-bound were simulated using the identical noise sequences as those used in computing the upper bound.

A question which naturally arises is how close to the optimal performance is the upper performance-bound as computed by the discretized Bucy representation procedure discussed above ? While a completely definitive answer to this question is impossible, a partial answer to the question can be arrived at by the following two arguments[1] . Firstly, that the values of the discretization parameters n_2 and n_3 are chosen in such a way that the upper performance bound is insensitive to these values[2] . Secondly, it has been observed that in five example problems (Sections 4.3 and 4.4) it was virtually impossible to better the upper performance-bound with that of any nonlinear filtering algorithm, no matter how complex and sophisticated the algorithm.

To those considering the use of the forementioned discrete-representation upper-bound procedure, we address the following suggestions and comments:

(i) Since the interval $[X_0, X_{n3}]$ must be chosen prior to computing $p(x(t)/Z_t)$, a relinearized Kalman filtering algorithm may be used to predict an appropriate interval size from the previously computed $p(x(s)/Z_s)$, $s < t$. One could take, for example, the 99% confidence-interval as predicted by a relinearized Kalman filter, the states of which would be updated at time t by the mean and variance of the newly computed $p(x(t)/Z_t)$.

1 The effects of too few Monte Carlo simulations are not considered here but are discussed in Section 4.3 .

2 This insensitivity suggests convergency to the continuous solution of the optimal performance function.

(ii) The values of the discretization parameters n_2 and n_3 should be chosen large enough so as to yield filter performance statistics which are relatively insensitive to the values of n_2 and n_3. Yet, the values of n_2 and n_3 should be kept as small as possible to minimize computing time.

(iii) The maximum time interval considered $[t_0, t_{max}]$, should be small enough so as to guarantee a sufficiently accurate backward numerical integration of the process state equation (4.3).

(iv) To minimize the computing time requirements of this method the two equations (4.6) and (4.7) should be computer programmed most efficiently, even resorting to machine language coding if possible.

(v) This upper performance-bound computing technique is limited by practical (not theoretical) considerations to those problems having one or two state variables, no process disturbances, and analytically computable a'priori probability density functions (although these functions could be computed via Monte Carlo simulations and orthogonal-function approximations)[1].

---

1 These limitations apply only to our method of computing upper performance-bounds, and not to the approximately optimal nonlinear filtering methods of Chapter III. This upper-bound technique is not proposed here as either a filtering scheme or a general design aid. It is strictly a research tool and nothing more.

(vi) The minimum variance optimality criterion is preferred to the maximum likelihood criterion since the maximization techniques available (e.g., steepest-descent, conjugate gradient, etc.) perform inefficiently and unreliably for the type of density functions of concern in filter theory.

4.3 Example E1 and Confidence Intervals

To demonstrate the upper performance-bound technique we consider in this section the scalar nonlinear filtering problem considered in Section 3.5 . We compare the simulated performance of two nonlinear filtering algorithms (the wide-sense and relinearized Kalman filters) with upper performance-bounds computed via the Bucy representation. Since these filter performances are computed from Monte Carlo simulations, questions arise regarding the confidence intervals for these computations. Experimentally derived curves are presented for this particular problem from which an estimate of confidence can be arrived at. Some numerical examples of the conditional probability density $p(x(t)/Z_t)$ are also presented.

Consider once again the filter problem

E1:

$$\frac{dx_1}{dt} = - x_1 \tag{3.12}$$

$$z(t) = x_1 + x_1^3 + R^{1/2}v(t) \tag{3.13}$$

$$p(x_1(0)) = N(m_{10}, \sigma_{10}^2). \tag{3.14}$$

Since the state equation (3.12) is linear and the initial state probability density is Gaussian the a'priori density of the state variable can be easily shown to satisfy

$$p(x_1(t)) = N(m_{10}\exp(-t),\ \sigma_{10}^2\exp(-2t)) \ . \qquad (4.12)$$

Bucy's representation (eq's. (4.1) - (4.3)) for problem E1 reduces to the following relations for the a'posteriori probability density, $p(x_1(t)/Z_t)$:

$$p(x_1(t) = X_1/Z_t) = \frac{\exp\{\phi/X_1 - \frac{1}{2}(X_1 - \bar{x}_1)^2/\sigma_1^2\}}{\int \exp\{\cdot\}dX_1} \qquad (4.13)$$

where

$$\bar{x}_1 = m_{10}\exp(-t)\ ,\ \sigma_1^2 = \sigma_{10}^2\ \exp(-2t)\ ,$$

and

$$\phi/X_1 = \int_{t_o}^{t} (x_1(s) + x_1^3(s))R^{-1}[z(s) - \tfrac{1}{2}(x_1(s) + x_1^3(s))]ds$$

with

$$\frac{dx_1}{ds} = -\ x_1(s)\ ;\ x_1(t) = X_1\ .$$

From Eq's. (3.1) and (3.2) the conditional moments m_1 and m_{11} for problem E1 are found to satisfy the equations

$$\frac{dm_1}{dt} = -\ m_1 + (m_{1111} + m_{11} + 3m_1m_{111} + 3m_1^2m_{11})R^{-1}\ \cdot$$

$$(z(t) - m_1 - m_1^3 - m_{111} - 3m_1m_{11}) \qquad (4.14)$$

and

$$\frac{dm_{11}}{dt} = -(m_{1111} + m_{11} + 3m_1m_{111} + 3m_1^2m_{11})^2R^{-1} - 2m_{11} + (m_{11111} + m_{111} + 3m_1m_{1111} + 3m_1^2m_{111} - m_{11}m_{111} - 3m_1m^2_{11})R^{-1}(z(t) - m_1 - m_1^3 - m_{111} - 3m_1m_{11}) \quad (4.15)$$

which yield the following nonlinear filter relations if we assume $p(x(t)/Z_t)$ to be Gaussian:

$$\frac{dm_1}{dt} = -m_1 + m_{11}(1 + 3m_{11} + 3m_1^2)R^{-1}(z - m_1 - m_1^3 - 3m_1m_{11}) \quad (4.16)$$

$$\frac{dm_{11}}{dt} = -m_{11}^2(1 + 3m_{11} + 3m_1^2)^2R^{-1} - 2m_{11} + 6m_1m_{11}^2R^{-1}(z - m_1 - m_1^3 - 3m_1m_{11}) \quad (4.17)$$

where

$$m_1(0) = m_{10}\ , \quad m_{11}(0) = \sigma_{10}^2\ .$$

Compare the Gaussian filter equations (4.16) and (4.17) with those of the relinearized Kalman filter; i.e.,

$$\frac{dm_1}{dt} = -m_1 + m_{11}(1 + 3m_{11} + 3m_1^2)R^{-1}(z - m_1 - m_1^3 - 3m_1m_{11}) \quad (4.18)$$

and

$$\frac{dm_{11}}{dt} = -m_{11}^2(1 + 3m_{11} + 3m_1^2)^2R^{-1} - 2m_{11} \quad (4.19)$$

where Eq's. (4.18) and (4.19) are determined by evaluating Eq's.

(3.8) and (3.9) for problem E1[1] .

Monte Carlo simulations were conducted for problem E1 in conjunction with a relinearized Kalman filter and a wide-sense Kalman filter (eq's. (3.17)). The error statistics for each filter were computed for up to 170 Monte Carlo simulations and these were compared with the corresponding error statistics of the upper performance-bound estimator. All simulations were conducted on the GE 4060 process control digital computer located in the Eckman Laboratory of the Systems Research Center. Rectangular integration was used in solving the filter equations and all arithmetic operations were performed in single precision.

Figure 4.2 contains the performance curves for the wide-sense and relinearized Kalman filters as well as the upper-bound estimator for two values of observation noise variance (R = .01 and .1) with the parameters $m_{10} = 1.0$ and $\sigma_{10} = 1.0$. The number of Monte Carlo simulations which was used to compute the squared-error statistics is noted on each figure along with the integration time step size, Δt .

In Monte Carlo experiments the question of confidence intervals for the results often arises. The sensitivity of the outcome of these experiments to the number of Monte Carlo experiments performed is an important factor and one which is not easily determined by analytical means. In Figure 4.2 we are primarily

---

1 In implementing the solutions to the filter equations (4.14)-(4.19), one should (theoretically, at least) include the Ito correction terms discussed in Sec. 2.2.5 to correct for the non-ideal nature of the actual observation noise process, v(t). In practice those terms have been found to appear to have no significant influence on filter performance.

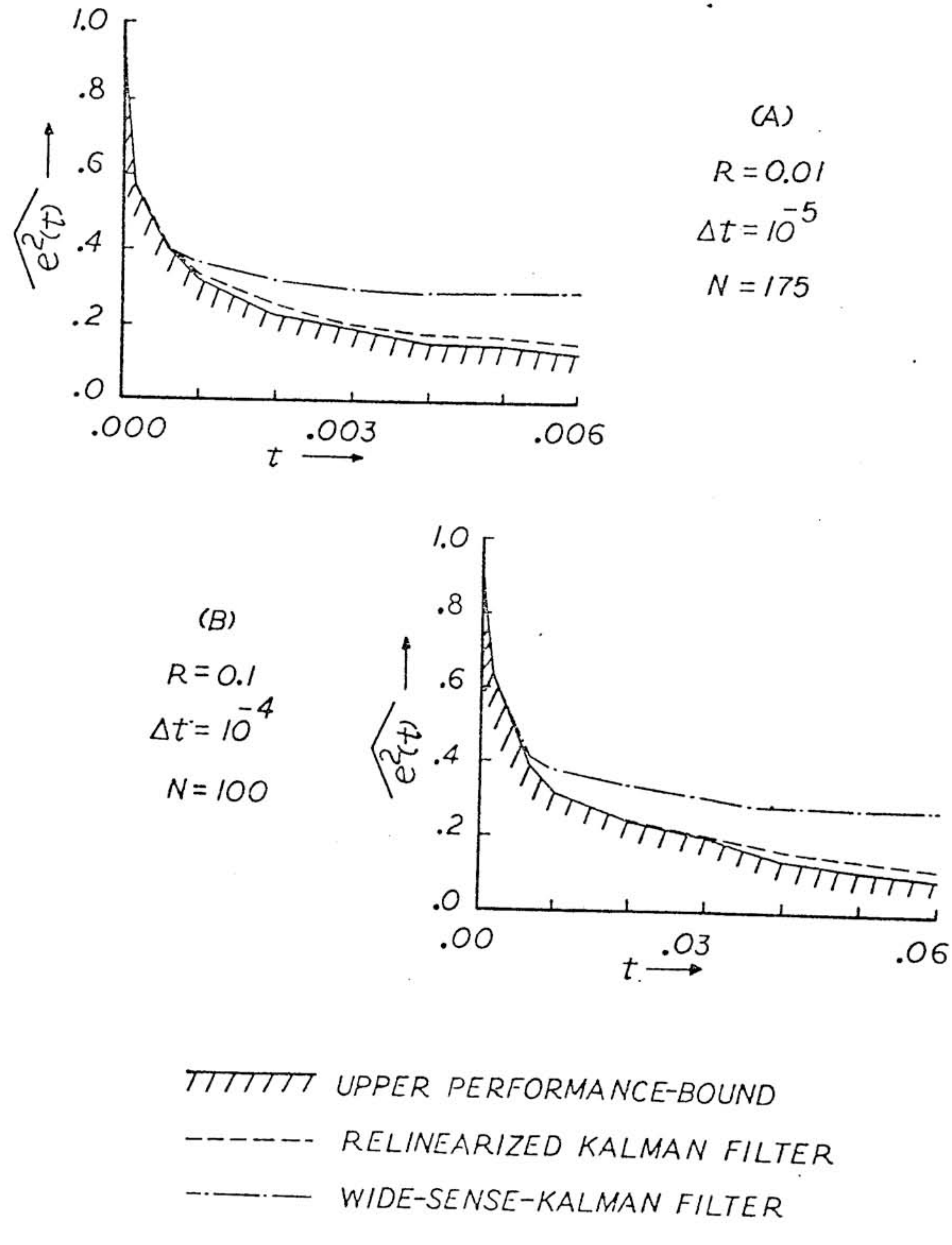


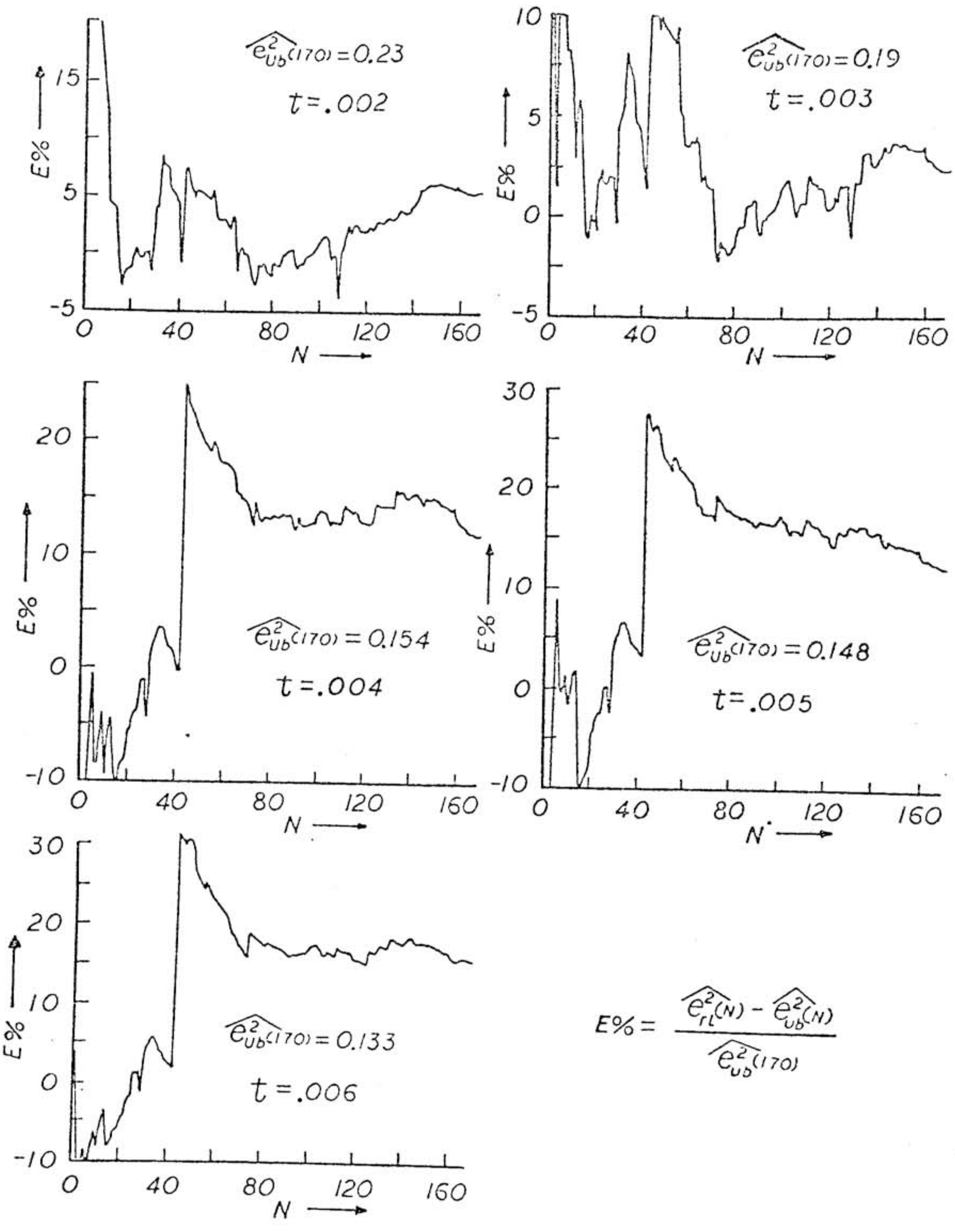
FIGURE 4.2 Average Squared-Errors for N Monte Carlo Simulations of Problem E1

concerned with the differences between the performance of the upper-bound and that of each filter tested, and it is the confidence in these computed differences which needs to be estimated, not the confidence in each individual performance curve.

The experimental curves plotted in Figure 4.3 demonstrate the effect the number of Monte Carlo simulations (N) has on the error-difference percentage factor

$$E\% = \frac{\widehat{e^2_{rl}(N)} - \widehat{e^2_{ub}(N)}}{\widehat{e^2_{ub}(170)}} \times 100\%$$

for the relinearized Kalman filter of problem E1 (R = .01) at five instants of time (.002, .003, .004, .005, and .006) . While it is doubtful that any analytically derived confidence intervals for this problem can be determined, the curves in Figure 4.3 provide an estimate of the confidence in our calculations. One may note from these figures that at times t = .004, .005, and .006 the error-difference factor becomes relatively insensitive to the number of Monte Carlo simulations performed for N greater than eighty. At earlier times (Figure 4.3), while this sensitivity is more sizeable, the error-difference factor is small enough (about five percent) so as to be of lesser importance than that at the later time instants. In general, the confidence in our Monte Carlo performance comparison calculations was considered satisfactory for one-hundred or more simulation runs.



$$E\% = \frac{\widehat{e^2_{rl}(N)} - \widehat{e^2_{ub}(N)}}{\widehat{e^2_{ub}(170)}}$$

FIGURE 4.3 Relinearized-Filter/Upper-Bound Performance Difference Ratio (E%) for N Monte Carlo Simulations of Problem E1 (R=0.01)

4.4 Experimental Problems and Results

We consider in this section four nonlinear filtering problems which were investigated in essentially the same manner as that of problem E1. For each problem, an upper performance-bound was determined via Bucy's representation and the performances (average squared - errors) were evaluated for a number of Monte Carlo simulations with the identical noise sequences being used in the simulations for each filtering technique investigated.

The maximum time intervals over which the experiments were conducted were limited by the following two factors:

(i) The intervals must be small enough to guarantee a sufficiently accurate backward numerical integration of the state equation in the upper-bound procedure.

(ii) The upper-bound procedure requires too much computer time if the maximum time interval is too great.

While it would have been desirable to utilize a larger maximum time interval in these experiments we feel that the sizeable reductions in average squared-errors which occurred over the intervals chosen justified our choice.

The four problems considered were the following[1]:

E2:

$$\frac{dx_1}{dt} = f(x_1)$$

$$z = x_1 + R^{1/2}v(t)$$

1 See Section 4.3 for problem E1.

$$f(x_1) = \begin{cases} -\frac{1}{4}x_1 \;, & x_1 \geq 0 \\ -\;\; x_1 \;, & x_1 \leq 0 \end{cases}$$

$$p(x_1(0)) = N(m_{10}, \sigma_{10}^2) .$$

E3:

$$\frac{dx_1}{dt} = f(x_1)$$

$$f(x_1) = \begin{cases} 2 + x_1 \;, & x_1 \leq -1 \\ -x_1 \;, & -1 \leq x_1 \leq 1 \\ -2 + x_1 \;, & 1 \leq x_1 \end{cases}$$

$$p(x_1(0)) = \begin{cases} 1/4 & , -2 \leq x_1(0) \leq 2 \\ 0 & , 2 < x_1(0) < -2 \end{cases}$$

$$z = x_1 + R^{1/2}v(t) .$$

E4[1]:

$$\frac{dx_1}{dt} = x_1x_2$$

$$\frac{dx_2}{dt} = 0$$

$$p(x_1(0),x_2(0)) = p(x_1(0))p(x_2(0))$$

$$p(x_i(0)) = N(m_{i0},\sigma_{i0}^2) \; ; \; i = 1,2 .$$

---

1 We are, in effect, estimating the bandwidth parameter of a system from noisy measurements of its output.

E5:

$$\frac{dx_1}{dt} = x_2$$

$$\frac{dx_2}{dt} = - x_1 - 2.5\, x_2$$

$$z = x_1 x_2 + R^{1/2} v(t)$$

$$p(x_1(0),x_2(0)) = p(x_1(0))p(x_2(0))$$

$$p(x_i(0)) = N(m_{i0},\sigma_{i0}^2) \quad i = 1,2 \; .$$

Figure 4.4 shows the results of the Monte Carlo experiments conducted with problem E2 for two values of observation noise variance ($R = 1.0, 0.1$). The performance of a relinearized Kalman filter for this problem was discovered to essentially duplicate that of the upper-bound and as such no other filtering techniques were investigated for problem E2. For comparison purposes, the a'priori performance curves for this problem are included in Figure 4.4. These curves represent the estimation accuracy one could achieve by utilizing only a'priori information and ignoring the information concerning the state provided by the noisy measurements.

