# APPROXIMATE SOLUTION FOR NONLINEAR FILTERING AND IDENTIFICATION PROBLEMS

by

Saleh M. Radaideh

A thesis submitted to the
School of Graduate Studies and
Research in partial fulfillment of the
requirements for the degree of
Doctor of philosophy

Ottawa-Carleton Institute for Electrical Engineering

Department of Electrical Engineering
Faculty of Engineering
University of Ottawa
May 1995

ISBN 0-612-15663-X

UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

I here by declare that I am the sole author of this thesis.
I authorize the University of Ottawa to lend this thesis to other institutions or individuals for the purpose of scholarly research.

**SALEH RADAIDEH**

I further authorise the University of Ottawa to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

**SALEH RADAIDEH**

# ACKNOWLEDGEMENT

The author expresses his deepest gratitude to his advisor Professor N.U. Ahmed for his enormous patience, generous encouragement, invaluable guidance and numerous hours of disscussions through this work, without which this thesis would have not been possible.

The author wishes to express his appreciation to the faculty and staff of the Department of Electrical Engineering, University of Ottawa, for their kindness and support.

Financial assistance provided by Jordan University of Science and Technology during the period of this research, is gratefully ackonwledged.

Finally, the author is highly indepted to his wife Halimeh Qawasmeh for her enormous patience, constant support and encouragment. I am specially grateful to my loving parents and my daughter Hebah and my son Osamah.

# Abstract

Filtering and identification problems of partially observable stochastic dynamical systems has been considered. A modification of the extended Kalman filter (EKF) called a modified extended Kalman filter (MEKF) and a theoretical justification for this modification have been investigated.

A simple but powerful numerical method for the approximation of the unnormalized conditional (probability) density of filtered diffusion process which satisfies Zakai equation ari..s from diffusion processes observed in correlated ( or uncorrelated) noises and solves the nonlinear filtering problem has been presented. Using Galerkin technique the solution of Zakai equation has been approximated by means of a sequence of nonstandard basis functions given by a parameterized family of Gaussian densities. The spatial domain for the solution of Zakai equation and the completeness of the Gaussian densities have been also investigated. The methods are illustrated by some numerical examples.

Techniques of optimal control theory as well as linear filter theory have been utilized in identifying the parameters of linear (partially observable) stochastic differential systems. Using the method of simulated annealing a computational algorithm for identifying the unknown parameters from the available observation has been derived. The results are illustrated by some examples.

# List of Symbols

| | |
|---|---|
| $\emptyset$ | the emptyset |
| $\equiv$ | identically equal to |
| $\simeq$ | approximately equal to |
| $R$ | the set of real numbers |
| $R^+$ | the set of non-negative real numbers |
| $(a,b)$ | the open interval $a < x < b$ in $R$ |
| $[a,b]$ | the closed interval $a \leq x \leq b$ in $R$ |
| $R^n$ | the $n$-dimensional Euclidean space |
| $(\cdot)'$ | the transpose of $(\cdot)$ |
| $(\cdot)^*$ | the formal adjoint of $(\cdot)$ |
| $A = [a_{ij}]$ | a matrix $A$ with $ij$th component $a_{ij}$ |
| $f = [f_i]$ | a vector $f$ with $i$th component $f_i$ |
| $C(R^n)$ | the space of continuous functions on $R^n$ |
| $C^k(R^n)$ | the space of $k$ times continuously differentiable functions on $R^n$ |
| $C^{1,k}(R^n)$ | the space of all real-valued functions which are $C^1$ in $t$ and $C^k$ in $x$ on $R^n$. |
| $C_b^k(R^n)$ | the subspace of functions in $C^k(R^n)$ which are bounded along with their derivatives of all orders up to and including $k$. |
| $O(r^p)$ | expression divided by $r^p$ remains bounded as $r \rightarrow 0$ |
| $o(r^p)$ | expression divided by $r^p$ converages to zero as $r \rightarrow 0$ |
| $L_2(R^n)$ | the space of all functions such that $\int_{R^n} \mid f(t) \mid^2 dt < \infty$. |
| $L_\infty([0,T], H)$ | the space of all $H$-valued functions which are essentially bounded on $[0,T]$. |
| $H^1(R^n)$ | the standard Sobolev space: $H^1(R^n) \equiv \{\varphi \in L_2(R^n) : D_{x_i}\varphi \in L_2(R^n)$, $i = 1,2,..,n\}$, where $D_{x_i}\varphi$ denotes the disributional derivative of $\varphi$ with respect to $x_i$ |
| $H^{-1}$ | the dual space of $H^1$. |
| $\mathcal{L}(H^1, H^{-1})$ | the bounded linear operator from $H^1$ to $H^{-1}$ |
| $< \cdot,\cdot >_{H^1,H^{-1}}$ | the duality product between $H^1$ and $H^{-1}$ |
| $(\Omega, \mathcal{B}, P)$ | the probability space |
| $\mathcal{F}_t^y$ | the $\sigma$- algebra generated by the process $y$ up to time $t$ |
| $i.i.d$ | independent identically distributed |
| $\mid x \mid$ | the norm of $x \in R^n$: $\mid x \mid^2 = \sum_{i=1}^n x_i^2$ |
| $trA$ | the trace of the matrix $A$: $trA = \sum_{i=1}^n a_{ii}$ |
| $I$ | the identity matrix |
| $x \cdot y$ | the scalar product of $x, y \in R^n$, $x \cdot y = \sum_{i=1}^n x_i y_i$ |
| $\nabla f$ | the gradient of the scalar function $f$ |
| $\Delta f$ | the Lablacian of $f$ |
| $LQR$ | Linear Quadratic Regulator |
| $diag(d)$ | a diagonal matrix; that is, the only nonzero entries are on the diagonal an |
| $D(B)$ | the domain of the operator $B$ |
| $\Omega$ | is the sample space of the elementary events. |
| $\mathcal{F}$ | is a class of subsets of the set $\Omega$. |
| $P$ | is a set function mapping $\mathcal{F}$ to $[0,1]$. |

# Contents

# List of Tables

# List of Figures

# Chapter 1

# INTRODUCTION

## 1.1 Introduction

Physical systems are designed and built to perform certain functions. For example, submarine, aircraft and spacecraft must navigate in their respective environment to accomplish their objectives: whereas an electric power system network must meet the load demand. In order to determine whether a system is performing properly, and ultimately one can control the system performance, the engineer must know the state of the system. In an electric power system, for example, the state may be taken as a number of generators in operation or the voltages and phase angles at the network nodes.