A number of nonlinear filters was considered for problem E3, the performance curves for two of which are presented in Figure 4.5 for two values of observation noise variance. A relinearized Kalman filter and a uniform-density filter (i.e., the conditional moment equations with $p(x/Z)$ approximated by a uniform probability density function) were discovered to be both the simplest and most efficient

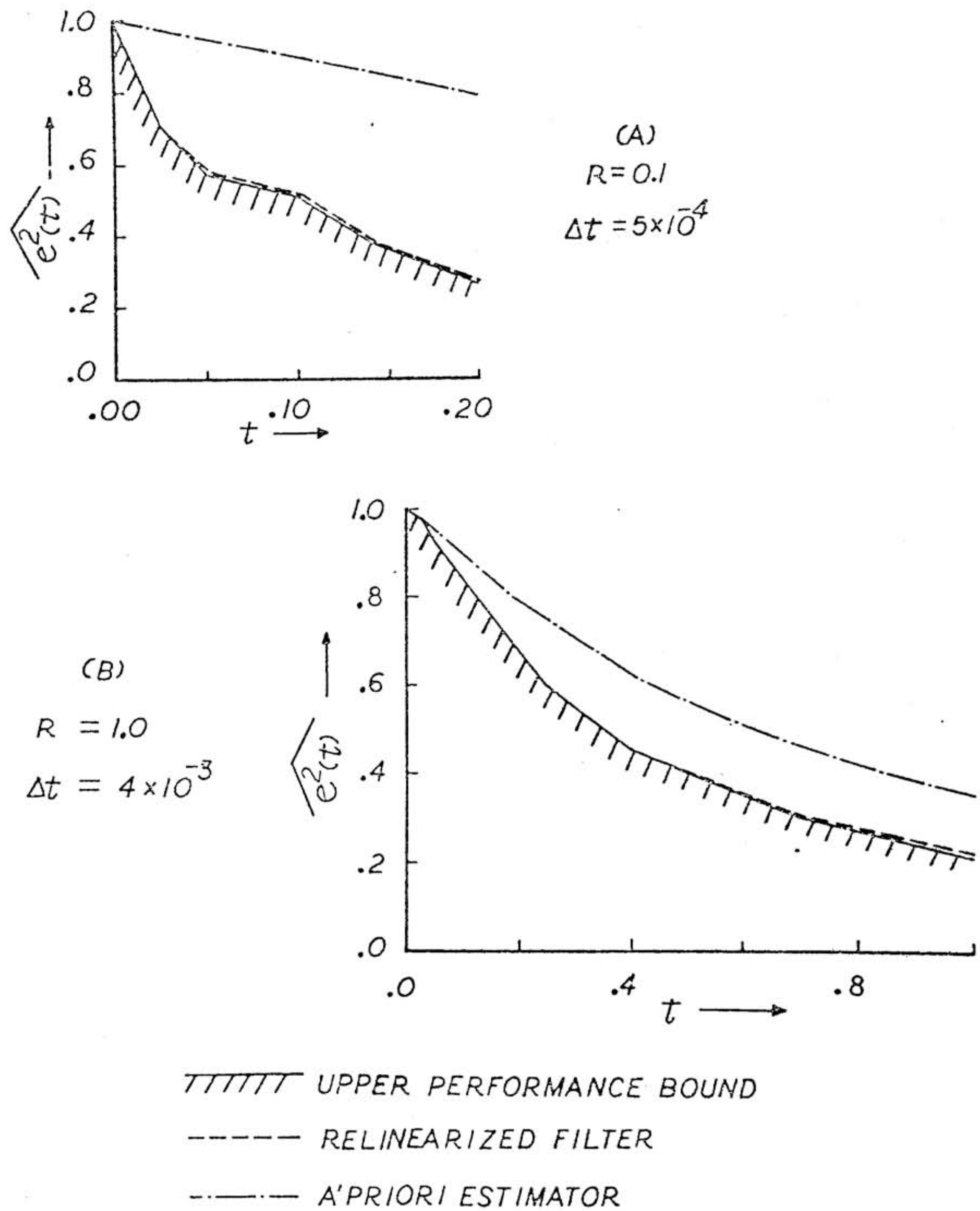


FIGURE 4.4 Average Squared-Errors for 100 Monte Carlo Simulations of Problem E2

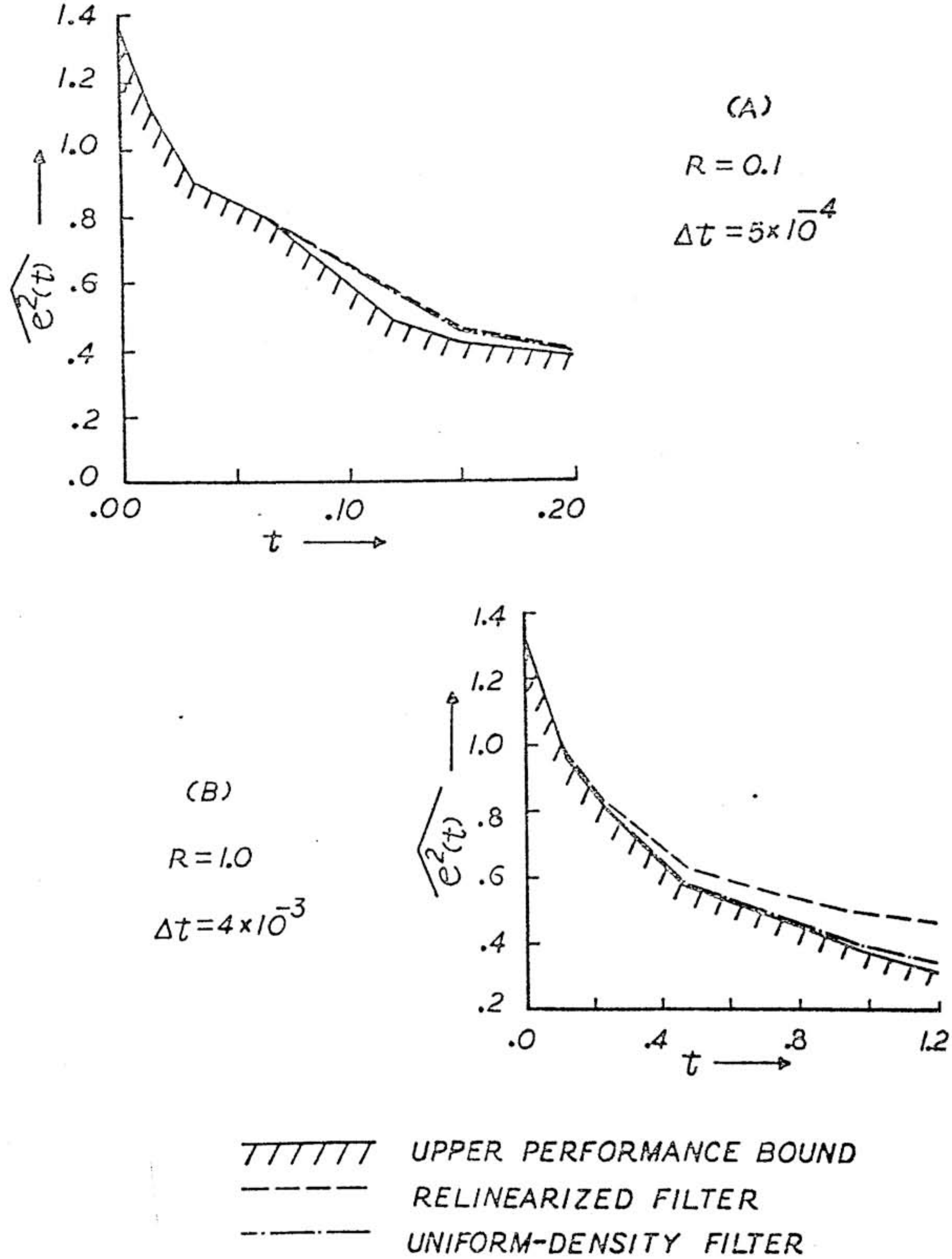


FIGURE 4.5 Average Squared-Errors for 100 Monte Carlo Simulations of Problem E3

of all the filters studied, some of which were quite complex and included third-order moment equations and approximate density functions of skewed-form (e.g., an asymmetric triangular density-function) .

In Figure 4.6 the weighted sum of the squared errors $\widehat{e_1^2} + 4\widehat{e_2^2}$ [1] is plotted against time for respectively, fifty and seventy-five Monte Carlo simulations of problem E4 and two different values of observation noise variance, R . Only a relinearized Kalman filter was investigated for problem E4 since its performance was essentially equal to that of the upper-bound. Note in Figure 4.6(A) that the relinearized filter actually achieved a slightly greater performance than that of the upper-bound. This is not really a contradiction since only fifty Monte Carlo simulations were performed and the confidence intervals for fifty runs were still sizeable enough to account for this result. Fifty simulations were considered satisfactory, however, since we were interested in detecting any significant differences which might exist between the upper performance-bound and the performance of the nonlinear filter; the one to two percent differences in performance indicated in Figure 4.6 were not considered significant.

The sum of the squared-errors $\widehat{e_1^2} + \widehat{e_2^2}$ versus time curves for fifty Monte Carlo simulations of problem E5 are graphed in Figure 4.7 for two separate values of observation noise variance.

---

1 The weighting ratio of 1:4 was chosen since the initial variance of x_1 is four times as great as that of x_2 .

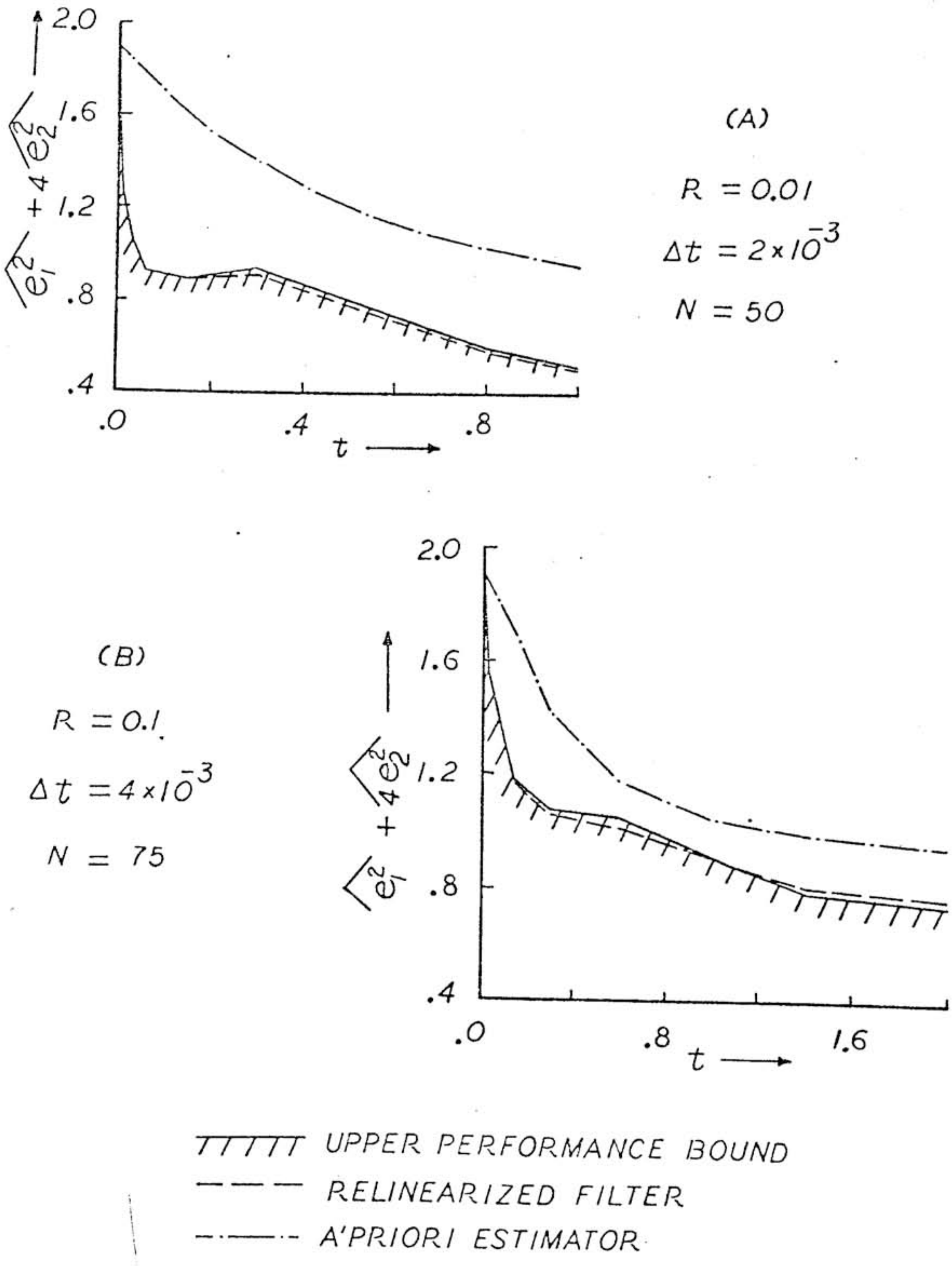


FIGURE 4.6 Weighted Average Squared-Errors for N Monte Carlo Simulations of Problem E4

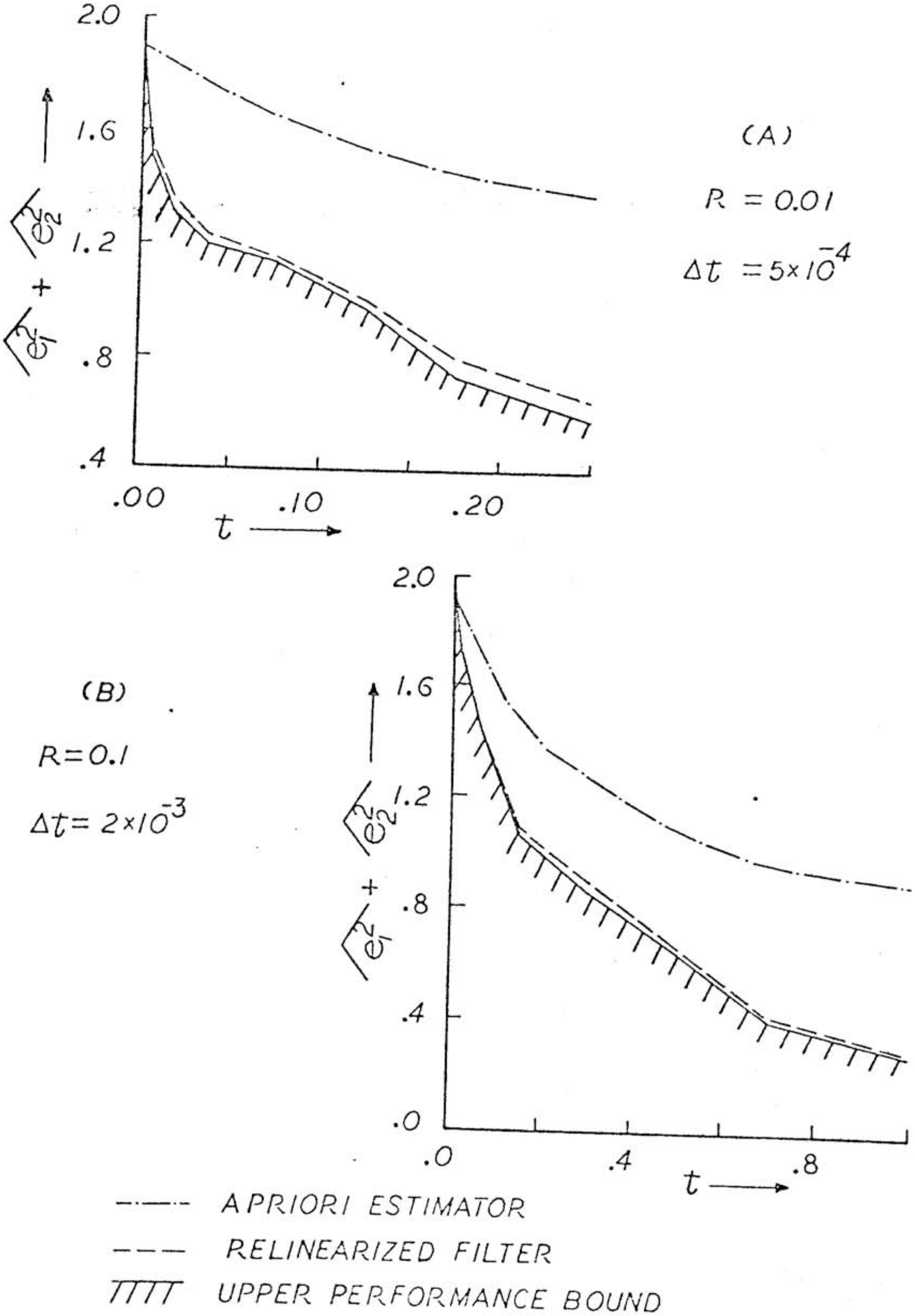


FIGURE 4.7 Summed Average Squared-Errors for 50 Monte Carlo Simulations of Problem E5

Once again the relinearized Kalman filter yielded near optimal performance and no alternative filtering techniques were investigated for this problem.

4.5 Concluding Remarks

A technique was introduced for computing numerical upper performance-bounds for relatively simple optimal nonlinear filtering problems. Using these performance bounds one is able to determine which nonlinear filtering techniques most closely approximate these optimal performances with the least amount of computational effort. Five contrived, low-dimension (one or two) optimal nonlinear filtering problems were investigated by computing their upper performance bounds and the performances of a number of nonlinear filtering algorithms.

For all five of the problems investigated the relinearized Kalman filtering method proved to be the most effective, yielding almost optimal performance with a relatively simple computational algorithm. More complex filtering methods were considered but in almost every case resulted in a lesser performance figure than that of the less complex relinearized method.

After considering the results of these five computational experiments, one might be tempted to conclude that the relinearized Kalman filtering method is the only such method which needs to be considered. This temptation, though strong, should be repressed and for a number of good reasons. First, generalizations should not be drawn from only five simple example problems; more such examples need to be studied using similar procedures to those discussed in this chapter. Secondly, the relinearized filtering technique requires that the partial derivatives in $\nabla h(x)$ and $\nabla f(x)$ exist; what does one do when they don't exist ? Finally, a fairly simple nonlinear filtering problem exists which can not be effectively handled by applying the relinearized Kalman method.

Consider the linear parameter estimation problem specified by

E6:

$$\frac{dx_1}{dt} = -x_1 + x_2^{1/2}w(t)$$

$$\frac{dx_2}{dt} = 0$$

$$p(x_1(0),x_2(0)) = p(x_1(0))p(x_2(0))$$

$$p(x_i(0)) = N(m_{i0},\sigma_{i0}^2)$$

$$z = x_1 + R^{1/2}v(t) \quad .$$

Then from Eq's. (2.33) - (2.35) the conditional moments for $p(x(t)/Z_t)$ can be easily shown to satisfy

$$\frac{dm_1}{dt} = - m_1 + m_{11}R^{-1}(z - m_1)$$

$$\frac{dm_2}{dt} = m_{12}R^{-1}(z - m_1)$$

$$\frac{dm_{11}}{dt} = - m_{11}^2R^{-1} - 2m_{11} + m_2 + m_{111}R^{-1}(z - m_1)$$

$$\frac{dm_{22}}{dt} = - m_{12}^2R^{-1} + m_{122} R^{-1}(z - m_1)$$

$$\frac{dm_{12}}{dt} = - m_{12}m_{11}R^{-1} - m_{12} + m_{112}R^{-1}(z - m_1)$$

$$\frac{dm_{111}}{dt} = - 3m_{111} + 3m_{12} - 3m_{11}R^{-1}m_{111} + (m_{1111} - 3m_{11}^2)R^{-1}(z - m_1)$$

$$\frac{dm_{112}}{dt} = - 2m_{112} + m_{22} - 2m_{11}R^{-1}m_{112} - m_{12}R^{-1}m_{111} + (m_{1112} - 3m_{11}m_{12})R^{-1}(z - m_1)$$

$$\frac{dm_{122}}{dt} = - m_1 m_{22} - 2m_{12}R^{-1}m_{112} - m_{11}R^{-1}m_{122} + (m_{1122} - m_{11}m_{22} - 2m_{12}^2)R^{-1}(z - m_1)$$

$$\vdots \qquad\qquad \vdots$$

The relinearized Kalman filter for problem E6 consists of the equations (see Eq's. (3.8) and (3.4))

$$\frac{dm_1}{dt} = - m_1 + m_{11}R^{-1}(z - m_1)$$

$$\frac{dm_2}{dt} = m_{12}R^{-1}(z - m_1) \qquad (4.20)$$

$$\frac{dm_{11}}{dt} = - m_{11}^2 R^{-1} - 2m_{11} + m_2$$

$$\frac{dm_{22}}{dt} = - m_{12}^2 R^{-1}$$

$$\frac{dm_{12}}{dt} = - m_{12}m_{11}R^{-1} - m_{12} \qquad (4.21)$$

where

$$m_i(0) = m_{i0}$$

$$m_{ii}(0) = \sigma_{i0}^2$$

$$m_{12}(0) = 0 \ .$$

Note from Eq. (4.20) that the estimated value of x_2 (m_2) can only change if the value of the cross-correlation, m_{12}, is nonzero. But since $m_{12}(0) = 0$, we see from Eq. (4.21) that dm_{12}/dt is initially and always of zero value. As a result, the relinearized Kalman filter provides no useful information regarding the value of x_2; a rather perplexing observation for such an apparently simple filtering problem.