In most cases, physical systems are subjected to random disturbances, so that the state of the system itself may be random. In order to determine the state of the system, the engineer should build a measuring device and take measurements (or observations) on his system. These measurements are generally contaminated with noise caused by the electronic and mechanical components of the measuring device. The problem of determining the state of the system from these noisy measurements is called the estimation problem. There are three types of estimation, which can be stated as follows. When the time at which an estimate is desired falls within the span of available measurements data, the problem is termed smoothing. If the time of interest occurs after the last measurements, the problem is referred to as prediction. If the time of interest coincides with the last available measurement point, the problem is called filtering, which is the main subject of this thesis. The problem of filtering random signals from noisy measurements arises in many engineering areas such as radar, communications, aerospace, and control. Therefore, the filtering problem is to estimate the conditional probability of the state given the observed data. In fact the filtering problem is of central importance; since the estimated states are required in monitoring, and for the control of the systems. Furthermore, large class of system identification can also be regarded as filtering problems [42].

## 1.2 A Brief Review

### 1.2.1 Filtering Problem

Over the last two decades, much attention has been focused on the study of filtering problem of linear as well as nonlinear dynamical systems [15,19,27,31,46,48,51,56,67, 74,91]. In 1960, Kalman and Bucy [51] used least square method to derive a system of linear stochastic and ordinary differential equations describing the evolution of the estimated state and the corresponding covariance matrix, respectively. In fact these filter equations were extensively used in practice for many engineering problems( see for example [48]). In 1960, Stratonovich [81] was also pioneering the development of the probabilistic approach to nonlinear filtering problem. In 1967, Kushner [56-58] obtained his equation, which is a nonlinear stochastic partial differential equation describing the flow of normalized conditional density of diffusion Markov processes, with the help of which the whole filtering problem is theoretically resolved. A more theoretically convenient filter equation was obtained by Zakai [91] in 1969. Zakai equation is a linear stochastic partial differential equation describing the flow of unnormalized conditional density of a diffusion Markov process.

After this pioneering work of Kushner and Zakai, the nonlinear filtering problem for diffusion processes was treated by many authors. In references [14,52,58,59], the problem of existence of solutions of Zakai and Kushner equations was considered. In [50], Kallianpur and Karandikar used Radon Nikodym derivative result obtained by Balakrishnan [14] and finitely additive white noise approach to derive a linear partial differential equation for the unnormalized conditional density in which the observation process (in finitely additive set up) occurs as a parameters. The extension of Kushner and Zakai results for the case where the state as well as the observed process are discontinuous, was also treated in [28,63,74,86]. The filter equations of Kushner and Zakai have also been applied to stochastic control problems of partially observable diffusion process by many authors such as Fleming [33-35], Kushner [54,58] where the existence of optimal control was proved given the observed path.

### 1.2.2 Identification Problem

Over the last several years, considerable attention has been focused on an identification problem of stochastic systems governed by linear or nonlinear Ito equations [8,26,41,63,83]. In [83], Identification problem for partially observed linear time-invariant systems was considered. Using linear filter theory, maximum likelihood approach, and the smoothness of solutions of an algebraic Riccati equation, sufficient conditions were obtained for the consistency of the likelihood estimate.

In [63], Liptser and Shiryayev considered the identification problem for a class of completely observed systems governed by a stochastic differential equation of the form

$$dX(t) = h(t, X(t))\alpha dt + dW(t), \quad t \geq 0, \tag{1.1}$$

where $X$ is a real-valued stochastic process and $\alpha$ is some unknown parameter. Using the maximum likelihood approach, the authors [63] obtained an explicit expression for the maximum likelihood estimate $\hat{\alpha}$. An extension of this result to a multi-parameter problem $\alpha \in R^m$ for stochastic systems in $R^n$ was considered by

Ahmed [8]. In [41], Gland considered the identification problem for a more general class of systems governed by stochastic differential equations of the form

$$dy(t) = h(\alpha, X(t))dt + dW(t), \quad t \geq 0. \tag{1.2}$$

where $\alpha$ is an unknown parameter and $X(t)$ is a diffusion process. Utilizing the maximum likelihood approach along with forward and backward Zakai equations, a numerical scheme was developed for computing the parameter $\alpha$ given the output history $\{y(s),\ s \leq t\}$.

In [26], Dabbous and Ahmed considered the problem of identification of drift and dispersion parameters for a general class of partially observed systems governed by the following system of Ito equations

$$\begin{aligned} dX(t) &= a(t, X(t), \alpha)dt + b(t, X(t), \alpha)dW(t), \quad t \in [0, T] \quad X(0) = x(0), \\ dy(t) &= h(X(t), \alpha)dt + \sigma_0(t, y(t))dW_0(t), \quad t \in [0, T], \quad y(0) = 0. \end{aligned} \tag{1.3}$$

Using the pathwise description of Zakai equation, they formulated the original identification problem as a deterministic control problem in which the unnormalized conditional density (solution of Zakai equation) is treated as a state, the unknown parameters as control and the likelihood ratio as an objective functional.

In [23], Bagchi considered a situation with an unknown observation covariance noise in which case the likelihood functional cannot be apparently defined, Bagchi proposed a functional analogous to the likelihood functional by giving an a priori guess of the observation covariance noise. However from the numerical point of view, an a priori guess should be close to the true value.

Newton's method is the usual procedure for computing the maximum likelihood estimates(MLE) [43] which involves recursive calculation of the gradient vector and Hessian matrix of the (MLE) at a fixed value of the parameter vector. The drawback of this method is that the convergence to the desired optimum fails whenever the Hessian matrix has some negative eigenvalues or is nearly singular.

In [4,63,83], identification of drift parameters for completely observed systems were considered. In [26], which considers partially observed identification problem, the authors used the Zakai equation as the basic state equation which, of course, is a partial differential equation. For n-dimensional problems, $n \geq 2$, the associated computational problem becomes nontrivial. It appears that for partially observed nonlinear problems there is no escape from PDE.

## 1.3 Formulation of Filtering Problem, Assumptions and Notations

We consider the standard filtering problem for a class of stochastic systems where the state is governed by Ito differential equations of the form

$$\begin{aligned} dX(t) &= f(X(t))dt + b(X(t))dW(t), \\ X(0) &= x(0), \end{aligned} \tag{1.4}$$

where $X$, $f$ are vectors in $R^n$; $b$ is $(n \times n)$ matrix and $W$ is $n$-dimensional standard Wiener process independent of the initial state $x(0)$. The output (observed) process $\{y(t),\ t \geq 0\}$, is assume to be the related to the state process $\{X(t),\ t \geq 0\}$, through the following stochastic differential equation

$$\begin{aligned} dy(t) &= h(X(t))dt + \sigma_0(t)dV(t), \\ y(0) &= 0, \end{aligned} \tag{1.5}$$

where $y$, $h$ are vectors in $R^m$, $\sigma_0$ is an $(m \times m)$ matrix and $V$ is an $m$-dimensional standard Wiener process independent of $W$ and the initial state $x(0)$.

**Remark 1.1.** Equation (1.5) is the mathematical way of expressing that some measurement [39]

$$z(t) = h(X(t)) + \sigma_0(t)\zeta(t) \tag{1.6}$$

is available at time $t$, where $\{\zeta(t), t \geq 0\}$ is a Gaussian white-noise process independent of $\{X(t), t \geq 0\}$. The measurements are coming continuously, it is worth to have a filter which can be revised step by step in order to take into account the new measurement. This can be done by solving the following Zakai or Kushner equation. Let $g$ be any bounded measurable function on $R^n$ with values in $R^k$. Let $\mathcal{F}_t^y$ denotes the $\sigma$- algebra generated by the process $y$ up to time $t$. The ultimate goal in the filtering is to estimate $g(X(t))$ given the history of $y$ up to time $t$, that is $\mathcal{F}_t^y$. Assuming that all random processes and vectors described above are defined on a complete probability space $(\Omega, \mathcal{B}_0, P)$, we state the filtering problem as follows.

Given any bounded measurable $R^k$-valued function $g$ on $R^n$, we find $\tilde{g}(t)$, which is $\mathcal{F}_t^y$-measurable, such that

$$E\{| g(X(t)) - \tilde{g}(t) |^2 / \mathcal{F}_t^y\} \tag{1.7}$$

is minimum. Defining

$$\hat{g}(t) \equiv E\{g(X(t))/\mathcal{F}_t^y\}, \tag{1.8}$$

one can rewrite (1.7) as

$$E\{| g(X(t)) - \tilde{g}(t) |^2/\mathcal{F}_t^y\} = E\{| g(X(t)) - \hat{g}(t) |^2 \; + \; 2(\hat{g}(t) - \tilde{g}(t)) \cdot (g(X(t)) - \hat{g}(t))$$
$$+ \; | \hat{g} - \tilde{g}(t) |^2/\mathcal{F}_t^y\}. \tag{1.9}$$

Using the definition (1.8) and the fact that the functions $\tilde{g}(t)$ and $\hat{g}(t)$ are $\mathcal{F}_t^y$-measurable, it follows from equation (1.9) that

$$E\{| g(X(t)) - \tilde{g}(t) |^2/\mathcal{F}_t^y\} = E\{| g(X(t)) - \hat{g}(t) |^2/\mathcal{F}_t^y\} + | \hat{g} - \tilde{g}(t) |^2. \tag{1.10}$$

Clearly, it follows from equation (1.10) that the quantity, given by (1.7), attains its minimum for

$$\tilde{g}(t) = \hat{g}(t) \equiv E\{g(X(t))/\mathcal{F}_t^y\}. \tag{1.11}$$

This shows that the conditional expectation gives the best (optimal) estimate in the mean square sense. Since

$$\hat{g}(t) \;\; = \;\; \int_{R^n} g(x) Pr\{X(t) \in dx/\mathcal{F}_t^y\}$$

$$\equiv \int_{R^n} g(x)P(t,dx/\mathcal{F}_t^y)$$
$$= \int_{R^n} g(x)p(t,x/\mathcal{F}_t^y)dx. \quad (1.12)$$

where $P(t,dx/\mathcal{F}_t^y)$ is a measure-valued $\mathcal{F}_t^y$- measurable random process, the whole filtering problem is resolved if the measure $P(t,dx/\mathcal{F}_t^y)$ (or the density $p(t,x/\mathcal{F}_t^y)$) can be computed (i.e., $\hat{g}(t)$ can be expressed as an output of a finite dimensional system of stochastic differential equations driven by the observed process $\{y(s), 0 \leq s \leq t\}$). In fact, Kushner [56] obtained a nonlinear stochastic partial differential equation for the conditional density $p(t,x/\mathcal{F}_t^y)$. This equation is given by

$$\begin{aligned} dp(t,x/\mathcal{F}_t^y) &= A^*p(t,x/\mathcal{F}_t^y)dt + (h(x) - \hat{h}(t)).\Gamma_0^{-1}[dy(t) - \hat{h}(t)dt]p(t,x/\mathcal{F}_t^y), \\ p(0,x) &= p_0(x), \end{aligned} \quad (1.13)$$

for all $t \in [0,T]$, where $\hat{h}(t) \equiv E\{h(X(t))/\mathcal{F}_t^y\}$, $\Gamma_0 \equiv \sigma_0\sigma_0'$ and $A^*$ is the formal adjoint of the operator $A$ given by

$$A\varphi = (f \cdot \varphi_x) + \frac{1}{2}tr(bb'\varphi_{xx}) \; , \varphi \in D(A), \quad (1.14)$$

where $\varphi_{xx}$ is the second derivative of $\varphi$ with respect to $x$. Theoretically, a more convenient equation was obtained by Zakai [91] which is a linear stochastic partial differential equation describing the flow of the unnormalized conditional density, denoted by $\Psi(t,\cdot)$. This equation is given by

$$d\Psi(t,x) = A^*\Psi(t,x)dt + \Psi(t,x)h(x).\Gamma_0^{-1}dy(t), \quad t \geq 0,$$

$$\Psi(0,x) \;=\; p_0(x). \tag{1.15}$$

The conditional density $p(t,x/\mathcal{F}_t^y)$ is related to the unnormalized conditional density $\Psi(t,\cdot), x \in R^n$, through the relation

$$p(t,x/\mathcal{F}_t^y) = \frac{\Psi(t,x)}{\int_{R^n} \Psi(t,x)dx}. \tag{1.16}$$

Therefore, solving equation (1.15) and using the relation (1.16), one obtains the conditional density $\{p(t,\cdot/\mathcal{F}_t^y),\ t \geq 0\}$, and hence $\hat{g}(t)$ can be computed using equation (1.12). It is clear from equation (1.13) and equation (1.15) that the basic data describing the model given by equation (1.4) and equation (1.5) are: the initial density $p_0(\cdot)$, the drift $f(\cdot)$, the diffusion matrix $bb'(\cdot)$, the sensor $h(\cdot)$, and the matrix $\Gamma_0 = \sigma_0\sigma_0'$. Note that if $h = 0$, equation (1.13) reduces, as expected, to the Fokker-Plank equation governing the unconditional density of $\{X(t), t \geq 0\}$. Equation (1.15) looks much simpler than equation (1.13) but still both of them are infinite dimensional, and also the normalized and the unnormalized conditional density depend on the a priori information of $\{X(t), t \geq 0\}$ provided by the infinitesimal generator $A$, and on the observation sample-path $\{y(t), t \geq 0\}$. Equation (1.13) yields, upon integration with respect to $x$, equations for the moments of the conditional density, in particular, for the optimal estimate $\hat{g}(t)$:

$$\begin{aligned} d\hat{g}(t) &= (\widehat{Ag})(t)dt + \left((\widehat{hg})(t) - \hat{h}(t)\hat{g}(t)\right)'\Gamma_0^{-1}[dy(t) - \hat{h}(t)dt], \\ \hat{g}(0) &= Eg(x(0)). \end{aligned} \tag{1.17}$$

Notice that the stochastic differential equation (1.17) is in general both infinite family of moments equations (i.e., the first moment requiring the information about the second, the second requiring the information about the third and so on), and nonrecursive (because of the occurrence of the expectations $\widehat{Ag}$, $\widehat{hg}$, $\hat{h}$, and $\hat{g}$).