An effective filtering algorithm for problem E6 was constructed from the eight conditional moment equations given above. The observation driving terms in the third-order moment equations were neglected by assuming that the conditional density $p(x(t)/Z_t)$ is essentially Gaussian; i.e.,

$$m_{1111} - 3m_{11}^2 \simeq 0$$

$$m_{1112} - 3m_{11}m_{12} \simeq 0$$

$$m_{1122} - m_{11}m_{22} - 2m_{12}^2 \simeq 0 .$$

Fifty Monte Carlo simulations of this filter for problem E6 were conducted and the average squared-error statistics for these simulations are presented in Figure 4.8 as a function of time. The necessary inclusion of third-order moments in the above filter can be understood when one considers that it is the correlation between the powers of the observation signal (i.e., x_1^2) and the disturbance process $x_2^{1/2}w(t)$ (i.e., x_2) which can be sensed, not the cross-correlation between the instantaneous values of $x_1(t)$ and x_2; $x_2^{1/2}w(t)$ being an ideal white-noise process.

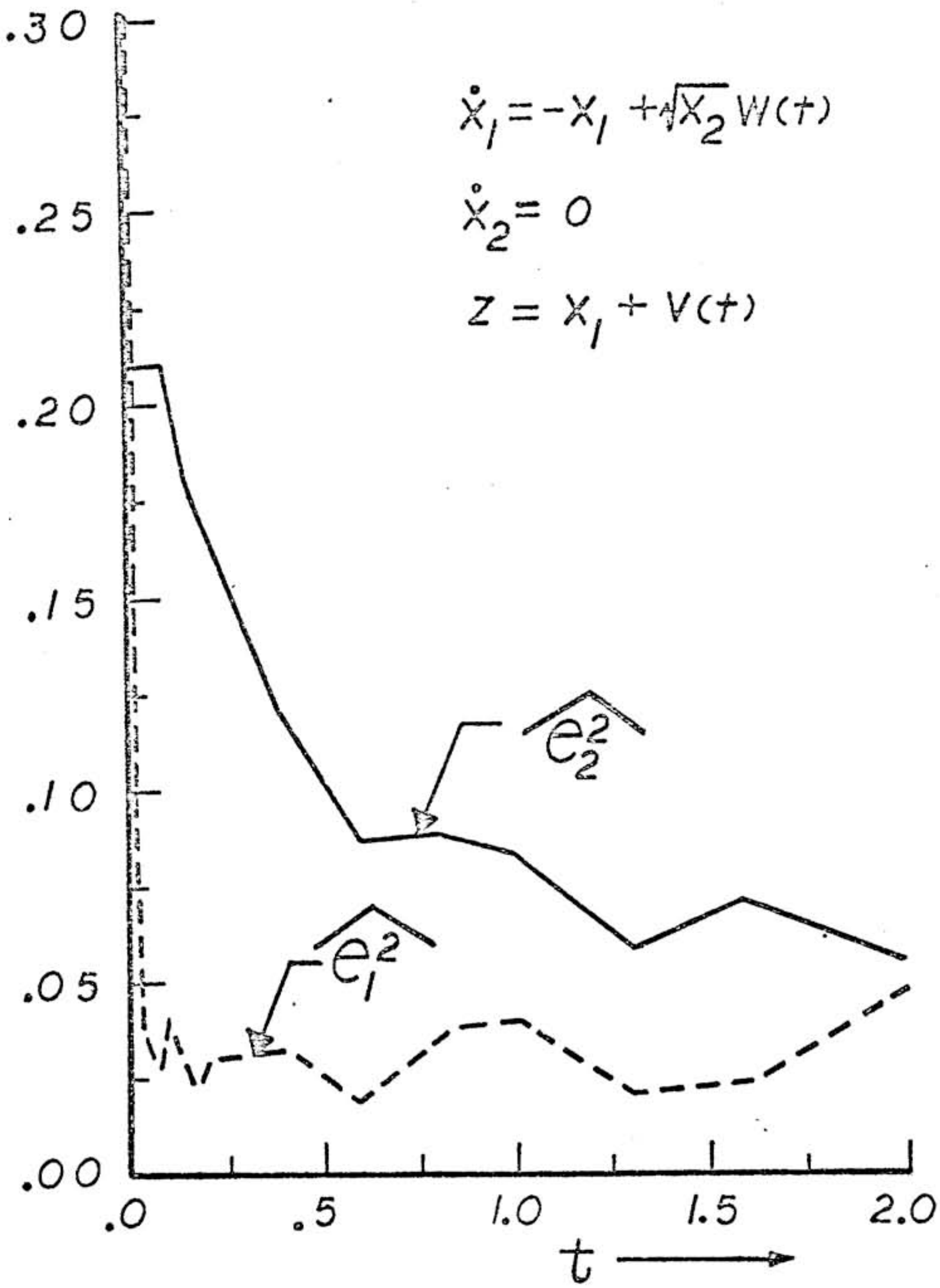


FIGURE 4.8 Average Squared-Errors for 50 Monte Carlo Simulations of Problem E6

While the relinearized Kalman filtering algorithm has proven to be a remarkably effective approximate solution to the optimal nonlinear filter problem, it can not be considered a panacean solution to this problem. It appears from the results of our experiments that the relinearized Kalman filter either works well or it doesn't work at all in actual use. From the limited number of filtering problems considered in this chapter it would be premature of us to conclude when or when not the relinearized Kalman filter should be used. Indications are, however, that this filter may be the best choice for those problems having these attributes:

- The system functions $f(x,t)$ and $h(x,t)$ are differentiable in the variable x.
- The disturbance function $g(x,t)$ is independent of the variable x.
- The measurement sampling rates (assuming a computer implementation) are high enough in value to be able to treat the problem as a continuous one.

CHAPTER V

ERROR SOURCES AND PERFORMANCE SENSITIVITY IN NONLINEAR FILTERING

5.1 Introduction

The approximations discussed in the previous chapters are but one source of error in nonlinear filtering, and are the source of error with which most of the pertinent literature has been concerned. This apparent preoccupation with mathematical approximations to the optimal solution to the nonlinear filtering problem might be attributed partly to the sincere belief held by some researchers that these approximations were the most significant source of filter error and partly to academic expediency by other concerned individuals. While the five particular filter problems investigated in the previous chapter are admittedly rather simple examples of a generally complex class of problems, an intuitive interpretation of the results of these investigations would lead one to believe that the differences between the optimal and nonoptimal filter performances are not really significant at all. There exist other sources of error in nonlinear filtering; sources which the filter designer should be aware of (as well as wary of).

In this chapter we enumerate some of the many error sources that can and in fact must exist in optimal nonlinear filtering algorithms. Of course, the critical question is how sensitive are filter performances to each error source? We consider two ways of computing these sensitivities for particular nonlinear filtering problems and present two example problems for discussion.

The significance of this chapter should not be overlooked. In attempting to successfully apply the optimal nonlinear filter theory to a real problem, one must consider more than just the mathematics of the theory. Accurate process models and statistical data are necessary ingredients to an effective filtering algorithm.

5.2 Computing Performance Sensitivities

Modelling stochastic physical processes by nonlinear differential equations with white-noise driving terms and random initial conditions provides far from ideal models. Still these same models when used in conjunction with a nonlinear filtering algorithm can provide quite satisfactory filter performance. Similarly, we can model the observation process as a nonlinear function of state plus additive white-noise and while the model may not give an accurate representation of reality it can be sufficiently accurate to allow a satisfactory filter performance to be achieved. In applying filter theory we must deal with uncertain process models and it behooves us to determine which of the inherent uncertainties can produce the greatest degradation in filter performance, and to minimize these particular uncertainties as much as possible.

We list here some of the more dominant sources of uncertainty in the process models used for nonlinear filtering applications, and while each source will have an effect on filter performance, the effect will not necessarily be a negative one.

(i) Inaccurate state space representations for process models (e.g., considering only the dominant dynamic modes).

(ii) Mathematical white-noise descriptions of physical stochastic processes possessing often significant auto-correlation coefficients.

(iii) Insufficient and inaccurate a'priori statistical knowledge about the state vector.

(iv) Inaccurate observation noise representations (Is it really additive ? Are its statistics constant or are they actually state dependent ? What are these statistics ? What about measurement bias ?).

These four items should not be passed over lightly for they do in fact determine whether or not a particular filtering application will be a success. Insufficient knowledge about the process can cause a far greater deterioration in filter performance than can mathematical approximations to the optimal filter solution.

In this section we are concerned with ways of computing the performance sensitivities of a filter to various process description uncertainties, such as items (i) - (iv) above. Two techniques are considered; the first being original and more sophisticated than the second which utilizes Monte Carlo simulations in a rather brute-force fashion.

5.2.1 A Fokker-Planck Method

Having chosen a particular filter configuration, we may desire to estimate the performance (e.g., mean squared-error) of that filter without performing tedious and often times difficult Monte Carlo simulations (see Sect. 5.2.2). In addition it would be desirable to include in this estimation the performance degradation effects of inaccuracies in noise models, process models, and initial statistics. The following method is considered.

The combined process and filter dynamics satisfy the following sets of differential equations:

$$\frac{dx}{dt} = f(x,t) + g(x,t)w(t) \quad : \quad \text{Process State Space}$$

$$z(t) = h(x,t) + R^{1/2}v(t) \quad : \quad \text{Observation Process}$$

$$\frac{dq}{dt} = F(q,t) + G(q,t)z(t) \quad : \quad \text{Filter Space}$$

where

$$E\{x(0)\} = m_o$$

$$E\{(x(0) - m_o)(x(0) - m_o)^T\} = \sigma_o^2$$

$$E\{w(t)w^T(s)\} = I\ \delta(t-s)$$

$$E\{v(t)v^T(s)\} = I\ \delta(t-s)$$

and the initial conditions of $q(t)$ are determined from our uncertain estimates of the a'priori statistics of $x(0)$. If the filter space is adjoined to the process state space a markov process results which satisfies the equations

$$\frac{dx}{dt} = f(x,t) + g(x,t)w(t)$$

$$\frac{dq}{dt} = F(q,t) + G(q,t)h(x,t) + G(q,t)R^{1/2}v(t)$$

which we denote by the single multidimensional differential equation

$$\frac{dy}{dt} = \mathfrak{F}(y,t) + \mathfrak{G}(y,t)u(t) \tag{5.1}$$

where

$$y \overset{\Delta}{=} (x,q)^T \; ; \; u \overset{\Delta}{=} (w,v)^T \; .$$

It may be desirable to replace x in Eq. (5.1) by the estimation error vector $e = x - m$ to make use of the desirable covariance properties of the $e(t)$ and $m(t)$ processes.

Then the probability density of the $y(t)$ process satisfies the Fokker-Planck equation (see Sect. 2.2.4)

$$\frac{\partial p(y(t))}{\partial t} = \mathcal{L}^*[p(y(t))] \tag{5.2}$$

and if $\phi(y)$ is any regular scalar function of the y vector, then

$$\frac{d\, E\{\phi(y)\}}{dt} = E\{\mathcal{L}[\phi(y)]\} \tag{5.3}$$

where

$$\mathcal{L}^*[\cdot] = - \sum_i \frac{\partial[\mathcal{F}_i(y,t)\;\cdot]}{\partial y_i} + \frac{1}{2} \sum_{i,j} \frac{\partial^2([\mathcal{G}\mathcal{G}^T]_{ij}\;\cdot)}{\partial y_i \partial y_j}$$

$$\mathcal{L}[\cdot] = \sum_i \mathcal{F}_i(y,t) \frac{\partial\,\cdot}{\partial y_i} + \frac{1}{2} \sum_{i,j} [\mathcal{G}\mathcal{G}^T]_{i,j} \frac{\partial^2\,\cdot}{\partial y_i \partial y_j} \; .$$

Equation (5.3) may be utilized to determine the differential equations satisfied by the statistical moment parameters of $p(y(t))$ (e.g., $E\{e(t)e^T(t)\} = E\{(x-m)(x-m)^T\}$). Unfortunately, these equations form an infinite coupled-set and as a result assumptions must be made regarding the form of the probability density $p(y(t))$. Because of the usually large dimension of the y-vector, there is only one such assumption which might prove useful and that is assuming

$p(y(t))$ is Gaussian. Then we need only determine the differential equations for the mean and variance parameters of $p(y(t))$. The technique is best demonstrated by an example.

E7:

Consider the dynamic system specified by the relations

$$\frac{dx_1}{dt} = - x_1^2 + q^{1/2}w_1(t) + 1.0 \qquad (5.4)$$

$$z(t) = x_1 + r^{1/2}(t)v(t) \qquad (5.5)$$

$$\frac{dv}{dt} = - bv + bw_2(t) \qquad (5.6)$$

where

$$p(x_1(0),v(0)) = p(x_1(0))p(v(0))$$

$$p(x_1(0)) = N(m_{10},\sigma_{10}^2) \; ; \; p(v(0)) = N(m_{vo},\sigma_{vo}^2)$$

$$r^{1/2}(t) = r_\infty^{1/2} \; \bar{x}_1(t)$$

$$\frac{d\bar{x}_1}{dt} = - \bar{x}_1^2 + 1.0 \; ; \; \bar{x}_1(0) = m_{10}$$

and $w_1(t)$ and $w_2(t)$ are independent unit-variance white-noise processes. We wish to investigate the error sensitivity of the relinearized Kalman filter for this system. The filter equations are given as

$$\frac{dm_1}{dt} = - (m_{11} + m_1^2) + m_{11}R^{-1}(t)(z-m_1) + 1.0 \qquad (5.7)$$

$$\frac{dm_{11}}{dt} = - m_{11}^2R^{-1}(t) - 4m_1m_{11} + Q \qquad (5.8)$$

where

$$m_1(0) = M_{10}\,, \quad m_{11}(0) = \Sigma_{10}^2$$

and the R, Q, M_{10}, and Σ_{10} are our estimated values of respectively r, q, m_{10}, and σ_{10}. Note that the observation noise process, v(t), is not a white-noise process even though we have assumed the contrary in deriving the filter equations. We also would like to investigate the effects of observation noise correlation on the filter's performance characteristics.

If in Eq's. (5.4) - (5.5) we replace the state variable x_1 by the estimation error variable $e = x_1 - m_1$ and adjoin the process and filter state spaces, the following adjoined state space equations result:

$$\frac{dy_1}{dt} = - y_1^2 - 2y_1y_2 + y_3 - y_1y_3R^{-1} - y_3y_4R^{-1}r^{1/2} + q^{1/2}w_1(t)$$

$$\frac{dy_2}{dt} = - y_3 - y_2^2 + y_1y_3R^{-1} + y_3y_4R^{-1}r^{1/2} + 1.0$$

$$\frac{dy_3}{dt} = - y_3^2R^{-1} - 4y_2y_3 + Q \tag{5.9}$$

$$\frac{dy_4}{dt} = - by_4 + bw_2(t)$$

where $$y = (e, m_1, m_{11}, v)^T .$$

If $\frac{dy}{dt} = \mathcal{F}(y) + \mathcal{G}(y)w(t)$, then Fokker-Planck principles dictate that

$$\frac{dc_i}{dt} = E\{\mathcal{F}_i\}$$

and

$$\frac{dc_{ij}}{dt} = E\{\mathcal{L}[(y_i - c_i)(y_j - c_j)]\}$$

where

$$c_i = \int y_i p(y)dy$$

$$c_{ij} = \int (y_i - c_i)(y_j - c_j)p(y)dy$$

$$\mathcal{L}[\cdot] = \sum_i \mathcal{F}_i \frac{\partial \cdot}{\partial y_i} + \frac{1}{2}\sum_{i,j} [\mathcal{G}\mathcal{G}^T]_{ij} \frac{\partial^2 \cdot}{\partial y_i \partial y_j} .$$

By assuming $p(y)$ to be Gaussian we can derive the following set of differential equations for c_i and c_{ij} from the relations above and the adjoined space equations (5.9):

$$\dot{c}_1 = - c_{11} - c_1^2 - 2c_{12} - 2c_1c_2 + c_3 - (c_{13} + c_1c_3)R^{-1} - (c_{34} + c_3c_4)R^{-1}r^{1/2}$$

$$\dot{c}_2 = - c_3 - c_{22} - c_2^2 + (c_{13} + c_1c_3)R^{-1} + (c_{34} + c_3c_4)R^{-1}\dot{r}^{1/2} + 1.0$$

$$\dot{c}_3 = - (c_{33} + c_3^2)R^{-1} - 4c_{23} - 4c_2c_3 + Q$$

$$\dot{c}_{11} = 2(-2c_1c_{11}-2(c_1c_{12}+c_2c_{11}) + c_{13} - (c_1c_{13} + c_3c_{11})R^{-1} - (c_3c_{14} + c_4c_{13})R^{-1}r^{1/2}) + q$$

$$\dot{c}_{22} = - 2c_{23} - 4c_2c_{22} + 2(c_1c_{23}+c_3c_{12})R^{-1} + (c_4c_{23}+c_3c_{24})R^{-1}r^{1/2}$$

$$\dot{c}_{33} = - 4c_3c_{33}R^{-1} - 8(c_3c_{23} + c_2c_{33})$$

$$\dot{c}_{44} = - 2bc_{44} + b^2$$

$$\dot{c}_{12} = - c_{13} - 4c_2c_{12} + (c_1c_{13}+c_3c_{11}-c_1c_{23}-c_3c_{12})R^{-1} + c_{23} +$$
$$(c_3c_{14}+c_4c_{13}-c_3c_{24}-c_4c_{23})R^{-1}r^{1/2} - 2c_1c_{12} - 2c_1c_{22}$$

$$\dot{c}_{13} = (-3c_3c_{13}-c_1c_{33})R^{-1} - 6c_2c_{13} - 4c_3c_{12} - 2c_1c_{13} - 2c_1c_{23} +$$
$$c_{33} - (c_3c_{34} + c_4c_{33})R^{-1}r^{1/2}$$

$$\dot{c}_{14} = - bc_{14} - 2c_1c_{14} - 2(c_1c_{24}+c_2c_{14}) + c_{34} -$$
$$(c_1c_{34} + c_3c_{14})R^{-1} - (c_4c_{34}+c_3c_{44})R^{-1}r^{1/2}$$

$$\dot{c}_{23} = (-2c_3c_{23}+ c_3c_{13}+c_1c_{33})R^{-1} - 6c_2c_{23} - 4c_3c_{22} - c_{33} +$$
$$(c_3c_{34} + c_4c_{33})R^{-1}r^{1/2}$$

$$\dot{c}_{24} = - bc_{24} - c_{34} - 2c_2c_{24} + (c_1c_{34} + c_3c_{14})R^{-1} +$$
$$(c_4c_{34} + c_3c_{44})R^{-1}r^{1/2}$$

$$\dot{c}_{34} = - bc_{34} - 2c_3c_{34}R^{-1} - 4(c_2c_{34} + c_3c_{24})$$

$$r = r_\infty \bar{x}^2 (t) \; ; \; R = R_\infty \bar{x}^2 (t)$$

where the initial conditions may be shown to satisfy

$$c_1(0) = m_{10} - M_{10} \; ; \; c_{11}(0) = \sigma_{10}^2$$
$$c_2(0) = M_{10} \; ; \; c_{44}(0) = b/2$$
$$c_3(0) = \Sigma_{10}^2 \; ; \; \text{all other } c_{ij} = 0 .$$
$$c_4(0) = 0$$

These fourteen ordinary differential equations were integrated numerically on the GE 4060 digital computer to determine the effects of parametric design errors in initial means (M_{10} vs. m_{10}), initial variance (Σ_{10}^2 vs. σ_{10}^2), and observation noise variance (R vs. r). In addition the equations were utilized to determine the effects of observation noise auto-correlation (b) and variance (r) on the filter performance. The results of these studies are presented in Figures 5.1 - 5.3 .

Figure 5.1 demonstrates the effect of observation noise variance, r = R, on the performance ($\widehat{e^2}$) versus time curves for the relinearized Kalman filter of problem E7 . Note the apparently asymptotic behavior of those performance curves as the value of R is increased. The limiting performance is the problem's a'priori (or measurement-less) performance.