Clearly, this means that the filtering problem of systems (1.4-1.5) is completely resolved provided that one can compute the normalized (or the unnormalized) conditional density of the process $\{X(t), t \geq 0\}$ given $\{y(t), t \geq 0\}$. However, in general, obtaining analytical solutions of Zakai or Kushner equations for nonlinear systems, is not an easy task and so numerical approximations are required.

Throughout this thesis, we will assume that the stochastic differential equations (1.4-1.5) has a unique strong solution . For this, we need the following assumptions and notations.

**Assumptions**

(A1) The functions $f(x)$ and $b(x)$ are continuous on $R^n$ and there exists a constant $\gamma > 0$ such that

$$\alpha = bb' \geq \gamma I$$

where $I$ denotes the identity matrix.

(A2) There exists a constant $k > 0$ such that

$$T \mid f(x) - f(y) \mid^2 + \parallel b(x) - b(y) \parallel^2 \leq k \mid x - y \mid^2,$$

for any $x, y \in R^n$.

(A3) The functions $f_i$, $\frac{\partial f_i}{\partial x_i}$, $a_{ij}$, $\frac{\partial a_{ij}}{\partial x_i}$, and $\frac{\partial^2 a_{ij}}{\partial x_i \partial x_j}$; $i, j = 1, 2, ..., n$ are bounded and satisfy Holder condition on $R^n$.

(A4) The function $h$ is continuous on $R^n$ and there exists a constant $k > 0$ such that

$$| h(x) - h(y) |^2 \leq k | x - y |^2, \quad x, y \in R^n,$$

and

$$E \int_0^t | h(X(s)) |^2 ds < \infty, \quad P - a.s \ (almost\ surely),$$

along any solution $X(t)$; $t \geq 0$, of equation (1.4).

(A5) The matrix function $\sigma_0$ is continuous, bounded and satisfies uniform Lipchitz condition on $R^m$ (see assumption (A2)).

**Notations:**

Let $\{\eta(t),\ t \geq 0\}$ be any random process and let $\sigma\{\eta(s),\ s \leq t\}$ denote the $\sigma$-field generated by the process $\eta$ up to time $t$. Define $\mathcal{F}_t^y \equiv \sigma\{y(s),\ s \leq t\}$, $\mathcal{F}_t^x \equiv \sigma\{x(s),\ s \leq t\}$, $\mathcal{F}_t^W \equiv \sigma\{W(s),\ s \leq t\}$, $\mathcal{F}_t^V \equiv \sigma\{V(s),\ s \leq t\}$, $\sigma(X(t)) \equiv \sigma\{X(t)\}$ and $\mathcal{F} \equiv \mathcal{F}_T^W \vee \mathcal{F}_T^V \vee \sigma(x_0) \subset \mathcal{B}_0$. Let $\mathcal{F}_t$, $t \geq 0$, be an increasing family of sub $\sigma$-fields contained in $\mathcal{F}$ such that for each $t \in [0, T]$, the processes $X(t)$ and $y(t)$ are $\mathcal{F}_t$-measurable. Let $\Omega$ denote the space of continuous functions on $[0, T]$ with values in $R^{n+m}$, and let $\mathcal{A}$ denote the Borel $\sigma$- algebra on $\Omega$. We call $(\Omega, \mathcal{A})$ the canonical sample space. We denote $\mathcal{B}(R^n)$ the Borel $\sigma$- field generated by the sub set of $R^n$.

**Remark 1.2**: The stochastic differentials which are interpreted in terms of stochastic integrals, do not transform according to the chain rule of classical calculus. Roughly speaking, the difference is due to the fact that the stochastic differential $(dW(t)dW'(t))$ is equal to $Idt$ in the mean square sense. The overall explanation of the stochastic differentials is in the following (Lemma 1.1,1.2).

**Lemma 1.1** (Ito differential, [42] corollary 1,p.70)

Let the $R^n$-valued process $\{X(t), t \geq 0\}$ be the solution of the following differential equation

$$dX(t) = f(X(t))dt + b(X(t))dW(t), \tag{1.18}$$

with initial value $X(0) = x(0)$, where $W$ is $n$-dimensional standard Wiener process independent of the initial state $x(0)$ and the functions $f$ and $b$ satisfy our basic assumptions. Let $g$ be any twice continuously differentiable function on $R^n$. Then the Ito differential of $g$ is given by

$$dg(X(t)) = (Ag)(X(t))dt + (b'g_x)(X(t)) \cdot dW(t), \tag{1.19}$$

where $g_x$ denotes the partial derivative of $g$ with respect to $x$ and $A$ denotes the (backward) Kolmogorov operator and can be determined with the help of the following Lemma.

**Lemma 1.2** (Infinitesimal Generator)

Consider the system (1.18) and let $g$ be any twice continuously differentiable function on $R^n$. Then

$$\lim_{\Delta t \to 0} E_x \frac{1}{\Delta t} \{g(X(t + \Delta t) - g(X(t))\} = \mathbf{A}g, \tag{1.20}$$

where

$$\mathbf{A}g \equiv g_x \cdot f(x) + \frac{1}{2} tr(bb' g_{xx}). \tag{1.21}$$

Here $E_x$ denotes the conditional expectation given $X(t) = x$.