A somewhat surprising result is obtained from the performance curves in Figure 5.2 for four different values of the observation noise correlation parameter, b, with the value of the equivalent white-noise variance R identical for each curve (the area under the v(t) process autocorrelation curve is independent of the value of b). It appears that as the observation noise process, v(t), becomes increasingly correlated (i.e., b decreases in value) the performance of the filter also increases. Hence, the approximation of a physical observation noise process by an equivalent ideal white-noise process will probably yield a conservative estimate of filter performance.

If our estimate, R, of the true observation noise variance, r, is in error, then the performance of the relinearized filter

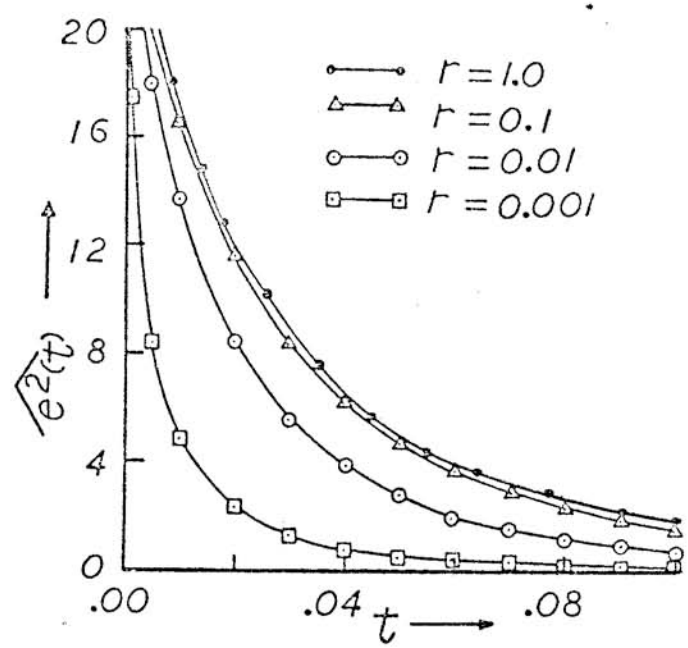


FIGURE 5.1 Mean Squared-Errors for Problem E7 with Four Values of Observation Noise Variance (r)

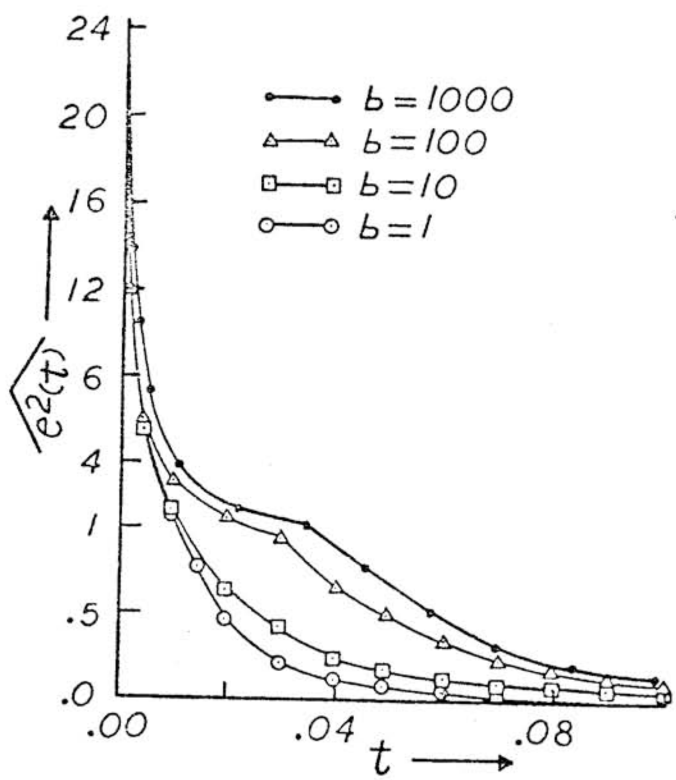


FIGURE 5.2 Mean Squared-Errors for Problem E7 with Four Different Values of Observation Noise Auto-Correlation (b) and r=0.001

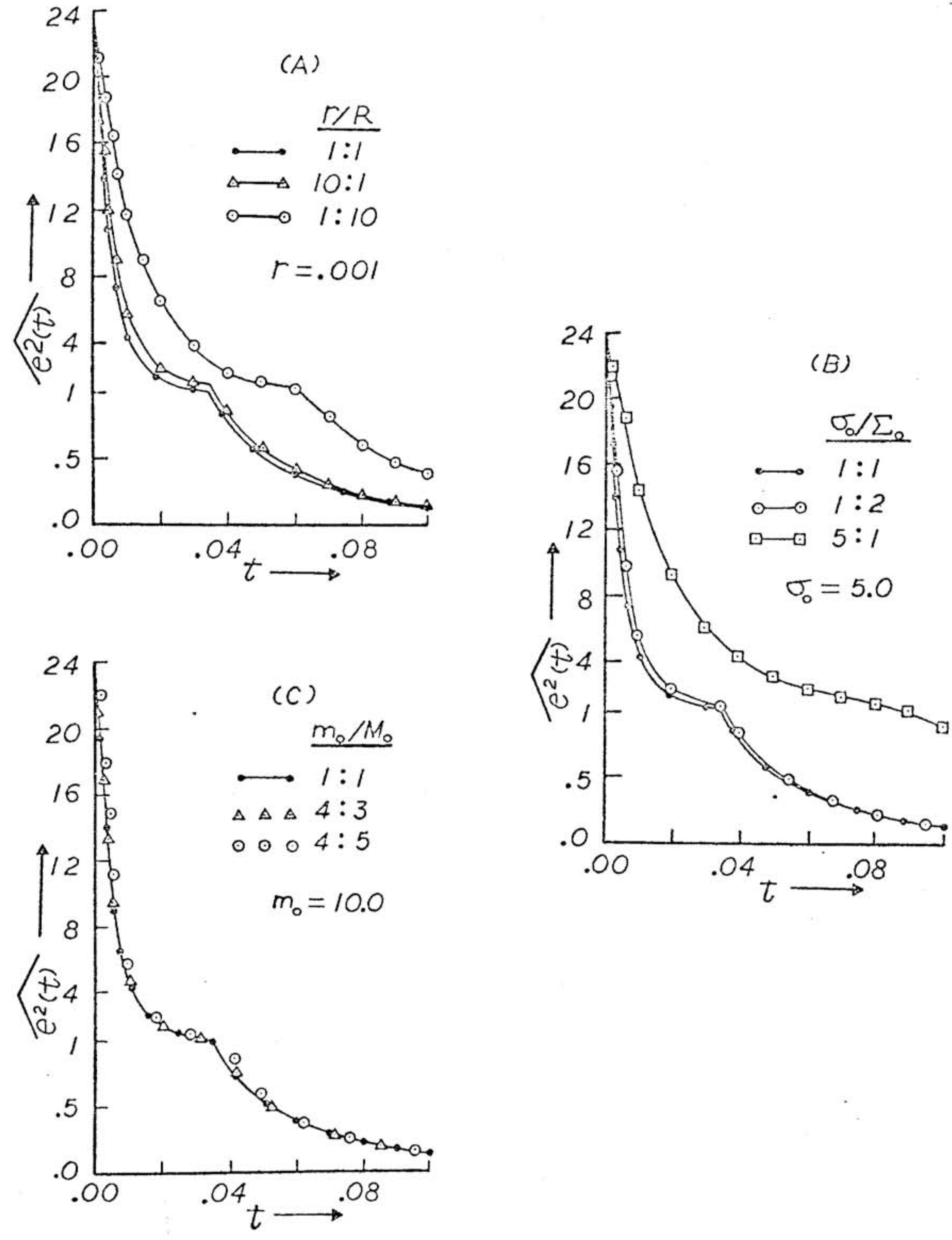


FIGURE 5.3 Mean Squared-Errors for Problem E7 with R, M_0, and Σ_0 Parameter Errors (All Parameters Nominal Valued Unless Otherwise Stated)

for problem E7 will be less than optimal. Figure 5.3(A) illustrates this effect quite clearly.

A similar behavior is demonstrated in Figure 5.3(B) which contains performance curves for various errors in our estimate, Σ_{10}^2 , of the initial state density variance, σ_{10}^2 . In the same way the influence the error in our estimate of the initial mean, m_{10} , has on filter performance is indicated in Figure 5.3(C).

5.2.2 A Monte Carlo Method

A superficial acquaintance with the forementioned Fokker-Planck method of computing performance sensitivities for nonlinear filters can be very misleading. Under scrutiny this technique reveals itself to be most impractical for applications to significant nonlinear filter problems. We have considered it in this thesis primarily to demonstrate the inherent difficulties associated with any "so-called" sophisticated solutions to the nonlinear filter sensitivity problem.

What are the difficulties in applying the Fokker-Planck sensitivity method to significant nonlinear filtering problems? Firstly, is the dimensionality problem for even if we can rightly restrict ourselves to the mean and covariance parameters of $p(y)$, the number of moment (e.g., c_1 and c_{ij}) equations would become quite prohibitive for even the simplest nonlinear filter problems. Secondly, is the requirement that the system functions $f(x)$, $h(x)$, and $g(x)$ be expressed as polynomials in x so that the expectations of those functions and their functionals can be computed under the assumption of a Gaussian probability density $p(y)$. Finally, and most importantly, the higher than second-order, odd moment parameters of $p(y)$ (i.e., c_{ijk}, etc.) can not always be neglected. This final difficulty was clearly demonstrated while applying the Fokker-Planck method to the following parameter estimation problem:

E8[1]:

$$\frac{dx_1}{dt} = x_1x_2 + x_3$$

$$\frac{dx_2}{dt} = 0$$

$$\frac{dx_3}{dt} = -bx_3 + q^{1/2}w(t) \qquad (b = q = 1.0)$$

$$z_1 = x_1 + r_1^{1/2}v_1(t)$$

$$z_2 = x_3 + r_2^{1/2}v_2(t)$$

where

$$p(x_1(0),x_2(0),x_3(0)) = p(x_1(0))p(x_2(0))p(x_3(0))$$

$$p(x_i(0)) = N(m_{io},\sigma_{io}^2) \quad i = 1,2,3$$

and the statistical parameters of the $x_3(t)$ process (i.e., b, q, m_{30}, and σ_{30}) are not known. The relinearized Kalman filter for problem E8 consists of the relations

$$\frac{dm_1}{dt} = m_1m_2 + m_{12} + z_2(t) + m_{11}R_1^{-1}(z_1-m_1)$$

$$\frac{dm_2}{dt} = m_{12}R_1^{-1}(z_1-m_1)$$

$$\frac{dm_{11}}{dt} = -m_{11}^2R_1^{-1} + 2m_1m_{12} + 2m_2m_{11} + R_2$$

---

1 Note that the $x_3(t)$ process represents a correlated disturbance process acting on the $x_1(t)$ process possessing an unknown bandwidth x_2. Noisy measurements of the disturbance, $x_3(t)$, are available.

$$\frac{dm_{22}}{dt} = - m_{12}^2 R_1^{-1} \quad ; \quad \frac{dm_{12}}{dt} = -m_{11} m_{12} R_1^{-1} + m_2 m_{12} + m_1 m_{22}$$

with

$$m_i(0) = M_{io}$$

$$m_{ii}(0) = \Sigma_i^2 \quad ; \quad i = 1,2$$

$$m_{12}(0) = \Sigma_{12} \; .$$

The above process and filter equations were adjoined and the differential equations for the first and second-order moment parameters of the adjoined state probability density, $p(y)$, were determined under the Gaussian assumption for $p(y)$. Certain of these equations were clearly unrealistic in that a number of moments c_{ij} remained at a zero value for all time even though physical reasoning required that these same c_{ij} become finite valued with increasing time. Odd-order moment variables (e.g., c_{ijk}) were unquestionably necessary in the moment equation for $p(y)$ but their inclusion increased the dimensionality of the problem to such an extent that the Fokker-Planck method was ruled out for determining the sensitivities of the nonlinear filter for problem E8.

As an alternative to the Fokker-Planck method we chose to investigate the parameter sensitivities of the nonlinear filter for problem E8 by means of Monte Carlo simulation. A nominal performance curve was first established by performing one hundred Monte Carlo simulations of a filter having the correct parameter values (i.e., $R = r$, $M_{10} = m_{10}$, $\Sigma_{10} = \sigma_{10}$, etc.). To determine the sensitivity of the filter performance to an error in one of the process parameters a certain percentage error in that parameter was assumed and the one-hundred simulations were repeated with the identical noise sequences as those used in the nominal performance curve calculations. The difference between the nominal and in-error performance curves provides an estimate of the filter's error sensitivity. Clearly, those errors to which the filter is most sensitive should be minimized as much as possible.

The results of the sensitivity study of the nonlinear filter for problem E8 are presented pictorially in Figures 5.4 - 5.10. In each figure the weighted sum of the squared-errors $\widehat{e_1^2} + 4\widehat{e_2^2}$ (since $\sigma_{10}^2/\sigma_{20}^2 = 4:1$) is plotted against time for one-hundred Monte Carlo simulations. Each figure deals with a single source of error common to all nonlinear filters, not just that of problem E8.

Figure 5.4 demonstrates how measurement-noise variance uncertainty affects filter performance. The curves in these figures reveal that twenty-five percent errors in our estimated values of r_1 and r_2 have no really significant effect on filter performance while a 300 percent over estimate of the value of r_1 (i.e., $R_1/r_1 = 4:1$) yields a filter performance which is still very tolerable.

A major source of uncertainty in any nonlinear filter exists in the statistics of the state vector at the time instant the filter is "turned-on". Two such statistics are the initial means m_{10} and m_{20} for problem E8. The filter's performance sensitivities to

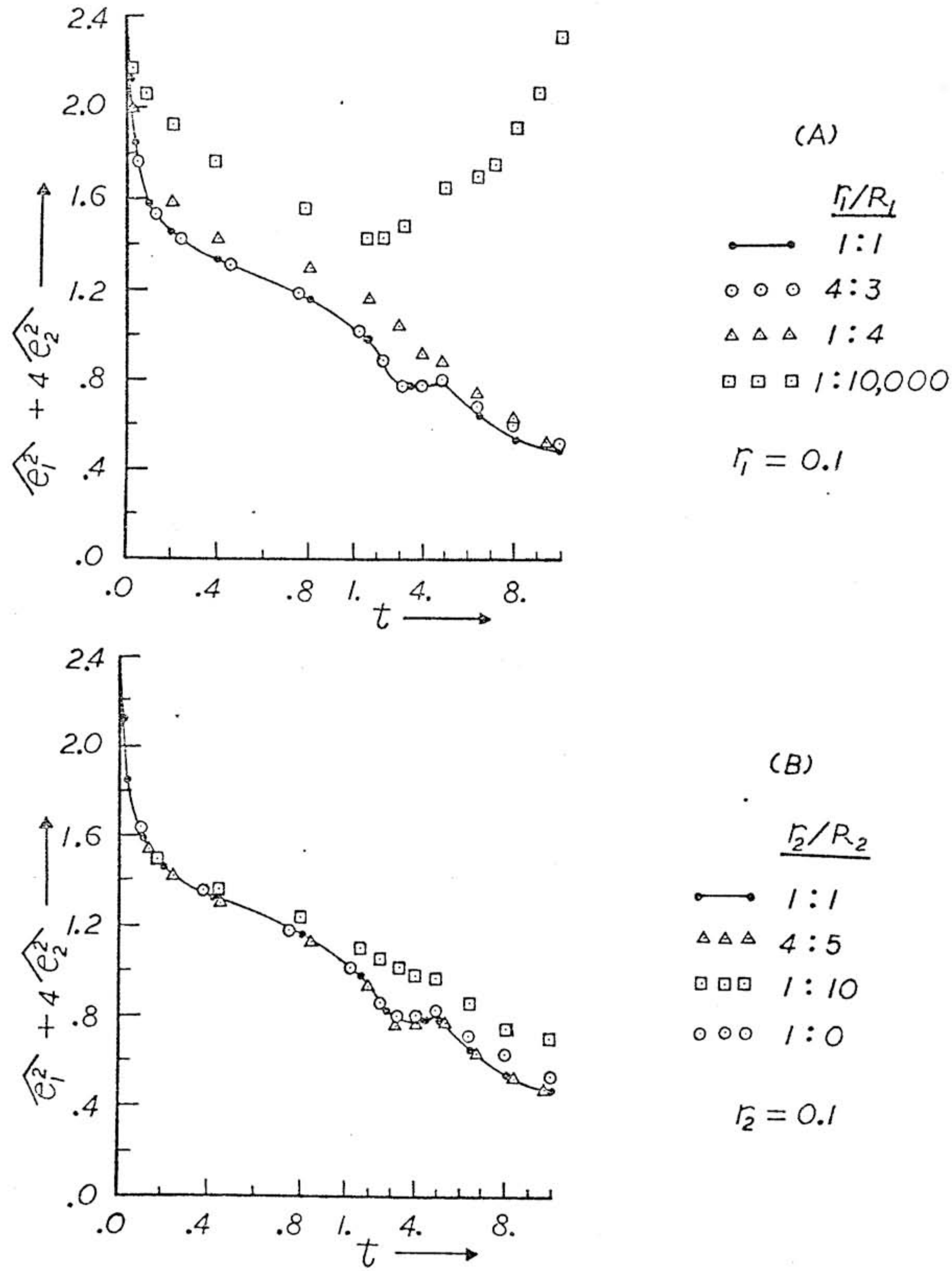


FIGURE 5.4 Weighted Average Squared-Errors for 100 Simulations of Problem E8 with R_1 and R_2 Parameter Errors

errors in our estimates (M_{10} and M_{20}) of those two parameters are revealed in Figure 5.5. One may observe from Figure 5.5(A) that the performance of the relinearized Kalman filter for this problem is relatively insensitive to errors in the initial mean-value estimates of x_1. Figure 5.5(B) on the other hand, reveals that the filter is far more sensitive to errors in the a priori estimated mean value of the second state variable, x_2. Note especially in Figure 5.5(B) the drastic reduction in performance which resulted from a one-hundred percent over estimate in the parameter m_{20} (i.e., $M_{20}/m_{20} = 2:1$). The a'priori estimated value of x_2 should not be taken lightly for this problem.

Uncertainty in our estimates of the variances of $p(x_1(0))$ and $p(x_2(0))$ proved to have an effect on filter performance which was similar to that of the mean value estimates of those probability densities. Figure 5.6(A) shows that the filter performance is relatively insensitive to error in estimates of σ_{10} while Figure 5.6(B) reveals σ_{20} to have a more influential effect on filter performance than that of σ_{10}. In determining the curves plotted in Figure 5.6(C) we wrongly assumed that a certain amount of correlation (σ_{12}) existed between the random variables $x_1(0)$ and $x_2(0)$. The curves in this figure show that a fairly large amount of initial covariance had to be assumed before a significant reduction in performance resulted.