**proof:** Using Taylor's series expansion, the function $g(X(t + \Delta t))$, $t \geq 0$, can be written as follows

$$\begin{aligned} g(X(t + \Delta t)) = g(X(t)) \quad &+ \quad (g_x \cdot \Delta X(t)) \\ &+ \quad \frac{1}{2}(g_{xx} \Delta X(t) \cdot \Delta X(t)) + o(| \Delta X(t) |^2), \end{aligned} \tag{1.22}$$

where $g_x$ and $g_{xx}$ denotes the first and second derivatives of $g$ with respect to $x$ and

$$\begin{aligned} \Delta X(t) &\equiv X(t + \Delta t) - X(t) \\ &\cong f(X(t))\Delta t + b(X(t))(W(t + \Delta t) - W(t)). \end{aligned} \tag{1.23}$$

Thus

$$\begin{aligned} E_x(g_x \cdot \Delta X(t)) &= (g_x \cdot f(x))\Delta t + b' g_x \cdot E_x(W(t + \Delta t) - W(t)) \\ &= (g_x \cdot f(x))\Delta t \end{aligned} \tag{1.24}$$

and

$$\begin{aligned} \frac{1}{2}E_x\{g_{xx}\Delta X(t)\cdot\Delta x(t)\} &= \frac{1}{2}E_x b' g_{xx} b\Big(W(t+\Delta t)-W(t))\cdot(W(t+\Delta t)-W(t)\Big) \\ &= \frac{1}{2}tr(bb'g_{xx})\Delta t. \end{aligned} \tag{1.25}$$

Utilizing equations (1.22), (1.24), and (1.25) in (1.20) and the facts that $o(|\Delta x|^2) = o(\Delta t)$ and $\frac{o(\Delta t)}{\Delta t} \to 0$, *as* $\Delta t \to 0$, the result of the lemma follows.

## 1.4 Outline of the Thesis

The thesis is organized as follows: Chapter 1 contains the motivation for filtering and identification problems and a brief review of the previous studies in these areas.

In Chapter 2, we present a various time discrete numerical methods which are appropriate for the simulation of stochastic differential equations sample paths on digital computers.

In Chapter 3, we consider the solution of the filtering problem in the case of mild nonlinearities and small noise intensities.

In Chapter 4, we present a finite dimensional approximation of Zakai equation arising from diffusion processes observed in uncorrelated noises.

In Chapter 5, we present a finite dimensional approximation of Zakai equation arising from diffusion processes observed in correlated noises.

In Chapter 6, we consider the identification problem for a system of partially observed linear stochastic differential equations.

## 1.5 Original Contributions in This Thesis

(1) We present a modification of the extended Kalman filter which we call modified extended Kalman filter (MEKF) and a theoretical justification for this modification; Section 3.4.

(2) We present a simple but powerful numerical method for the approximation of the unnormalized conditional (probability) density of filtered diffusion process which satisfies Zakai equation arising from diffusion processes observed in correlated ( or uncorrelated) noises; Chapter 4 and Chapter 5.

(3) We formulate the identification problem for a system of partially observed linear stochastic differential equations as deterministic control problem. We present a result whereby one can determine all the system parameters including the covariance matrices of the noise processes; Chapter 6.

# Chapter 2

# NUMERICAL METHODS FOR STOCHASTIC DIFFERENTIAL EQUATIONS AND ZAKAI EQUATION

## 2.1 Introduction

In the following Chapters of this thesis we have proposed some iterative numerical schemes to compute the optimal estimate. For preparation, we will discuss briefly in this Chapter various numerical approaches that solve Ito stochastic differential equations and the nonlinear filtering problem. In the literature there exists now several different approaches that have been suggested for the numerical solution of stochastic differential equations (SDEs). Maruyama [72] was the first who proposed a scheme to approximate the solutions of (SDEs); it is a natural stochastic analogue of the Euler scheme for deterministic differential equations, and it converges in the mean- square sense. Another important contribution belongs to Milstein

[73] who improved the order of convergence by adding a second order term in the approximation of the stochastic Ito integral. Among more recent contributions one should mention those of Wagner and Platen [66] and Platen [66] who have derived a formula of the Taylor type expanding the solution of a stochastic differential equations about the points of a time partition. In a series of papers , Pardoux and Talay [76] performed a systematic study of basic problems in stochastic numerical integration. Kushner [60] proposed the discretization of both time and space variables, so the approximating processes are then finite Markov chains. These can be handled on digital computers through their transition matrices. The most efficient and widely applicable approach to solve SDEs seems to be the simulation of sample paths of time discrete approximations on digital computers. This is based on a finite discretization of the time interval under consideration and generates approximate values of the sample paths step by step at the discretization times. The simulated sample paths can then be analyzed by usual statistical methods to determine how good the approximation is and in what sense it is close to the exact solution.

In contrast to the deterministic differential equation case, where different numerical methods converge (if they are convergent) to the same solution, in the case of stochastic differential equations different schemes can converge to different solutions for the same noise sample and initial condition.

In the case of stochastic Ito differential equations one can work with deterministic Fokker Plank equations for the probability density. While such partial (deterministic) differential equations can be, in principle, integrated numerically with standard procedures, in practice there occur great difficulties. So, direct numerical methods for stochastic equations seem to be important and promising tools in characterization of the behavior of complex systems governed by stochastic differential equations.

## 2.2 Numerical Approaches to the Nonlinear Equations

The most commonly model used in the study of the state of the stochastic systems is the following stochastic differential equation (or Ito equation):

$$\begin{aligned} dX(t) &= f(t, X(t))dt + b(t, X(t))dW(t), \\ X(0) &= x(0), \end{aligned} \tag{2.1}$$

where $X$, and $f$ are vectors in $R^n$; $b$ is $(n \times p)$ matrix and $W$ is $p$-dimensional standard Wiener process independent of the initial state $x(0)$. One can write the integral form of equation (2.1) as

$$X(t) = X(0) + \int_0^t f(s, X(s))ds + \int_0^t b(s, X(s))dW(s) \tag{2.2}$$

for $t \geq 0$. It consists of an initial value $X(0) = x(0)$, which may be random, a slowly varying continuous component called the drift and rapidly varying continuous component called diffusion. The second integral in equation (2.2) is an Ito stochastic integral with respect to the Wiener process $W = \{W(t), t \geq 0\}$. Assume that $x(0)$ is independent of $W(t)$, and the drift $f$ and diffusion $b$ satisfy all conditions guaranteeing the existence and uniqueness of a solution. Since in practice the observation process is usually of discrete type, it is convenient to write the stochastic differential equation (2.1) in a discrete form. The most widely applicable numerical approximations to the solutions of the (SDEs) are the time discrete approximation or difference methods, in which the continuous time stochastic differential equation (2.2) is replaced by a discrete-time difference equation generating values $Z(1)$, $Z(2)$,...,$Z(N)$ to approximate $X(t_1)$, $X(t_2)$,....,$X(t_N)$ at given discretization times $0 < t_1 < t_2 < ..... < t_N = T$. Once the initial value $Z(0)$ has been specified, usually $Z(0) = x(0)$, the approximation values can be calculated recursively. For simplicity we will consider $\{0 = t_0 < t_1 < ... < t_N = T\}$ as a uniform partition of the interval $[0, T]$ with increment $\delta = \frac{T}{N}$. Let $\Delta W(t_i) \equiv W(t_{i+1}) - W(t_i)$ be the increments of the Wiener process $W = \{W(t), t \geq 0\}$, for $i = 0, 1, ..., N-1$. It is known that these increments are i.i.d Gaussian random variables with means and variances

$$E(\Delta W(t_i)) = 0, \quad E(\Delta W(t_i))^2 = \delta,$$

respectively, for $i = 0, 1, ..., N-1$. For numerical purposes, we can generate such random variables from independent, uniformly distributed random variables on