So far it was assumed that the observation noise sequences used in the simulations were uncorrelated from sample to sample. In physical processes of course there exists a certain and sometimes

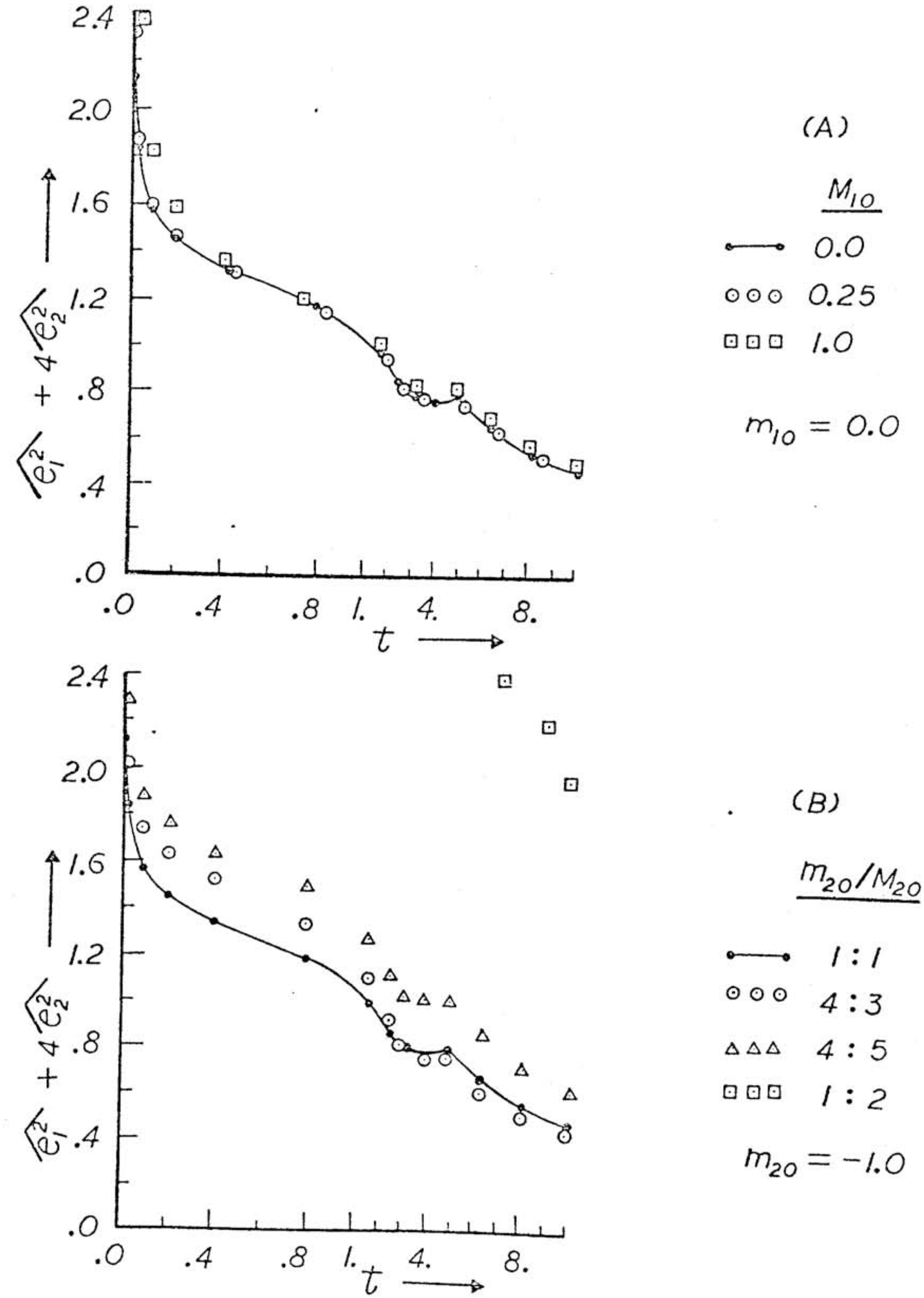


FIGURE 5.5 Weighted Average Squared-Errors for 100 Simulations of Problem E8 with M_{10} and M_{20} Parameter Errors

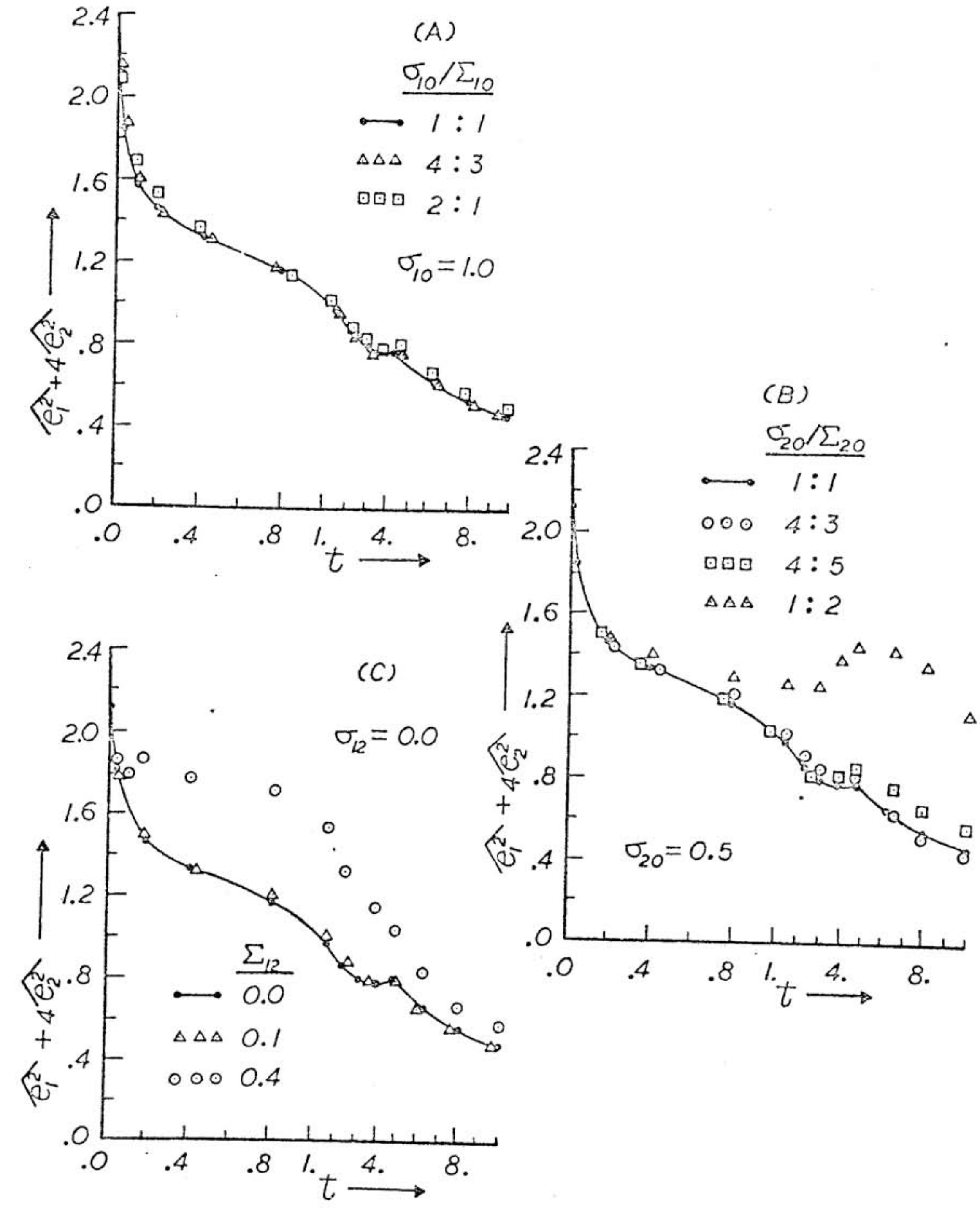


FIGURE 5.6 Weighted Average Squared-Errors for 100 Simulations of Problem E8 with Σ_{10}, Σ_{20}, and Σ_{12} Parameter Errors

sizeable amount of correlation between separate samples. We investigated the effects of non-white observation noise on filter performance for problem E8 by replacing the output observation equation by the "equivalent" correlated version

$$z_1 = x_1 + R_1^{1/2} v(t)$$

where

$$\frac{dv}{dt} = -cv + cv_1(t) \; .$$

Two values of c were considered and the performance curves resulting with each are plotted in Figure 5.7 . One may note from this figure that the effects of observation noise correlation are, if anything, positive effects since the performance appears to increase with increasingly correlated (i.e., smaller values of c) observation noise, $v(t)$.

Another major source of error in applications of nonlinear filter theory is the process model simplifications which are sometimes necessary. For our study of problem E8 such an error source was created by assuming that the actual process state equations were given as

$$\frac{dx_1}{dt} = x_1 x_2 + x_4$$

$$\frac{dx_2}{dt} = 0$$

$$\frac{dx_3}{dt} = -b\,x_3 + q^{1/2} w_1(t)$$

$$\frac{dx_4}{dt} = -d\,x_4 + d\,x_3$$

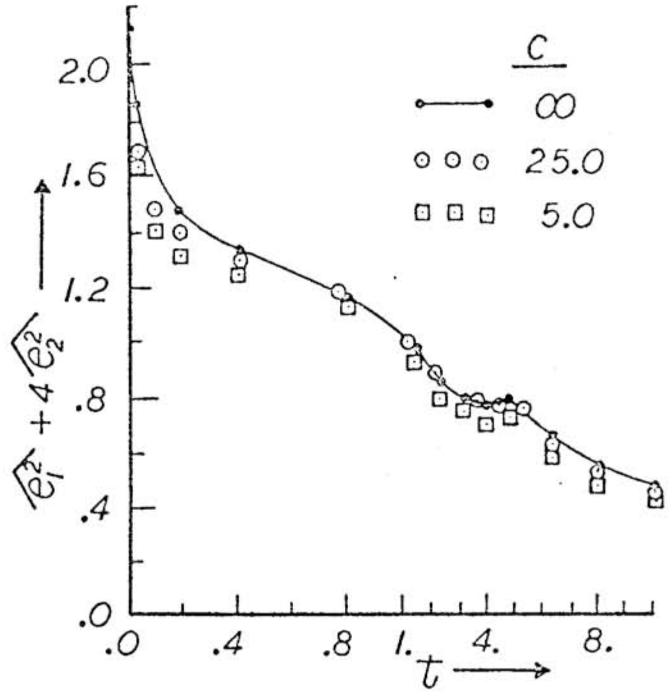


FIGURE 5.7 Weighted Average Squared-Errors for 100 Simulations of Problem E8 with Three Values of Observation-Noise Correlation (c)

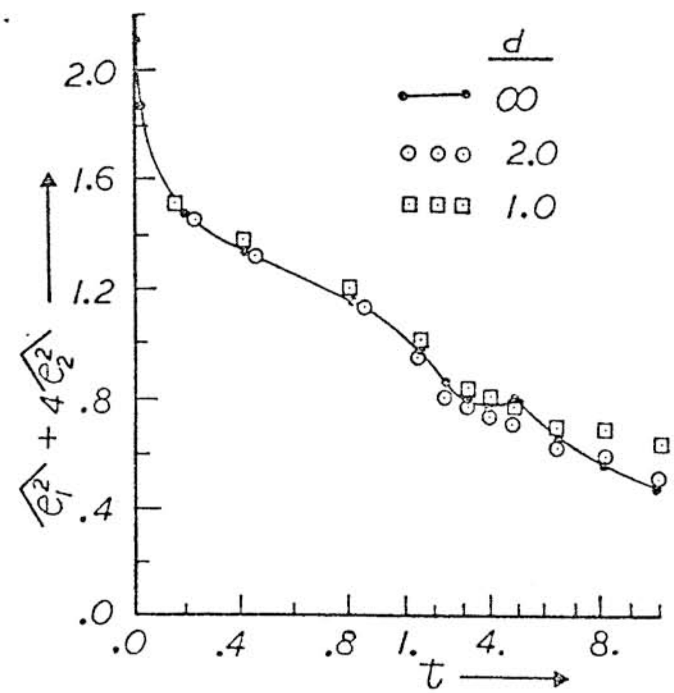


FIGURE 5.8 Weighted Average Squared-Errors for 100 Simulations of Problem E8 with 3 Values of Neglected Process Bandwidth (d)

rather than their simplified forms assumed in the derivation of the nonlinear filter for this problem. The influence of the neglected process lag, d, on filter performance is indicated in the curves of Figure 5.8 . For the most part, the effects of this lag were negligible, a result which was probably due to the relatively small bandwidth of the input process $x_3(t)$ $(b = 1.0)$.

One basic assumption in the theory of nonlinear filtering is that the observation noise process is Gaussian distributed. What is the effect of approximating a non-Gaussian observation noise by an "equivalent" Gaussian process ? To assist in answering this question 200 Monte Carlo simulations of problem E8 were performed with an observation noise process $v_1(t)$ which was uniformly distributed with a variance equal to that of the previously studied Gaussian distributed $v_1(t)$ process. The relinearized Kalman filtering algorithm used remained the same as that utilized in the Gaussian noise case. The two simulated performance curves for the Gaussian and uniform-density observation noise processes are presented in Figure 5.9 . One will note from this figure that there appears to be only an inconsequential difference between the two curves. This result is encouraging in that the filter designer does not have to concern himself with the particular type of probability distribution function the observation noise possesses; assuming it to be Gaussian appears satisfactory.

The final source of error considered in this abbreviated study was that of state-dependent observation noise. For most measurement

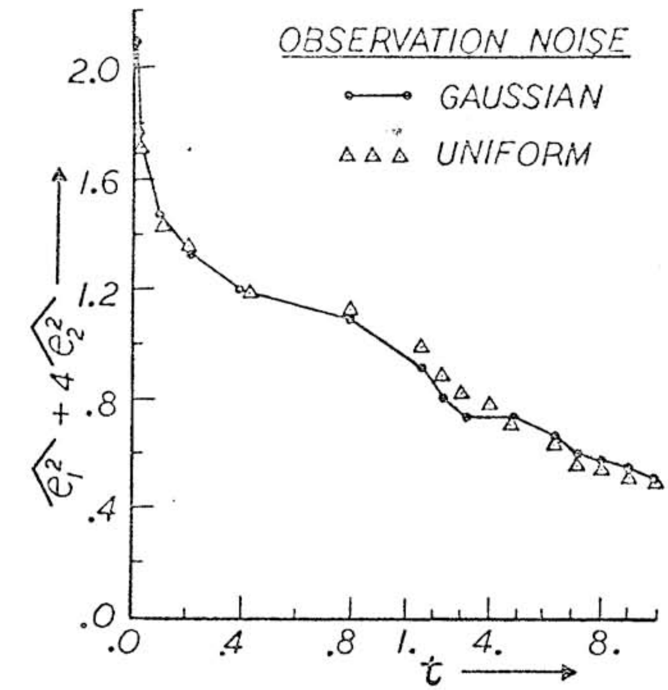


FIGURE 5.9 Weighted Average Squared-Errors for 200 Simulations of Problem E8 with "Equivalent" Gaussian- and Uniform-Observation Noise (v_1)

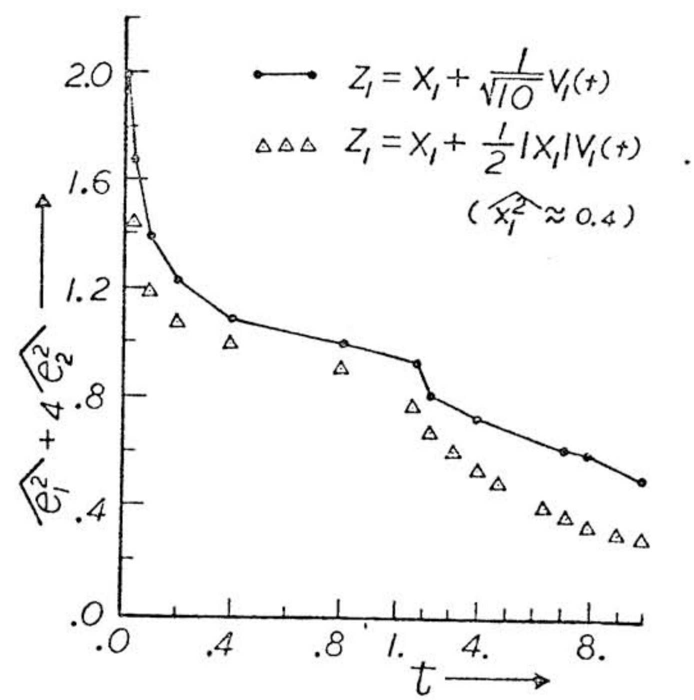


FIGURE 5.10 Weighted Average Squared-Errors for 200 Simulations of Problem E8 with "Equivalent" State-Dependent and-Independent Noise $v_1(t)$

instruments their accuracy is specified in terms of percentage of reading. Thus, the amount of uncertainty in measurements made with these instruments is proportional to the average value of the measurements, and is not of constant value as assumed in the general formulation of the nonlinear filtering problem. To investigate the effects of approximating a state-dependent observation noise process by an equivalent state-independent noise process we replaced the output observation equation in problem E8,

$$z_1 = x_1 + r_1^{1/2} v_1(t) \; ; \; r_1 = 0.1 \, ,$$

by the "equivalent " process equation

$$z_1 = x_1 + 0.5 |x_1| v_1(t) \, ,$$

where the average squared-value of the $x_1(t)$ process was determined to be approximately four-tenths of a squared unit. Two-hundred Monte Carlo simulations of the relinearized Kalman filter were conducted in conjunction with, first, the state-independent noise process and, second, the state-dependent noise process. The simulated performance curves which resulted for both cases are plotted in Figure 5.10 . The results are somewhat surprising in that the performance of the filter with a state-dependent observation noise was greater than the filter performance with a state-independent noise. How general this particular result is remains in question.

5.3 Concluding Remarks

We considered in this chapter two techniques for computing the sensitivity of nonlinear filter performance to various errors which exist in all nonlinear filtering problems. While the first technique is quite sophisticated, it requires the numerical solution of such a large number of differential equations as to be virtually useless for any practical filtering problems which may be of interest. This problem of large dimensionality is inherent to any analytical solution of the nonlinear filter sensitivity problem. The second technique discussed was simply that of "brute-force" Monte Carlo simulations with purposefully introduced modelling and parameter errors. While lacking analytical sophistication, this latter technique is most effective and highly recommended.

While we will not attempt to draw any general conclusions from the two admittedly simple filter problems investigated in this chapter, the error sources discussed for each example are common to all nonlinear filter problems and the significance of each and others not mentioned must be ascertained by the filter designer if a truly effective filtering performance is to be achieved. The approximation problem in optimal nonlinear filtering is not restricted to just the mathematical approximations discussed in the preceding chapters of this thesis, but includes all the modelling and statistical approximations which must be applied to a process in the interests of practicality.

CHAPTER VI

AN APPLICATION OF NONLINEAR FILTERING: TIME-DELAY ESTIMATION

6.1 Introduction

Two major approximation problems associated with most practical applications of optimal nonlinear filter theory are concerned with simplifying the process models and simplifying the optimal solution to the filter problem. The latter problem was considered in Chapters III and IV of this thesis while the former was discussed in the preceding chapter. An application which clearly illustrates how these problems may affect the performance of a nonlinear filter is that of on-line time-delay estimation via nonlinear filtering. We consider this particular application in this chapter.

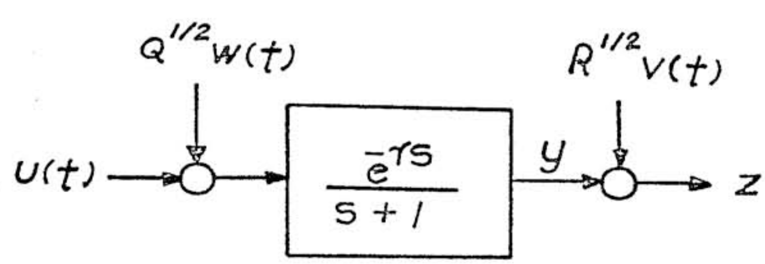


FIGURE 6.1 Block Diagram of Time-Delay Estimation Problem

The linear dynamic system specified by the block diagram in Figure 6.1 could represent a simplified dynamic model of a distributed parameter system (e.g., a distillation column or a heat exchanger) acted upon by the measurable disturbance variable $u(t)$ and having the output variable $y(t)$. An unknown time-delay, γ, exists between the input and output variables and for adaptive control purposes it is desired to develop a nonlinear filter to improve our estimate of the value of γ by processing the input-output records $u(t)$ and $z(t)$. The process disturbance noise $Q^{1/2}w(t)$ is assumed to represent both the unmeasurable inputs which may act upon the process and any $u(t)$ measurement-noise which may exist. For this study we consider only the case in which $u(t)$ is a unit-step function; the more general random function case while of greater importance, is also of greater difficulty and could easily be the sole concern of an entire thesis. Similarly, we assumed for simplicity (not by necessity) that the effective time constant of the process is constant and equal to unity.

Then the process state and observation equations for the system of Figure 6.1 are given as

$$\frac{dx_1}{dt} = - x_1 + u(t - x_2) + Q^{1/2}w(t) \tag{6.1}$$

$$\frac{dx_2}{dt} = 0 \tag{6.2}$$

$$z = x_1 + R^{1/2}v(t) \tag{6.3}$$

where

$$x = (y,\gamma)^T$$

$$u(\tau) = \begin{cases} 1 \ , & \tau \geq 0 \\ 0 \ , & \tau < 0 \end{cases}$$

and

$$p(x_1(0),x_2(0)) = p(x_1(0))p(x_2(0))$$

$$p(x_i(0)) = N(m_{io},\sigma_{io}^2)$$

with

$$m_{10} = 0 \ ; \ m_{20} = 1.0$$

$$\sigma_{10} = 0.25 \ ; \ \sigma_{20} = 0.50 \ .$$

One may inquire as to why a nonlinear filter is required to estimate the time-delay, γ. Wouldn't the amount of dead-time present be apparent from the step response of the process ? If the noise intensities (i.e., Q and R) were low enough and the step-input amplitude high enough the answer would be in the affirmative. Unfortunately, such conditions exist primarily in textbooks since in most on-line applications the step-input amplitude must be kept as small as possible so that the process output does not stray significantly from the desired operating point. For in the control of physical systems, linear models are usually perturbation models and the accuracy of these models varies inversely with the perturbation size. Low amplitude step-inputs, of course, will produce low signal-to-noise ratios for which it is the function of the nonlinear filter to compensate as optimally as possible. Nonlinear filters do not ignore nor nullify the deleterious effects of noise but instead recognize and minimize those effects as much as statistically possible.