[0, 1]. In problems such as those that we shall consider in the next Chapters involving direct simulations, filtering of Ito processes, it is important that the trajectories (the sample paths) of the approximation be close to those of Ito process. This suggests that a criterion involving some form of strong convergence should be used. Mathematically it is advantageous to consider the absolute error at the final time instant $T$. We shall say that an approximation process $Z$ converges in the strong sense with order $\gamma > 0$ if there exists a finite constant $K$ and a positive constant $\delta_0$ such that

$$E(\mid X(T) - Z(N) \mid) \leq K\delta^\gamma$$

for any time discretization with maximum step size $\delta \in (0, \delta_0)$. This is a criterion for the closeness of the sample paths of the Ito process $X$ and the approximation $Z$ at time $T$.

Here, we shall survey various time discrete numerical methods which are appropriate for the simulation of sample paths on digital computers. In the following schemes, we will use $i$ to denote $t_i$, and $g = g(t_i, Z(i))$ for each function defined on $R^+ \times R^n$ and $i = 0, 1, ...N - 1$ for the simplicity of presentation.

### 2.2.1 The Euler Scheme

The simplest numerical scheme for the Ito equation (2.2) is the stochastic analogue of the Euler method. The Euler scheme converges with strong order $\gamma = 0.5$ under Lipschitz and bounded growth conditions on the coefficients $f$ and $b$.

(i) In the 1-dimensional case, $d = p = 1$, the Euler scheme has the form

$$Z(i+1) = Z(i) + f\delta + b\Delta W(i),$$
$$Z(0) = x(0), \; i = 0, 1, ..., N-1. \tag{2.3}$$

(ii) In the multi-dimensional case with scalar noise, $n = 1, 2, ..$ and $p = 1$, the $k$th component of the Euler scheme is given by

$$Z_k(i+1) = Z_k(i) + f_k\delta + b_k\Delta W(i),$$
$$Z_k(0) = x_k(0), \; i = 0, 1, ..., N-1, \tag{2.4}$$

for $k = 1, 2, ..., n$.

(iii) For multi-dimensional case with $n = p = 1, 2, ..$ the $k$th component of the Euler scheme has the form

$$Z_k(i+1) = Z_k(i) + f_k\delta + \sum_{j=1}^{p} b_{kj}\Delta W_j(i),$$
$$Z_k(0) = x_k(0), \; i = 0, 1, ..., N-1, \tag{2.5}$$

for $k = 1, 2, ..., n$. Here,

$$\Delta W_j(i) = W_j(i+1) - W_j(i) \tag{2.6}$$

is the increment of the $j$th component of the $p$-dimensional standard Wiener process $W$, and $\Delta W_{j_1}(i)$ and $\Delta W_{j_2}(i)$ are independent for $j_1 \neq j_2$, $i = 0, 1, ..., N-1$. Note that the different components of the time discrete approximation $Z$ are coupled through the drift and diffusion coefficients, as in the corresponding stochastic differential equation (2.1). In special cases the Euler scheme may actually achieve a higher order of strong convergence. For example, when the noise is additive, that is when the diffusion coefficient has the form

$$b(t,x) \equiv b(t) \tag{2.7}$$

for all $(t,x) \in R^+ \times R^n$ and under smoothness assumptions on $f$ and $b$ it turns out that the Euler scheme has an order of strong convergence $\gamma = 1.0$. Usually the Euler scheme gives good numerical results when the drift and diffusion coefficients are nearly constant. In general, it is not particularly satisfactory and the use of higher order schemes is recommended.

### 2.2.2 The Milstein Scheme

The Milstein scheme [73] has the order of strong convergence $\gamma = 1.0$ under the assumption that $f \in C^{1,1}(R^+ \times R^n)$ and $b \in C^{1,2}(R^+ \times R^n)$. To simplify the notations in the following schemes, we shall use the following operators,

$$L^0 = \frac{\partial}{\partial t} + \sum_{i=1}^{n} f_i \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,l=1}^{n} \sum_{j=1}^{p} b_{ij} b_{lj} \frac{\partial^2}{\partial x_i \partial x_l}, \tag{2.8}$$

$$L^j = \sum_{i=1}^{n} b_{ij} \frac{\partial}{\partial x_i}, \tag{2.9}$$

for $j = 1, 2, ..p$.

(i) In the 1-dimensional case with $n = p = 1$, the Milstein scheme has the form

$$Z(i+1) = Z(i) + f\delta + b\Delta W(i) + \frac{1}{2} b \frac{\partial b}{\partial x} \{(\Delta W(i))^2 - \delta\},$$

$$Z(0) = x(0),\ i = 0, 1, \ldots, N-1. \qquad (2.10)$$

Thus, with the addition of just one term to the Euler scheme to form the Milstein scheme we increase the strong convergence order from $\gamma = 0.5$ to $\gamma = 1.0$.

(ii) In the multi-dimensional case with $p = 1$, and $n = 1, 2, ..$ the $k$th component of the Milstien scheme is given by

$$\begin{aligned} Z_k(i+1) &= Z_k(i) + f_k\delta + b_k\Delta W(i) + \frac{1}{2}\left(\sum_{j=1}^{n} b_j \frac{\partial b_k}{\partial x_j}\right)\{(\Delta W(i))^2 - \delta\} \\ Z_k(0) &= x_k(0),\ i = 0, 1, \ldots, N-1, \end{aligned} \qquad (2.11)$$

for $k = 1, 2, \ldots, n$.

(iii) In the general multi-dimensional case with $n, p = 1, 2, ..$ the $k$th component of the Milstein scheme has the form

$$\begin{aligned} Z_k(i+1) &= Z_k(i) + f_k\delta + \sum_{j=1}^{p} b_{kj}\Delta W_j(i) + \sum_{j_1, j_2=1}^{p} L^{j_1} b_{kj_2} I(j_1, j_2), \\ Z_k(0) &= x_k(0),\ i = 0, 1, \ldots, N-1, \end{aligned} \qquad (2.12)$$

for $k = 1, 2, \ldots, n$, where $I(j_1, j_2)$ is the multiple stochastic integrals and given by

$$I(j_1, j_2) = \int_{t_i}^{t_{i+1}} \int_{t_i}^{s_1} dW_{j_1}(s_2) dW_{j_2}(s_1). \qquad (2.13)$$

We remark that for $j_1 \neq j_2$ the multiple stochastic integrals appearing in the scheme (2.12) can not be so easily expressed in terms of the increments $\Delta W_{j_1}$ and $\Delta W_{j_2}$ as in the case $j_1 = j_2$.

$$I(j_1, j_1) = \tfrac{1}{2}\{(\Delta W_{j_1}(i))^2 - \delta\}.$$

In many practical problems the diffusion coefficients have special properties which allow the Milstein scheme to be simplified in a way that avoids the use of double stochastic integrals involving different components of the Wiener process. For instance, with additive noise (2.7) the diffusion coefficients depend at most on time $t$ and not on $x$ variable and the Milstein scheme reduces to the Euler scheme, which involves no double stochastic integrals.

Another important special case is that of diagonal noise, where $n = p$ and each component $X_k$ of the Ito process $X$ is disturbed only by the corresponding component $W_k$ of the Wiener process $W$ and the diagonal diffusion coefficient $b_{kk}$ depends only on $x_k$, that is

$$b_{kj}(t,x) \equiv 0 \quad and \quad \frac{\partial b_{jj}}{\partial x_k}(t,x) \equiv 0 \tag{2.14}$$

for each $(t,x) \in R^+ \times R^n$ and $j,k = 1,2,..p$ with $j \neq k$. Thus for diagonal noise the components of the Ito process are coupled only through the drift term. It is easy to see that the Milstein scheme for diagonal noise reduces to

$$\begin{aligned} Z_k(i+1) &= Z_k(i) + f_k\delta + b_{kk}\Delta W_k(i) + \frac{1}{2}b_{kk}\frac{\partial b_{kk}}{\partial x_k}\{(\Delta W_k(i))^2 - \delta\}, \\ Z_k(0) &= x_k(0),\ i = 0,1,\dots N-1, \end{aligned} \tag{2.15}$$

for $k = 1,2,\dots n$. It should be noticed that the Euler and Milstein schemes described above produce values of the approximation only at the discretization times. If values are required at intermediate instants, then either piecewise constant values from the proceeding discretization point or some interpolation, especially a linear interpolation, of the values of the two immediate enclosing discretization points could be used.

### 2.2.3 Platen Scheme

Platen obtained his scheme [66] by including further multiple stochastic integrals. These multiple stochastic integrals contain additional information about the sample path of the Wiener process.

(i) In the 1-dimensional case with $n = p = 1$, Platen scheme has the following form

$$\begin{aligned} Z(i+1) = Z(i) &+ f\delta + b\Delta W(i) + \frac{1}{2}b\frac{\partial b}{\partial x}\{(\Delta W(i))^2 - \delta\} \\ &+ \frac{\partial f}{\partial x}b\Delta\tilde{W}(i) + \frac{1}{2}(f\frac{\partial f}{\partial x} + \frac{1}{2}b^2\frac{\partial^2 f}{\partial x^2})\delta^2 \\ &+ (f\frac{\partial b}{\partial x} + \frac{1}{2}b^2\frac{\partial^2 b}{\partial x^2})\{\Delta W(i)\delta - \tilde{W}(i)\} \\ &+ \frac{1}{2}b(b\frac{\partial^2 b}{\partial x^2} + (\frac{\partial b}{\partial x})^2)\{\frac{1}{3}(\Delta W(i))^2 - \delta\}\Delta W(i). \end{aligned} \quad (2.16)$$

for $i = 0, 1, ..., N-1$, and initial value $Z(0) = x(0)$. Here the additional random variable $\Delta\tilde{W}(i)$ is required to represent the double stochastic integral

$$\Delta\tilde{W}(i) = \int_{t_i}^{t_i+1}\int_{t_i}^{s_2} dW(s_1)ds_2. \quad (2.17)$$

It is clear that $\Delta\tilde{W}(i)$ is normally distributed with mean $E(\Delta\tilde{W}(i)) = 0$, variance $E((\Delta\tilde{W}(i))^2) = \frac{1}{3}\delta^3$ and covariance $(\Delta\tilde{W}(i)\Delta W(i)) = \frac{1}{2}\delta^2$, for all $i = 1, 2, ..N-1$. In the simulation, the pair of correlated normally distributed random variables $(\Delta\tilde{W}(i), \Delta W(i))$ can be generated from two independent $N(0,1)$ distributed random variables $U_1(i)$ and $U_2(i)$ by means of the transformation

$$\Delta W(i) = U_1(i)\sqrt{\delta}, \quad \Delta\tilde{W}(i) = \tfrac{1}{2}\sqrt{\delta^3}(U_1(i) + \tfrac{1}{\sqrt{3}}U_2(i)),$$

for $i = 0, 1, ..., N-1$.

(ii) In the multi-dimensional case with $n = 1, 2, ..$ and $p = 1$, that is with scalar noise, the $k$th component of the Platen scheme is given by

$$\begin{aligned} Z_k(i+1) = Z_k(i) &+ f_k\delta + b_k\Delta W(i) \\ &+ \frac{1}{2}L^1 b_k\{(\Delta W(i))^2 - \delta\} + L^1 f_k \Delta\tilde{W}(i) \\ &+ L^0 b_k\{\Delta W(i)\delta - \Delta\tilde{W}(i)\} + \frac{1}{2}L^0 f_k \delta^2 \\ &+ \frac{1}{2}L^1 L^1 b_k\{\frac{1}{3}(\Delta W(i))^2 - \delta\}\Delta W(i), \end{aligned} \tag{2.18}$$

for $i = 0, 1, .., N-1$, $Z(0) = x_k(0)$, and $k = 1, 2, ..., n$.