Two nonlinear filtering techniques for the time-delay estimation problem are considered in this chapter. The first technique models the time-delay in the process by a Padé approximation and utilizes a relinearized Kalman filter to estimate the time-delay parameter, γ . In the second technique no approximations were made to the basic process model and a uniform-density nonlinear filtering algorithm was developed for the estimator. The results of simulation studies conducted for both these techniques were somewhat surprising and also encouraging.

6.2 A Padé Model Filter

Since the function $u(t - x_2)$ in Eq. (6.1) is nondifferentiable at time zero a direct application of the relinearized Kalman filter algorithm to the time-delay estimation problem is impossible. As an alternative method we chose to approximate the dead-time transfer function of the problem by a first order Padé approximation.[1] The block diagram for the approximate process model is given in Figure 6.2 . Then the process and observation equations

---

1 Another possibility would have been that of approximating the step-function $u(t)$ by the exponential function $1 - \exp(-ct)$ where the value of ϵ would be chosen to maximize performance.

for the approximate model of Figure 6.2 are given as

$$\frac{d\tilde{x}_1}{dt} = -\tilde{x}_1\tilde{x}_3 + \tilde{x}_2(1 + \tilde{x}_3) - u(t) - Q^{1/2}w(t)$$

$$\frac{d\tilde{x}_2}{dt} = -\tilde{x}_2 + u(t) + Q^{1/2}w(t)$$

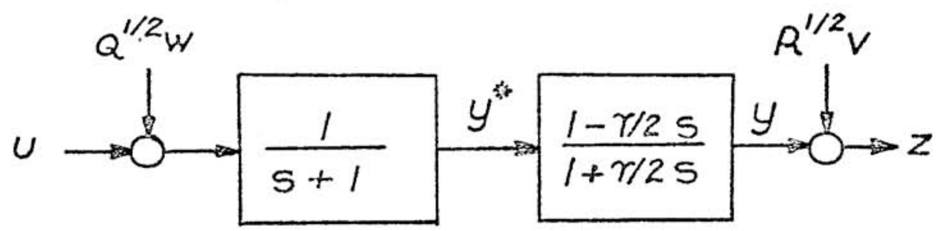


FIGURE 6.2 Block Diagram for Padé Model of Time-Delay Estimation Problem

$$\frac{d\tilde{x}_3}{dt} = 0$$

$$z = \tilde{x}_1 + R^{1/2}v(t)$$

where

$$x = (y, y^*, 2/\gamma)^T$$

$$p(\tilde{x}_1(0),\tilde{x}_2(0),\tilde{x}_3(0)) = p(\tilde{x}_1(0))p(\tilde{x}_2(0))p(\tilde{x}_3(0))^{1}$$

$$p(\tilde{x}_1(0)) = p(\tilde{x}_2(0)) = N(m_{10},\sigma_{10}^2)$$

$$p(\tilde{x}_3(0)) = p(\gamma) \left| \frac{d\gamma}{dx_3} \right|_{\gamma = \frac{2}{\tilde{x}_3(0)}} .$$

The relinearized Kalman filter for the above approximate time-delay process model consists of the following equations:

$$\frac{dm_1}{dt} = - m_1m_3 - m_{13} + m_2(1 + m_3) + m_{23} - u(t) + m_{11}R^{-1}(z - m_1)$$

$$\frac{dm_2}{dt} = - m_2 + u(t) + m_{12}R^{-1}(z - m_1)$$

$$\frac{dm_3}{dt} = m_{13}R^{-1}(z - m_1)$$

$$\frac{dm_{11}}{dt} = - m_{11}^2R^{-1} - 2m_1m_{13} - 2m_3m_{11} + 2m_{12} + 2m_2m_{13} + 2m_3m_{12} + Q$$

$$\frac{dm_{22}}{dt} = - m_{12}^2R^{-1} - 2m_{22} + Q$$

$$\frac{dm_{33}}{dt} = - m_{13}^2R^{-1}$$

$$\frac{dm_{12}}{dt} = - m_{11}m_{12}R^{-1} - m_{12} - m_1m_{23} - m_3m_{12} + m_{22} + m_2m_{23} + m_3m_{22} - Q$$

---

1 To justify the assumption that $\tilde{x}_1(0)$ and $\tilde{x}_2(0)$ are independent note that prior to the application of the unit step $\tilde{x}_1(t)$ and $\tilde{x}_2(t)$ respond solely to the random disturbances $w(t)$. It is reasonable to expect that $\tilde{x}_1(t)$ and $\tilde{x}_2(t+\gamma)$ were uncorrelated for the relatively large values of time-delay (γ) and $t \leq 0$.

$$\frac{dm_{13}}{dt} = - m_{11}m_{13}R^{-1} - m_3m_{13} - m_1m_{33} + m_{23} + m_3m_{23} + m_2m_{33}$$

$$\frac{dm_{23}}{dt} = - m_{12}m_{13}R^{-1} - m_{23}$$

where

$$m_1(0) = m_{10} \quad , \quad m_2(0) = m_{10}$$

$$m_3(0) \simeq \frac{1}{\sigma_{20}} \ln \left| \frac{m_{20} + \sigma_{20}}{m_{20} - \sigma_{20}} \right| \qquad (6.4)$$

$$m_{11}(0) = m_{22}(0) = \sigma_{10}^2$$

$$m_{33}(0) \simeq \frac{4}{(m_{20} + \sigma_{20})(m_{20} - \sigma_{20})} - m_3^2(0) \qquad (6.5)$$

$$m_{12}(0) = m_{13}(0) = m_{23}(0) = 0$$

and

$$E\{\gamma/Z_t\} = \frac{1}{\sqrt{m_{33}}} \ln \left| \frac{m_3 + \sqrt{m_{33}}}{m_3 - \sqrt{m_{33}}} \right| , \qquad (6.6)$$

The approximate relations (6.4) - (6.6) were obtained by assuming the appropriate probability density functions $(p(\gamma)$ or $p(\tilde{x}_3))$ to be uniform with the given mean and variance parameters.

One-hundred Monte Carlo simulations of the above filter acting in conjunction with the process equations (6.1) - (6.3) were conducted and the average performance of the filter was computed for these simulations. The results achieved were, to say the least, disappointing. No additional statistical information about the time-delay parameter, γ, was obtained with the relinearized Kalman filter in that the average squared-error of the estimates remained essentially constant over the entire three-unit time interval the simulations were conducted for. This poor filter performance can probably be attributed to the sizeable modelling errors introduced by the Padé approximation, an approximation which becomes less and less accurate with increasing values of time-delay. In effect, any positive information regarding the value of γ which was obtained from the measurements, $z(t)$, was nullified by the negative information introduced by the Padé approximation of the process model.

The performance achieved with this nonlinear filter was not only disappointing but also somewhat surprising in that it was thought that since the process model is part of a feedback loop in the filter, the filter would not be particularly sensitive to modelling inaccuracies. Such was clearly not the case and a further study was initiated to determine an effective filtering algorithm for the time-delay estimation problem.

6.3 Uniform-Density Filter

To design a proper filter we consider the conditional moment equations of $p(x(t)/Z_t)$ for the process representation specified by Eq's. (6.1) - (6.3). These moment equations are given as

$$\frac{dm_1}{dt} = - m_1 + \widehat{u(t - x_2)} + m_{11}R^{-1}(z - m_1)$$

$$\frac{dm_2}{dt} = m_{12}R^{-1}(z-m_1)$$

$$\frac{dm_{11}}{dt} =- m_{11}^2R^{-1} - 2m_{11} + 2\ \widehat{x_1u(t - x_2)} - 2m_1\widehat{u(t - x_2)} + Q + m_{111}R^{-1}(z - m_1)$$

$$\frac{dm_{22}}{dt} =- m_{12}^2R^{-1} + m_{122}R^{-1}(z - m_1)$$

$$\frac{dm_{12}}{dt} =- m_{11}m_{12}R^{-1} - m_{12} + \widehat{x_2u(t - x_2)} - m_2\widehat{u(t - x_2)} + m_{112}R^{-1}(z - m_1)$$

$\cdot \qquad \cdot$

$\cdot \qquad \cdot$

$\cdot \qquad \cdot$

where

$$\widehat{u(t - x_2)} = \int u(t - x_2)p(x/Z)dx$$

$$\widehat{x_1u(t - x_2)} = \int x_1u(t - x_2)p(x/Z)dx$$

$$\widehat{x_2u(t - x_2)} = \int x_2u(t - x_2)p(x/Z)dx$$

If we make the assumption that $p(x/Z_t)$ is a symmetric uniform-density function $\tilde{p}(x/Z_t)$, then symmetry requires that $m_{111} = m_{112} = m_{122} = 0$, and the moment equations become

$$\frac{dm_1}{dt} =-m_1 + \widehat{u(t - x_2)} + m_{11}R^{-1}(z - m_1)$$

$$\frac{dm_2}{dt} = m_{12}R^{-1}(z - m_1)$$

$$\frac{dm_{11}}{dt} = - m_{11}^2 R^{-1} - 2m_{11} + 2\,\widehat{x_1 u(t-x_2)} - 2m_1\,\widehat{u(t-x_2)} + Q$$

$$\frac{dm_{22}}{dt} = - m_{12}^2 R^{-1}$$

$$\frac{dm_{12}}{dt} = - m_{11}m_{12}R^{-1} - m_{12} + \widehat{x_2 u(t-x_2)} - m_2\widehat{u(t-x_2)}$$

where

$$\widehat{u(t-x_2)} \tilde{=} \int u(t-x_2)\tilde{p}(x/Z_t)dx \qquad (6.7)$$

$$\widehat{x_1 u(t-x_2)} \tilde{=} \int x_1 u(t-x_2)\tilde{p}(x/Z_t)dx \qquad (6.8)$$

$$\widehat{x_2 u(t-x_2)} \tilde{=} \int x_2 u(t-x_2)\tilde{p}(x/Z_t)dx \;. \qquad (6.9)$$

The analytical evaluation of integrals (6.7) - (6.9) in terms of the conditional moments m_1, m_2, m_{11}, m_{22}, and m_{12} is a necessary and somewhat difficult task. To simplify this task we make the following transformation of variables; let

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = G(x - m)$$

where

$$x = (x_1, x_2)^T \;, \; m = (m_1, m_2)^T \;,$$

$$G = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

and

$$\theta = \frac{1}{2} \tan^{-1} \frac{2m_{12}}{m_{22} - m_{11}}$$

Then

$$x = G^{-1}y + m$$

where

$$G^{-1} = G^T = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} .$$

Also,

$$\widehat{xu(t-x_2)} = G^T \widehat{yu(t-x_2)} + m\, \widehat{u(t-x_2)}$$

and, hence,

$$\widehat{x_1u(t-x_2)} = \widehat{y_1u(t-x_2)}\cos\theta + \widehat{y_2u(t-x_2)}\sin\theta + m_1\widehat{u(t-x_2)}$$

$$\widehat{x_2u(t-x_2)} = -\widehat{y_1u(t-x_2)}\sin\theta + \widehat{y_2u(t-x_2)}\cos\theta + m_2\widehat{u(t-x_2)} .$$

The transformed nonlinear filter equations become

$$\frac{dm_1}{dt} = -m_1 + \widehat{u(t-x_2)} + m_{11}R^{-1}(z-m_1) \tag{6.10}$$

$$\frac{dm_2}{dt} = m_{12}R^{-1}(z-m_1) \tag{6.11}$$

$$\frac{dm_{11}}{dt} = -m_{11}^2R^{-1} - 2m_{11} + Q + 2\widehat{y_1u(t-x_2)}\cos\theta + 2\widehat{y_2u(t-x_2)}\sin\theta \tag{6.12}$$

$$\frac{dm_{22}}{dt} = -m_{12}^2R^{-1} \tag{6.13}$$

$$\frac{dm_{12}}{dt} = -m_{11}m_{12}R^{-1} - m_{12} + \widehat{y_2u(t-x_2)}\cos\theta - \widehat{y_1u(t-x_2)}\sin\theta \tag{6.14}$$

where

$$m_i(0) = m_{io}\ , \qquad i = 1,2$$
$$m_{ii}(0) = \sigma_{io}^2\ ,$$
$$m_{12}(0) = 0\ ,$$

and

$$\widehat{u(t-x_2)} = \int u(t-x_2)\tilde{p}(y/Z_t)\,dy \qquad (6.15)$$

$$\widehat{y_1u(t-x_2)} = \int y_1u(t-x_2)\tilde{p}(y/Z_t)\,dy \qquad (6.16)$$

$$\widehat{y_2u(t-x_2)} = \int y_2u(t-x_2)\tilde{p}(y/Z_t)\,dy \qquad (6.17)$$

with

$$x_2 = -y_1\sin\theta + y_2\cos\theta + m_2\ .$$

The integrations in equations (6.15) - (6.17) were carried out analytically as discussed in Appendix II . The expressions for $\widehat{u(\ \cdot\ )}$, $\widehat{y_1u(\ \cdot\ )}$, and $\widehat{y_2u(\ \cdot\ )}$ were determined in terms of the parameters σ_1 , σ_2 , y_a , y_b , y_c , and y_d for each mode and are presented in Tables 6.1 - 6.4 , where:

$$\sigma_1{}^2 = \tfrac{3}{2}[m_{11} + m_{22} - \sqrt{4m_{12}^2 + (m_{22} - m_{11})^2}]$$

$$\sigma_2{}^2 = \tfrac{3}{2}[m_{11} + m_{22} + \sqrt{4m_{12}^2 + (m_{22} - m_{11})^2}]$$

$$y_a = (t - m_2)/\cos\theta - \sigma_1/\tan\theta$$

$$y_b = -\sigma_1 - (\sigma_2 + y_a)/\tan\theta$$

$$y_c = y_a + 2\sigma_1 \tan\theta$$

$$y_d = y_b + 2\sigma_2 / \tan\theta \ ,$$

and the modes are defined in Table 6.1[1].

Then the uniform-density nonlinear filter consists of the differential equations (6.10) - (6.14) for the expectation variables

TABLE 6.1 Uniform-Density Time-Delay Estimator Mode Numbers

y_c \ y_a	$\cdot \le -\sigma_2$	$-\sigma_2 < \cdot \le \sigma_2$	$\sigma_2 < \cdot$
$\cdot \le -\sigma_2$	1	2	3
$-\sigma_2 < \cdot \le \sigma_2$	4	5	6
$\sigma_2 < \cdot$	7	8	9.

$\widehat{u(\cdot)}$, $\widehat{y_1 u(\cdot)}$, and $\widehat{y_2 u(\cdot)}$ determined from Tables 6.1 - 6.4 as indirect functions of the moment variables. While the resulting filtering algorithm may appear somewhat complex, it is really quite simple to implement on a digital computer in that it is totally recursive and requires no iterative solutions of nonlinear algebraic equations. Simulation studies of this filter have revealed that the filter operates in only modes 1, 5, and 9 so that the filter

---

1 The mode of the filter is used here to refer to one of a number of possible regions of the filter state space in which certain filter relations hold true.

TABLE 6.2 Analytical Expressions for $\widehat{u(t-x_2)}$ for Modes 1-9 of the Uniform-Density Nonlinear Filter.

MODE NUMBER	$\widehat{u(t-x_2)}$
1	0
2	$\frac{(\sigma_1 + y_b)(\sigma_2 + y_a)}{8\sigma_1 \sigma_2}$
3	$\frac{2\sigma_1 + y_d + y_b}{4\sigma_1}$
4	$\frac{(\sigma_1 - y_b)(\sigma_2 + y_c)}{8\sigma_1 \sigma_2}$
5	$\frac{2\sigma_2 + y_a + y_c}{4\sigma_2}$
6	$1 - \frac{(\sigma_1 - y_d)(\sigma_2 - y_c)}{8\sigma_1 \sigma_2}$
7	$\frac{2\sigma_1 - y_d - y_b}{4\sigma_1}$
8	$1 - \frac{(\sigma_2 - y_a)(\sigma_1 + y_d)}{8\sigma_1 \sigma_2}$
9	1

TABLE 6.3 Analytical Expressions for $\widehat{y_1 u(t-x_2)}$ for Modes 1-9 of the Uniform-Density Nonlinear Filter.

MODE NUMBER	$\widehat{y_1 u(t-x_2)}$
1	0
2	$\dfrac{(\sigma_2 + y_a)(\sigma_1 + y_b)(y_b - 2\sigma_1)}{24\sigma_1 \sigma_2}$
3	$\dfrac{-3\sigma_1^2 + (y_d + y_b)^2 - y_d y_b}{12\sigma_1}$
4	$\dfrac{(y_c + \sigma_2)(\sigma_1 - y_b)(2\sigma_1 + y_b)}{24\sigma_1 \sigma_2}$
5	$\dfrac{\sigma_1 (y_c - y_a)}{12\sigma_2}$
6	$\dfrac{-(\sigma_2 - y_c)(\sigma_1 - y_d)(2\sigma_1 + y_d)}{24\sigma_1 \sigma_2}$
7	$\dfrac{3\sigma_1^2 - (y_d + y_b)^2 + y_d y_b}{12\sigma_1}$
8	$\dfrac{(y_a - \sigma_2)(\sigma_1 + y_d)(y_d - 2\sigma_1)}{24\sigma_1 \sigma_2}$
9	0

TABLE 6.4 Analytical Expressions for $\widehat{y_2 u(t-x_2)}$ for Modes 1-9 of the Uniform-Density Nonlinear Filter.

MODE NUMBER	$\widehat{y_2 u(t-x_2)}$
1	0
2	$\dfrac{(\sigma_1 + y_b)(\sigma_2 + y_a)(y_a - 2\sigma_2)}{24\sigma_1 \sigma_2}$
3	$\dfrac{-\sigma_2(y_b - y_d)}{12\sigma_1}$
4	$\dfrac{(\sigma_1 - y_b)(y_c + \sigma_2)(y_c - 2\sigma_2)}{24\sigma_1 \sigma_2}$
5	$\dfrac{-3\sigma_2^2 + (y_a + y_c)^2 - y_a y_c}{12\sigma_2}$
6	$\dfrac{-(\sigma_1 - y_d)(\sigma_2 - y_c)(2\sigma_2 + y_c)}{24\sigma_1 \sigma_2}$
7	$\dfrac{-\sigma_2(y_d - y_b)}{12\sigma_1}$
8	$\dfrac{(y_a - \sigma_2)(\sigma_1 + y_d)(y_a + 2\sigma_2)}{24\sigma_1 \sigma_2}$
9	0

relations for the remaining six modes could be neglected in any actual application of the filter. A closer examination of the filter relations would probably reveal other possible simplifications to the algorithm.

The Monte Carlo simulated performance ($\widehat{e_2^2}$ vs. time) curves for the uniform-density nonlinear filter time-delay estimator are presented in Figure 6.3 for four separate values of observation noise variance ($R = .001, .01, .1, \infty$) and 100 simulation experiments. Note how the performance approaches a non-zero constant value asymptote with increasing time. To increase the time-delay estimation accuracy further another step-input would have to be applied to the process and the filter reactivated (i.e., t reset to zero) after it reaches a steady state. Figure 6.3 also indicates that the performance curve will approach some limiting curve as the value of the observation noise variance approaches zero; this agrees with what one would expect intuitively. When contrasted with that of the Padé approximation filter, the performance of the uniform-density filter was almost startling.

Time-response curves for two simulation experiments with the uniform-density filter are presented in Figure 6.4 . Note from this figure how the covariance variable m_{22} approaches a nonzero steady-state value, as does the time-delay estimation error variable. The influence of the cross correlation variable m_{12} on the filter behavior is clear in that the most significant reduction in estimation error occurs when the value of m_{12} is the greatest. A

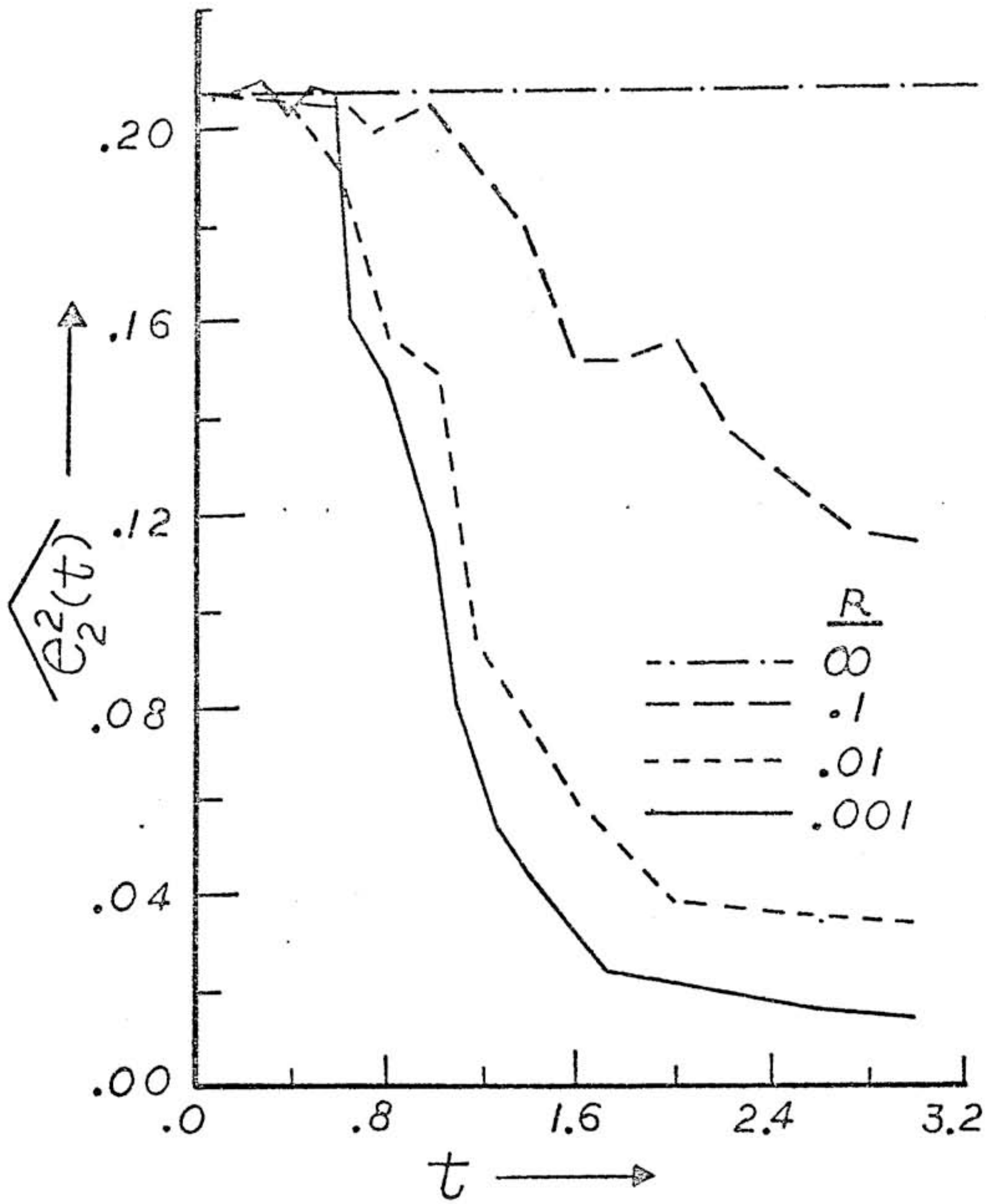


FIGURE 6.3 Average Squared-Errors for 100 Simulations of Time-Delay Estimation Problem with Four Values of R

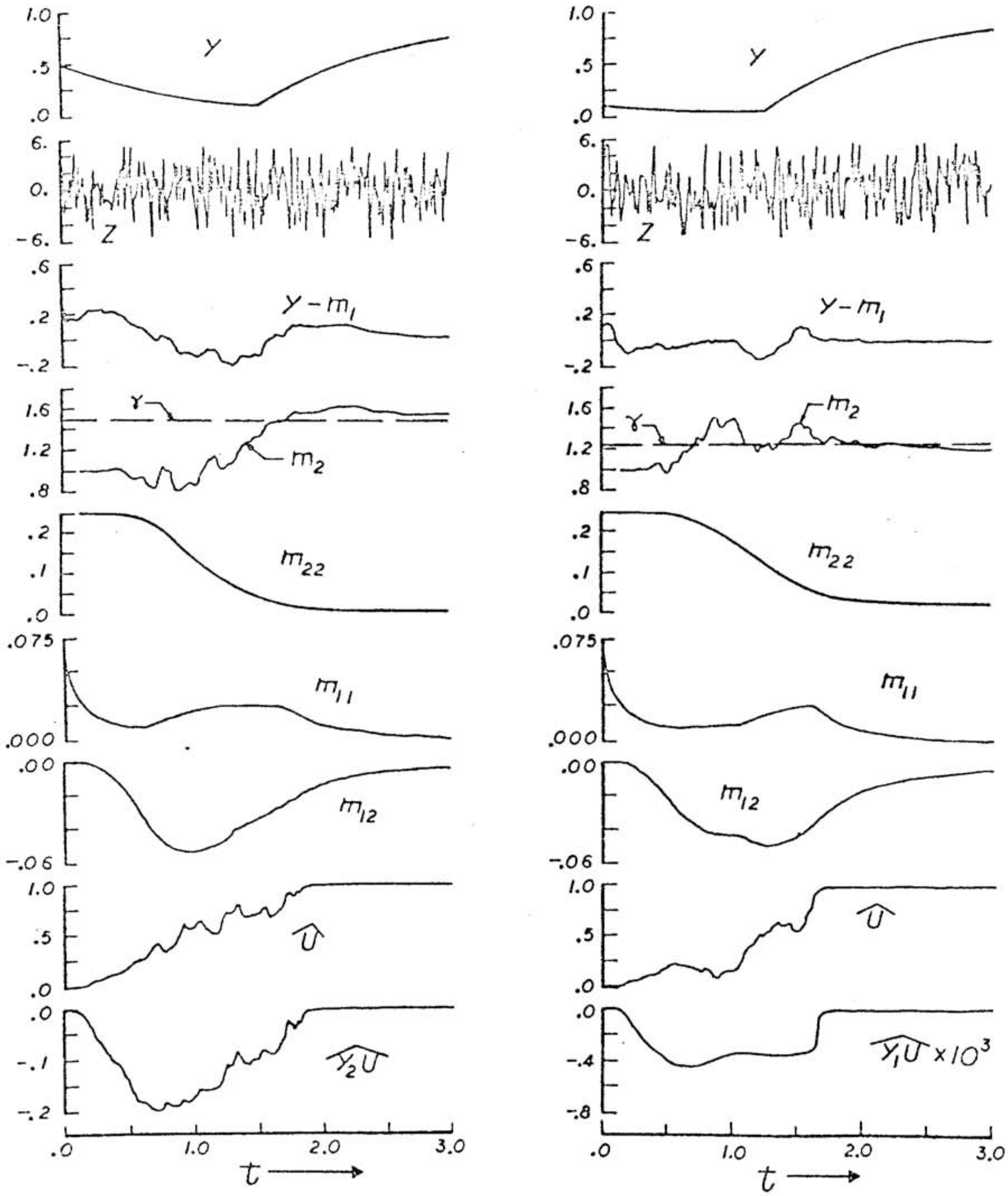


FIGURE 6.4 Two Uniform-Density Time-Delay Estimator Time Responses; R =0.01, m_{10} = 0.0, m_{20} = 1.0, σ_{10} = 0.25, σ_{20} = 0.5

discontinuity exists in the $\widehat{y_1 u(\cdot)}$ variable response curve and is caused by a change in filter modes (a result of the uniform-density approximation for $p(x/Z)$) .

6.4 Concluding Remarks

The dynamics of many industrial processes are such that sizeable time-delays can exist between the input and output variables of those processes. Sophisticated control strategies can be employed to compensate for the effects of these time-delays but such strategies require a knowledge of the amount of dead-time present, an amount which may be randomly varying with time. We've considered in this chapter two possible techniques for obtaining on-line estimates of time-delay parameter values from step-response measurements. Each of these techniques was based upon a particular method of approximately optimal nonlinear filtering but only one technique proved effective.

It was shown that while a Padé approximation of the process dead-time transfer function made it possible to apply the relinearized Kalman filter to the time-delay estimation problem, the resulting filter proved to be totally ineffective as a solution to this problem. A second nonlinear filtering algorithm was investigated and consisted of the differential equations for the conditional mean and variance parameters of $p(x/Z_t)$ under the assumption of a uniform-density for that function. The performance of this filter was discovered to be very acceptable.

While the time-delay estimation problem is only one application of nonlinear filtering, from it, two important general remarks regarding nonlinear filtering can be made:

- The effects of process model simplifications should not be ignored, especially when the process dynamics contain significant amounts of dead-time.
- The filter designer should become familiar with the conditional moment equations for a particular filter problem and should not become totally reliant upon any single method of approximately optimal nonlinear filtering (e.g., the relinearized Kalman filter) as it usually places many hidded restrictions on its application.

CHAPTER VII

SUMMARY AND EXTENSIONS

7.1 Summary

Because the theoretical solution to an optimal nonlinear filtering problem is generally of infinite dimensional form, in practice a finite-dimensional approximate solution to the problem must be found. There exist in the literature many approximate methods of optimal nonlinear filtering, the absolute effectiveness of all of which has been ignored because of the difficulty in computing the optimal filter performance. Those responsible for some of these methods imply but do not show that by increasing the complexity (i.e., dimensionality) of their particular filtering techniques the performance of the filter can be made arbitrarily close to the performance of the theoretically optimal nonlinear filter. One could only speculate as to how much additional filter complexity produced how much additional filter performance.

By way of examples, we have considered in this thesis the question of how much less than optimal is the performance of approximate methods of optimal nonlinear filtering? To evaluate the effectiveness of various approximate filters a numerical technique was developed for computing the optimal performance (a so-called upper performance-bound) for simple nonlinear filtering problems. This technique was based upon a representation theorem of Bucy and was applied to the meticuluous study of five nonlinear filtering problems. Results of these studies showed that essentially optimal performance could be achieved with a fairly simple nonlinear filtering algorithm; the relinearized Kalman filter. Other, more complex methods of nonlinear filtering were investigated and most performed less effectively than the simpler relinearized filter.

The relinearized Kalman filter, though highly effective, is not recommended as a universal solution to the general optimal nonlinear filtering problem. Besides requiring the differentiability of all the system functions (i.e., $f(x)$, $h(x)$, and $g(x)$), the relinearized filter appears to be ineffective when the process disturbance driving functions ($g(x,t)$) are functions of the state, x. This weakness of the relinearized Kalman filter was clearly demonstrated by a rather simple linear disturbance-parameter estimation problem. A satisfactory solution to this problem was achieved by considering the differential equations for the moments of the conditional probability density $p(x/Z_t)$, and making appropriate simplifications to these equations.

It is our recommendation that increased emphasis be given in the literature to the conditional moment equations of nonlinear filtering problems. We have found that a very effective method of nonlinear filtering is to utilize the conditional moment equations under the assumption of a particular form for the density function $p(x/Z_t)$. In fact, this particular method is the only one which can effectively handle filter problems in which some of the system functions are non-differentiable.

Contrary to what the literature might lead you to believe, it was pointed out in this thesis that the approximation problem of optimal nonlinear filtering is not just concerned with the deleterious effects of replacing the infinite-dimensional optimal filter by a finite-dimensional approximation. This particular approximation is only one of many influences on the performance of a nonlinear filter. Improper modelling assumptions, inaccurate estimates of noise levels, and insufficient knowledge regarding the initial state-vector statistics also contribute to the less than optimal performance of a nonlinear filter and should be given at least as much consideration by the filter designer as the mathematical approximations associated with that filter. The question of filter performance sensitivity to modelling errors was considered in this thesis and it appears that in practice its answer can only be determined through Monte Carlo simulation. Identifying those model errors to which a filter is most sensitive is a matter of great importance to a nonlinear filter designer.

The potential applications of optimal nonlinear filtering theory are many. One such application which was given considerable attention in this thesis was that of on-line time-delay parameter estimation for dynamic processes. However, the importance of this particular application was not due so much to the usefulness of the estimator, as it was to the demonstration it gave of the way approximations can affect the performance of nonlinear filters. Two filtering methods are considered, one relying on process model simplifications and the other relying on simplifications to the optimal nonlinear filter solution. The choice between these two filtering schemes was clear in that the former scheme yielded no useful information at all about the time-delay parameter value while the latter scheme performed superbly.

7.2 Extensions

It is doubtful that the theoretical questions concerning optimal nonlinear filtering which remain can be answered by way of some general analytical procedure. Instead, we feel that answers to these questions can only be arrived at through carefully conducted studies of an extensive number of relatively simple nonlinear filter problems. From the results of these studies it is hoped that a valid set of general guidelines for nonlinear filter designers can be established. With this in mind, we recommend that more studies such as those discussed in Chapter IV of this thesis be carried out.

While the primary concern of this thesis has been with the continuous optimal nonlinear filter problem, a parallel effort could be carried out which deals with the discrete optimal filter problem. Upper performance bounds could be computed with a discrete version of the Bucy representation theorem and the performances of the many available discrete filtering methods could be compared with these upper bounds. It is felt, however, that a more important class of nonlinear filtering problems are those which are described by continuous process equations and discrete observation equations (see Section 2.4). The means of establishing an upper performance bound for this particular class of filter problems is not at all clear at this time.

To those whose interests are concerned more with practical questions than with theoretical questions, we address the following comments. A real need exists for carefully thought out applications of nonlinear filtering. All too often, practical applications of a theory have been conceived and carried out by individuals who did not possess a sufficient understanding of the basic theory involved. For this reason, we recommend that a person wishing to apply the theory of nonlinear filtering to real problems first acquire a basic understanding of the theory; the derivation and meaning of the conditional moment equations should be a minimal requirement.

Some potential areas of application for nonlinear filters are the following:

- Chemical reaction kinetic coefficient estimation from input-output measurements.
- Time-delay parameter tracking for distributed parameter systems by signal injection and by naturally occurring disturbance measurements.
- Low frequency, small amplitude frequency response analysis in a real-time environment.
- An optimal phase-lock loop.
- An adaptive feed-forward control parameter estimator.
- A temperature and carbon concentration estimator for the basic oxygen steel making process.

The above problems should be investigated via hybrid simulation with process and observation dynamics simulated on an analog computer and the nonlinear filter implemented in real time on a digital computer, preferably of the process control type. Simulation studies are a necessity for these studies in order that the true values of the variables being estimated be known.

One important practical aspect of nonlinear filtering which needs investigating is that of determining the statistical information about a process which is required for the filtering algorithm. In particular, how can and should the initial state vector statistics and the observation noise parameters be determined by off-line analysis for a particular filter application? Is field measurement data required to estimate these parameters or should the unknown filter parameters be determined by "tuning" the filter on a simulated process or even the process itself?

APPENDIX I

CONDITIONAL CENTRAL MOMENT EQUATIONS

A procedure outlined by Kushner [31] is followed in determining the stochastic differential of the second conditional central moment parameter m_{ij}, where

$$m_{ij} \stackrel{\Delta}{=} \int (x_i - m_i)(x_j - m_j) p(x/Z_t) dx \quad . \qquad (I.1)$$

The same procedure may be utilized to find the differential expressions for any of the higher-order conditional central moments.

Let $\phi \equiv m_{ij}$ and $P \equiv p(X/Z_t)$ and compute for a differential increment in time dt the corresponding differential change in ϕ, where

$$dm_k = \widehat{f_k}\, dt + (\widehat{x_k h} - \widehat{h}\, m_k)^T R^{-1}(d\zeta(t) - \widehat{h}\, dt)$$

$$= \left[\widehat{f_k} + (\widehat{x_k h} - \widehat{h}\, m_k)^T R^{-1}(h - \widehat{h})\right] dt + \qquad (I.2)$$

$$(\widehat{x_k h} - \widehat{h}\, m_k)^T R^{-1} R^{1/2} da(t) \; ; \; k = i, j$$

and

$$dP = \left[\mathcal{L}^*[p] + P(h - \widehat{h})^T R^{-1}(h - \widehat{h}) \right] dt + da^T R^{1/2} R^{-1}(h - \widehat{h})P \quad . \qquad (I.3)$$

Since $a(t)$ is ideal brownian-motion it possesses the two properties

$$E\{da\ dt\} = 0$$

and

$$E\{(da)(da)^T\} = Idt \quad , \qquad (I.4)$$

and as a result the Taylor series differential expansion of ϕ about t must include second-order terms; i.e.,

$$d\phi = \frac{\partial\phi}{\partial m_i}(dm_i) + \frac{\partial\phi}{\partial m_j}(dm_j) + \frac{\partial\phi}{\partial P}(dP) + \frac{1}{2}\frac{\partial^2\phi}{\partial m_i^2}(dm_i)^2 +$$

$$\frac{1}{2}\frac{\partial^2\phi}{\partial m_j^2}(dm_j)^2 + \frac{1}{2}\frac{\partial^2\phi}{\partial P^2}(dP)^2 + \frac{\partial^2\phi}{\partial m_i \partial m_j}(dm_i dm_j) +$$

$$\frac{\partial^2\phi}{\partial m_i \partial P}(dm_i dP) + \frac{\partial^2\phi}{\partial m_j \partial P}(dm_j dP) \quad . \qquad (I.5)$$

From the defining relation (I.1) the partial derivatives in Eq. (I.5) can be computed to be

$$\frac{\partial\phi}{\partial m_i} = -\int(x_j - m_j)P(x)dx \ ; \quad \frac{\partial\phi}{\partial m_j} = -\int(x_i - m_i)P(x)dx \ ;$$

$$\frac{\partial\phi}{\partial P} = \int(x_i - m_i)(x_j - m_j)dx \ ; \quad \frac{\partial^2\phi}{\partial m_i^2} = \frac{\partial^2\phi}{\partial m_j^2} = \frac{\partial^2\phi}{\partial P^2} = 0 \ ;$$

$$\frac{\partial^2\phi}{\partial m_i \partial m_j} = \int P(x)dx \ ; \quad \frac{\partial^2\phi}{\partial m_i \partial P} = -\int(x_j - m_j)dx \ ; \quad \frac{\partial^2\phi}{\partial m_j \partial P} = \int(x_i - m_i)dx$$

which when substituted into Eq. (I.5) yields

$$d\phi = -\int(dm_i)(x_j - m_j)Pdx - \int(dm_j)(x_i - m_i)Pdx +$$

$$\int(dP)(x_i - m_i)(x_j - m_j)dx + \int(dm_i dm_j)Pdx - \int(dm_i dP)(x_j - m_j)dx -$$

$$\int(dm_j dP)(x_i - m_i)dx \qquad (I.6)$$

the first two integrals of which disappear; dm_i and dm_j being independent of x.

From Eq. (I.2) we get

$$dm_i dm_j = (\widehat{x_i h} - \widehat{hm_i})^T R^{-1} R^{1/2}(da)(da)^T R^{1/2} R^{-1}(\widehat{x_j h} - \widehat{hm_j}) +$$

$$\mathcal{O}(dt^2, dadt)$$

which, since $E\{(da)(da)^T\} = Idt$ and $E\{dadt\} = 0$, simplifies in the limit and in the mean to the relation

$$dm_i dm_j = (\widehat{x_i h} - \hat{h} m_i)^T R^{-1}(\widehat{x_j h} - \hat{h} m_j)dt \; . \qquad (I.7)$$

Similarly, it can be shown that

$$dm_i dP = (\widehat{x_i h} - \hat{h} m_i)^T R^{-1}(h - \hat{h})P\, dt \qquad (I.8)$$

and

$$dm_j dP = (\widehat{x_j h} - \hat{h} m_j)^T R^{-1}(h - \hat{h})P\, dt \; . \qquad (I.9)$$

Substitution of (I.3), (I.7), (I.8), and (I.9) into Eq. (1.6) yields

$$d\phi = \int (x_i - m_i)(x_j - m_j)(\mathcal{L}^*[P]dt + P(h - \hat{h})^T R^{-1}(d\zeta - \hat{h}dt))dx +$$

$$(\widehat{x_i h} - \hat{h} m_i)^T R^{-1}(\widehat{x_j h} - \widehat{hm_j})dt - dt\int (x_j - m_j)(\widehat{x_i h} - \widehat{hm_i})^T \cdot$$

$$\cdot R^{-1}(h - \hat{h})Pdx - dt\int (x_i - m_i)(\widehat{x_j h} - \widehat{hm_j})^T R^{-1}(h - \hat{h})Pdx \ . \quad (I.10)$$

Using the $\mathcal{L}$ operators' adjoint property and identifying expectations we get from (I.10)

$$d\phi = \widehat{\mathcal{L}[(x_i - m_i)(x_j - m_j)]}dt + (d\zeta - \hat{h}dt)^T R^{-1}\widehat{(h-\hat{h})(x_i-m_i)(x_j-m_j)} +$$

$$(\widehat{x_i h} - \widehat{hm_i})^T R^{-1}(\widehat{x_j h} - \widehat{hm_j})dt - (\widehat{x_i h} - \widehat{hm_i})^T R^{-1}(\cdot$$

$$\cdot \ \widehat{(x_j - m_j)(h-\hat{h})}\,dt - (\widehat{x_j h} - \widehat{hm_j})^T R^{-1}\,\widehat{(x_i - m_i)(h-\hat{h})}\,dt$$

which simplifies to the final desired relation

$$dm_{ij} \equiv d\phi = \widehat{\mathcal{L}[(x_i-m_i)(x_j-m_j)]}dt - (\widehat{x_i h} - \widehat{hm_i})^T R^{-1}(\widehat{x_j h} - \widehat{hm_j})dt +$$

$$(d\zeta - \hat{h}dt)^T R^{-1}\ \widehat{(h - \hat{h})(x_i - m_i)(x_j - m_j)} \ .$$

APPENDIX II

TIME-DELAY FILTER FUNCTIONS

To completely determine the uniform-density time-delay estimator discussed in Chapter VI it is necessary to evaluate analytically the conditional expectations of the functions $u(t-x_2)$, $y_1u(t-x_2)$, and $y_2u(t-x_2)$ in terms of the conditional moment parameters m_1, m_2, m_{11}, m_{22}, and m_{12}. In particular, we must evaluate the integrals in the defining relations

$$\widehat{u(t-x_2)} = \int u(t-x)\tilde{p}(y)dy$$

$$\widehat{y_1u(t-x_2)} = \int y_1u(t-x_2)\tilde{p}(y)dy$$

$$\widehat{y_2u(t-x_2)} = \int y_2u(t-x_2)\tilde{p}(y)dy \quad .$$

where $\tilde{p}(y) \equiv \tilde{p}(y_1,y_2)$ is a zero mean, zero covariance, uniform density function with variances $\sigma_1^2/3$ and $\sigma_2^2/3$; i.e.,

$$\tilde{p}(y_1,y_2) = \begin{cases} 1/4\sigma_1\sigma_2, & -\sigma_1 \leq y_1 \leq \sigma_1 \text{ and} \\ & -\sigma_2 \leq y_2 \leq \sigma_2 \\ 0 & , \text{ otherwise} \end{cases} \quad .$$

Computing the second moments of $\tilde{p}(y)$ in terms of those of $p(x/Z)$

yields the identities

$$\sigma_1^2 = \frac{3}{2}\{m_{11} + m_{22} - \sqrt{4m_{12}^2 + (m_{22} - m_{11})^2}\}$$

and

$$\sigma_2^2 = \frac{3}{2}\{m_{11} + m_{22} + \sqrt{4m_{12}^2 + (m_{22} - m_{11})^2}\} .$$

Then the conditional expectation of $u(t-x_2)$ can be determined approximately from the following relations:

$$\widehat{u(t-x_2)} = \widehat{u(t + y_1\sin\theta - y_2\cos\theta - m_2)}$$

$$\approx \int_{-\sigma_1}^{\sigma_1}\int_{-\sigma_2}^{\sigma_2} \frac{u(t + y_1\sin\theta - y_2\cos\theta - m_2)}{4\sigma_1\sigma_2} dy_2 dy_1$$

where

$$u(\tau) = \begin{cases} 1 , & \tau \geq 0 \\ 0 , & \tau < 0 \end{cases} ,$$

$$\tau = t + y_1\sin\theta - y_2\cos\theta - m_2 .$$

Referring to the sketch in Figure II.1 , we recognize that the piecewise nature of the function $u(t-x_2)$ requires that a separate set of equations for the terms $\widehat{u(\cdot)}$, $\widehat{y_1u(\cdot)}$, and $\widehat{y_2u(\cdot)}$ be derived for certain sets of values of the parameters y_a and y_c (or equivalently y_b and y_d) . These sets will be

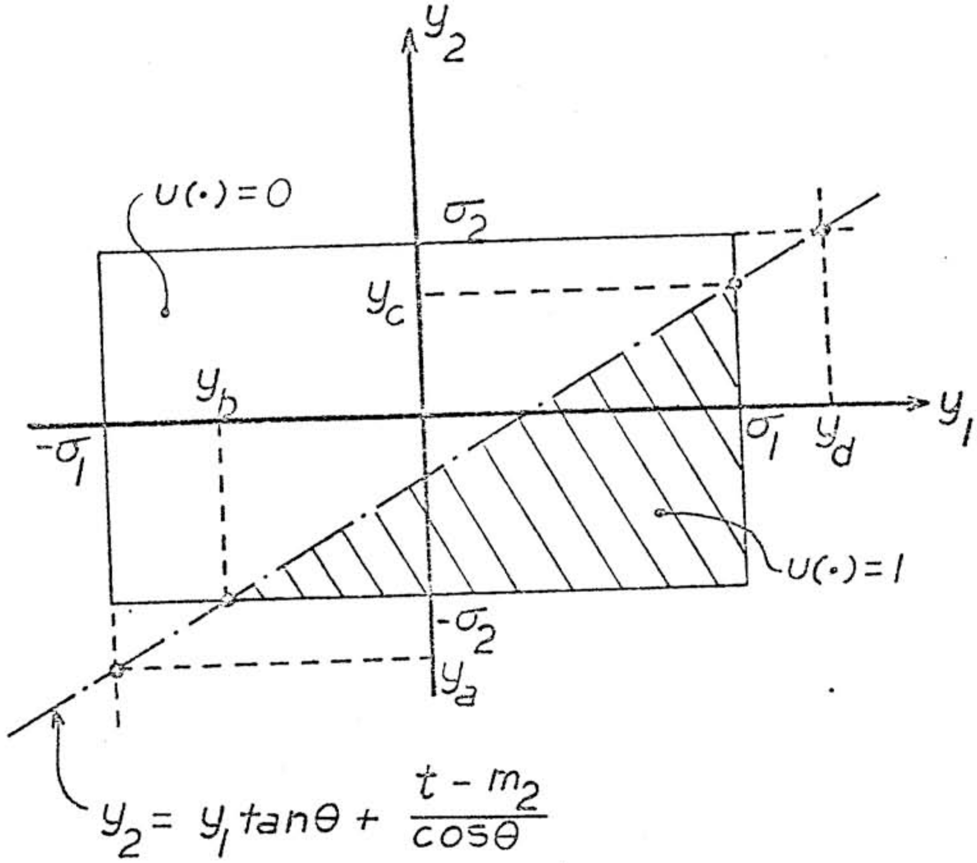


FIGURE II.1 $u(t + y_1 \sin\theta - y_2 \cos\theta - m_2)$ Versus y_1 and y_2

referred to as modes (1 - 9) as defined in Table 6.1 , where

$$y_a = \frac{t - m_2}{\cos\theta} - \sigma_1 \tan\theta$$

and

$$y_c = y_a + 2\sigma_1 \tan\theta \ .$$

In addition, it may be easily shown that

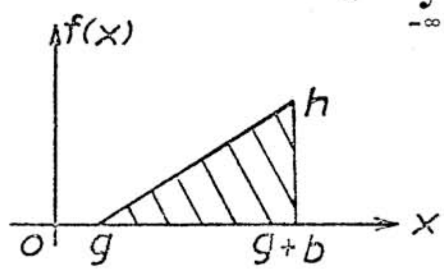
$$y_b = -\sigma_1 - (\sigma_2 + y_a)/\tan\theta$$

and

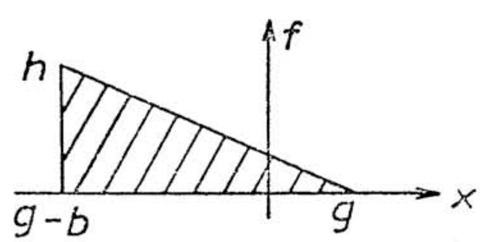
$$y_d = y_b + 2\sigma_2/\tan\theta \ .$$

To determine the analytical expressions for $\widehat{u(\cdot)}$, $\widehat{y_1 u(\cdot)}$, and $\widehat{y_2 u(\cdot)}$ for each of the nine modes, we make use of the following simple area-moment relations:

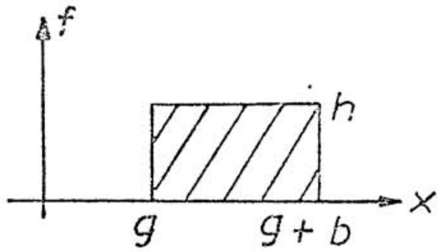
$$\mu_x \stackrel{\Delta}{=} \int_{-\infty}^{\infty} x \ f(x) \ dx$$



$$\mu_x = hb \left(\frac{b}{3} + \frac{g}{2}\right) \qquad \text{(II.1)}$$



$$\mu_x = hb \left(\frac{g}{2} - \frac{b}{3}\right) \qquad \text{(II.2)}$$



$$\mu_x = \frac{hb}{2}(b + 2g) \qquad \text{(II.3)}$$

Consider, for example, mode 2 ; i.e., $y_c \leq -\sigma_2$ and $-\sigma_2 < y_a \leq \sigma_2$. From the area of the triangle in Figure II.2 we get

$$\widehat{u(t - x_2)} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(t-x_2)\tilde{p}(y_1,y_2)dy_1dy_2$$

$$= (\sigma_2 + y_a)(\sigma_1 + y_b)/8\sigma_1\sigma_2 \ ,$$

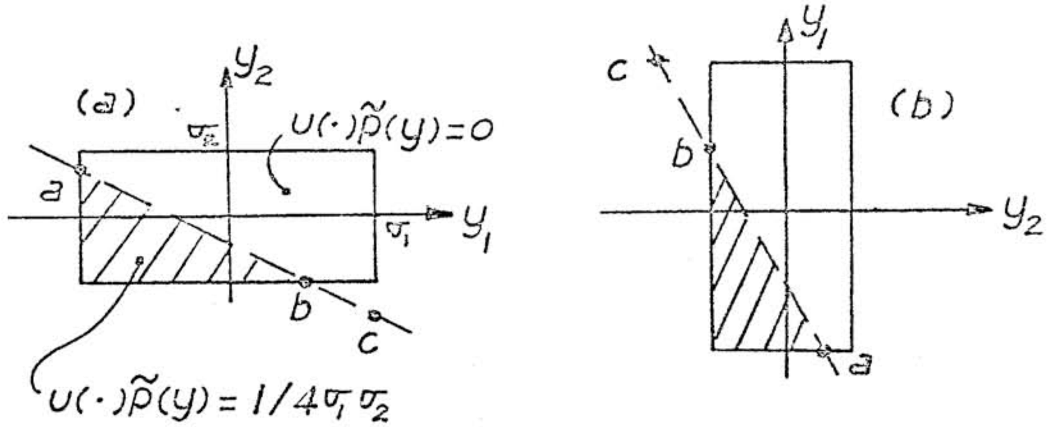


Figure II.2 - Mode 2

and from moment relation (II.2) we find that

$$\widehat{y_1u(t - x_2)} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_1u(t-x_2)\tilde{p}(y_1,y_2)dy_1dy_2$$

$$= (\sigma_2 + y_a)(\sigma_1 + y_b)(y_b - 2\sigma_1)/24\sigma_1\sigma_2 \ .$$

In a similar way it can be shown from moment relation (II.2) and Figure II.2(b) that

$$\widehat{y_2 u(t-x_2)} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_2 u(t-x_2)\tilde{p}(y_1, y_2)dy_1 dy_2$$

$$= (\sigma_1 + y_b)(\sigma_2 + y_a)(y_a - 2\sigma_2)/24\sigma_1\sigma_2 \ .$$

The expressions for $\widehat{u(\cdot)}$, $\widehat{y_1 u(\cdot)}$, and $\widehat{y_2 u(\cdot)}$ were determined in terms of y_a, y_b, y_c, y_d, σ_1, and σ_2 for each mode and are presented in Tables 6.2 - 6.4 of this thesis.

REFERENCES

[1] Athans, M., Wishner, B. P., and Bertolini, A., "Suboptimal State Estimation for Continuous-Time Nonlinear Systems from Discrete Noisy Measurements", JACC Proceedings, June 1968, pp. 364-382.

[2] Bass, R. W., Norum, V. D., and Schwartz, L., "Optimal Multichannel Nonlinear Filtering", Journal of Mathematical Analysis and Applications, Vol. 16, No. 1, October 1966, pp. 152-164.

[3] Bellman, R., Kagiwada, H. H., Kalaba, R., and Sridhar, R., "Invariant Imbedding and Nonlinear Filtering Theory", RM-4374-PR, Rand Corp., Santa Monica, Calif., Dec. 1964.

[4] Bucy, R. S., "Nonlinear Filtering Theory", IEEE Trans. on Auto. Control, Vol. 10, April 1965, pg. 198.

[5] Bucy, R. S., "Recent Results in Linear and Nonlinear Filtering", Symposium: Stochastic Problems in Control, pp. 87-105, June 1968 JACC.

[6] Bucy, R. S., and Joseph, P. D., "Filtering for Stochastic Processes with Applications to Guidance", Interscience Publishers, N. Y. 1968.

[7] Carney, T. M., and Goldwyn, R. M., "Numerical Experiments with Various Optimal Estimators", Journal of Optimization Theory and Applications, Vol. 1, No. 2, Sept. 1967, pp. 113-130.

[8] Cramer, H., "Mathematical Methods of Statistics", Princeton University Press, N. J. 1946.

[9] Dashevskii, M. L., and Lipster, R. S., "Application of Conditional Semi-Invariants in Problems of Nonlinear Filtering of Markov Processes", Automation and Remote Control, No. 6, June 1967, pp. 912-921.

[10] Dashevskii, M. L., and Lipster, R. S., "Approximate Analysis of Nonstationary Nonlinear Dynamic Systems", Automation and Remote Control", 1967, pp. 1110-1120.

[11] Dashevskii, M. L., "Approximate Analysis of the Accuracy of Nonstationary, Nonlinear Systems, Using the Method of Semi-Invariants", Automation and Remote Control, No. 11, Nov. 1967, pp. 1673-1690.

[12] Dashevskii, M. L., "Method of Semi-Invariants in the Nonlinear Filtering of Markov Processes", Automation and Remote Control, No. 7, July 1968, pp. 1035-1042.

[13] Dashevskii, M. L., "Equations for the Semi-Invariants of a Nonlinear Dynamic System", Automation and Remote Control, No. 10, Oct. 1968, pp. 1605-1613.

[14] Detchmendy, D. M., and Sridhar, R., "Sequential Estimation of States and Parameters in Noisy Nonlinear Dynamic Systems", JACC, June 1965.

[15] Doob, J. L, "Stochastic Processes", John Wiley & Sons, N. Y. 1953.

[16] Fisher, J. R., "Conditional Probability Density Functions and Optimal Nonlinear Estimation", Ph.D. Thesis, University of California at Los Angeles, 1966.

[17] Fisher, J. R., and Stear, E. B., "Optimal Nonlinear Filtering For Independent Increment Processes - Part I, II", IEEE Transactions on Information Theory, Vol. 3, No. 4, Oct. 1967, pp. 558-578.

[18] Fisher, J. R., "Optimal Nonlinear Filtering", Advances in Control Systems, Vol. 5, 1967, Academic Press, N. Y., pp. 197-300.

[19] Gnedenko, B. V., "The Theory of Probability", Chelsea Pub. Co., N. Y. 1962.

[20] Gray, A. H., and Caughey, T. K., "A controversy in Problems Involving Random Parametric Excitation", Journal of Mathematics and Physics, Vol. 44, No. 3, Sept. 1965, pp. 288-296.

[21] Ho, Y. C., "The Method of Least Squares and Optimal Filtering Theory", RM-3329-PR, The Rand Corp., Santa Monica, Calif., Oct. 1962.

[22] Ito, K., "On a Stochastic Differential Equation", Memoirs American Mathematical Society, No. 4, 1951, pp. 51-89.

[23] Kailath, T., and Frost, P., "Mathematical Modeling of Stochastic Processes", Symposium: Stochastic Problems in Control, pp. 1-38, June 1968, JACC.

[24] Kalman, R. E., "A New Approach to Linear Filtering and Prediction Problems", Trans. ASME, Ser. D, Vol. 82, No. 1, March 1960, pp. 35-45.

[25] Kalman, R. E., and Bucy, R. S., "New Results in Linear Filtering and Prediction Theory", Trans. ASME, Ser. D, Journal Basic Eng., Vol. 83, No. 1, March 1961, pp. 95-107.

[26] Kalman, R. E., "New Methods and Results in Linear Prediction and Filtering Theory", Tech. Rept. 61-1, RIAS, 1961.

[27] Kushner, H. J., "On the Differential Equations Satisfied by Conditional Probability Densities of Markov Processes with Applications", Journal SIAM Control, Vol. 2, No. 1, 1964, pp. 106-119.

[28] Kushner, H. J., "On the Dynamical Equations of Conditional Probability Density Functions, With Applications to Optimal Stochastic Control Theory", Journal of Math. Analysis and Applications, Vol. 8, No. 2, April 1964, pp. 332-344.

[29] Kushner, H. J., "Dynamical Equations for Optimal Nonlinear Filtering", Journal of Differential Equations, Vol. 3, No. 2, April 1967, pp. 179-190.

[30] Kushner, H. J., "Nonlinear Filtering: The Exact Dynamical Equations Satisfied by the Conditional Mode", IEEE Trans./Auto. Cont., Vol. 12, No. 3, June 1967, pp. 262-267.

[31] Kushner, H. J., "Approximations to Optimal Nonlinear Filters", IEEE Trans./Auto. Cont., Vol. 12, No. 5, Oct. 1967, pp. 546-556.

[32] Kroy, W. H., and Stubberud, A. R., "Identification via Nonlinear Filtering", International Journal Control, Vol. 6, No. 6, 1967, pp. 499-522.

[33] Lee, E. S., "Quasilinearization and Invariant Imbedding, With Applications to Chemical Engineering and Adaptive Control", Academic Press, N. Y. 1968.

[34] Lo, J. T., "Finite Dimensional Sensor Orbits and Optimal Nonlinear Filtering", Ph.D. Thesis, University of Southern Calif., August 1969.

[35] Mortenson, R. E., "Optimal Control of Continuous-Time Stochastic Systems", Ph.D. Thesis, University of California at Berkeley, No Date.

[36] Schwartz, L., and Stear, E. B., "A Valid Mathematical Model for Approximate Nonlinear Minimal-Variance Filtering", Journal of Mathematical Analysis and Applications, Vol. 21, No. 1, January 1968, pp. 1-6.

[37] Schwartz, L., "Approximate Continuous Nonlinear Minimal-Variance Filtering", Ph.D. Thesis, University of California at Los Angeles, 1966.

[38] Schwartz, L., and Stear, E. B., "A Computational Comparison of Several Nonlinear Filters", IEEE Trans./Auto. Cont., Feb. 1968, pp. 83-86.

[39] Sorenson, H. W., and Stubberud, A. R., "Nonlinear Filtering by Approximation of the A'Posteriori Density", International Journal of Control, Vol. 8, No. 1, pp. 33-51.

[40] Stratonovich, R. L., "On the Theory of Optimal Non-Linear Filtration of Random Functions", Theory of Probability and Its Applications, Vol. 4, 1959, pp. 223-225.

[41] Stratonovich, R. L., "Conditional Markov Processes", Theory of Probability and Its Applications, Vol. 5, No. 2, 1960, pp. 156-178.

[42] Stratonovich, R. L., "A New Representation for Stochastic Integrals and Equations", Journal SIAM Control, Vol. 4, No. 2, May 1966, pp. 362-371.

[43] Stratonovich, R. L., "Conditional Markov Processes and Their Application to the Theory of Optimal Control", American Elsevier Publishing Co., N. Y. 1968.

[44] Sunahara, Y., "An Approximation Method of State Estimation for Nonlinear Dynamical Systems", Tech. Rept. 67-8, Center for Dynamical Systems, Brown University, Dec. 1967.

[45] Tribus, M., and Kornblum, R. J., "The Use of Bayesian Inference in the Design of an Endpoint Control System for the Basic Oxygen Steel Furnace", Unpublished Paper, Thayer School of Engineering, Dartmouth College, 1968.

[46] Wong, E., and Zakai, M., "On the Relation Between Ordinary and Stochastic Differential Equations", International Journal Engineering Science, Vol. 3, No. 2, pp. 213-229.

[47] Wonham, W. M., "Some Applications of Stochastic Differential Equations to Optimal Nonlinear Filtering", Journal SIAM Control, Vol. 2, No. 3, 1965, pp. 347-369.