(iii) In the case of additive noise (2.7) with multi- dimensional $n, p = 1, 2, ..$ the $k$th component of the Platen scheme has the following form

$$\begin{aligned} Z_k(i+1) &= Z_k(i) + f_k\delta + \sum_{j=1}^{p} b_{kj}\Delta W_j(i) + \frac{1}{2}L^0 f_k \delta^2 \\ &+ \sum_{j=1}^{p}[L^j f_k \Delta\tilde{W}_j(i) + \frac{\partial b_{kj}}{\partial t}\{\Delta W_j(i)\delta - \Delta\tilde{W}_j(i)\}], \end{aligned} \tag{2.19}$$

for $i = 0, 1, .., N-1$, $Z_k(0) = x_k(0)$, and $k = 1, 2, .., n$, where

$$\Delta\tilde{W}_j(i) = \int_{t_i}^{t_{i+1}} \int_{t_i}^{s_2} dW_j(s_1) ds_2 \; for \; j = 1, ..p. \tag{2.20}$$

Platen scheme converges with a strong order $\gamma = 1.5$ when the coefficients $f$ and $b$ are sufficiently smooth and satisfy Lipschitz and bounded growth conditions, and includes (in addition to $\Delta W(i)$) the $\Delta \tilde{W}(i)$. We have used this scheme to simulate the sample paths of the stochastic systems given in Chapters 3 and 4 of this thesis.

## 2.3 Numerical Approach to the Linear Equations

As a special case, let us consider the following linear stochastic differential equation

$$\begin{aligned} dX(t) &= A(t)X(t)dt + B(t)dW(t), \\ X(0) &= x(0), \end{aligned} \tag{2.21}$$

where $A$ is a $n \times n$ matrix, $B$ is $n \times p$, and $W$ is $p$-dimensional Wiener process and $x(0)$ independent of the process $W$. Clearly, the solution of equation (2.21) is given by

$$X(t) = \phi(t,0)x(0) + \int_0^t \phi(t,\theta)B(\theta)dW(\theta), \tag{2.22}$$

where $\phi(t,\theta)$, $0 \leq \theta \leq t$, denote the transition matrix corresponding to the equation (2.21), and satisfies the equations

$$\tfrac{d}{dt}\phi(t,\theta) = A(t)\phi(t,\theta), \quad \phi(t,t) = I.$$

Then, it is clear from equation (2.22) that

$$X(t_{i+1}) = \phi(t_{i+1},t_i)X(t_i) + \int_{t_i}^{t_{i+1}} \phi(t_{i+1},\theta)B(\theta)dW(\theta). \tag{2.23}$$

In the following we will use $i$ to denote $t_i$, $i = 0, 1, 2, ..$ for the simplicity of presentation. For sufficiently small $\delta$, equation (2.23) can be approximately written as

$$X(i+1) = X(i) + \phi(i+1,i)X(i) + \phi(i+1,i)B(i)\Delta W(i) \tag{2.24}$$

We have used this scheme to simulate the sample paths of the stochastic systems given in Chapters 5 and, 6 of this thesis.

## 2.4 A Brief Survey of Numerical Approximation of Zakai Equation

Computing an approximate solution of the Zakai equation is one possible way of solving a practical nonlinear filtering problem, at least when the dimension of the diffusion process is small. Since, in general, it is impossible to obtain analytical solution of the Zakai equation for nonlinear systems, it is reasonable to seek a numerical schemes that are implementable in straight forward manner and provide accurate solutions. In the following we will describe some of the various algorithms that approximate the solution of Zakai equation which arises from the standard nonlinear filtering problem given by equation (1.4) and equation (1.5).

### 2.4.1 Time Discretization of Zakai Equation

In this kind of discretization two steps are considered: sampling of the observation, and discretization of Zakai equation.

**a-Sampling of the observation sample-path**

This is the process by which the information contained in the whole sample-path

$\{y(s), 0 \leq s \leq t\}$ is replaced by some simpler information contained in a finite collection of random variables. The sampling time step, denoted by $\delta$, is a parameter associated to the sampling of the observation. There are different sampling strategies depending on the finite collection of random variables considered. If one only uses the increments of the observation process on time intervals of length $\delta$, one gets the following simple sampling

$$\Delta y(i) \equiv y(i+1) - y(i) = \int_{t_i}^{t_{i+1}} z(s)ds, \tag{2.25}$$

which is the mean value of the actual measurements (equation (1.6)) on the time interval $t_i \leq s \leq t_{i+1}$. If one considers in addition quantities like [39]

$$\Delta y^1(i) = \frac{1}{\delta}\int_{t_i}^{t_{i+1}} (s - t_i)dy(s) \tag{2.26}$$

$$\Delta y^2(i) = \frac{1}{\delta}\int_{t_i}^{t_{i+1}} (t_{i+1} - s)dy(s) \tag{2.27}$$

which are two other different ways of computing some mean value of the actual measurements (equation (1.6)) on the time interval $t_i \leq s \leq t_{i+1}$. Note that $\Delta y(i) = \Delta y^1(i) + \Delta y^2(i)$. The simple scheme (2.25) has been used in all simulation introduced in this thesis.

### b-Discretization of Zakai equation

For notational convenience, from now on , we shall use $\Psi(t)$ to denote $\Psi(t, x)$ and $h$ to denote $h(x)$, $x \in R^n$. Following Milstein scheme given in the previous section, a time- discretized approximation of Zakai equation (1.15) (when the observation process (equation (1.5)) takes values in $R^1$) is given by [32,39,38,40,41,74,78] as follows

$$\Psi(i+1) = \Psi(i) + A^*\Psi(i+1)\delta + \frac{1}{\sigma_0^2}h\Psi(i)\Delta y(i) + \frac{1}{2\sigma_0^2}h^2\Psi(i)[(\Delta y(i))^2 - \delta] \quad (2.28)$$

or even better

$$(I - \delta A^*)\Psi(i+1) = \Psi(i)exp[\frac{1}{\sigma_0^2}h\Delta y(i) - \frac{h^2}{2\sigma_0^2}\delta]. \quad (2.29)$$

Then, since $\Psi(0,x) \geq 0$, $\Psi(t_i,x) \geq 0$, for all $i$ and $x$. In the scheme defined by equation (2.28) (or equation (2.29)) the transition from $\Psi(t_i)$ to $\Psi(t_{i+1})$ reflects the following situation: A new measurement $\Delta y(i)$ is available, which is interpreted as a noisy nonlinear observation of $X(i+1)$, and combined with the current estimate $\Psi(i)$ of $X(i)$ to produce an estimate $\Psi(i+1)$ of $X(i+1)$.

The next step of course is to discretize the space, in order to get an algorithm that can be used on a computer.

### 2.4.2 Space Discretization of Zakai Equation

The first step here is to restrict the whole state space to a fixed bounded subregion (unless the state space was already bounded itself). In the one-dimensional case, this is just an interval, whose characteristics (location and width) are just considered as parameters. In higher dimension, one will have to describe more explicitly the shape of this subregion (more details given in Chapter 4). The boundary conditions can be of either Dirichlet type (stopped) or Numann type (reflected), where both refer to the behavior of the diffusion process with respect to the boundary.

Then one has to design a grid , whose mesh size denoted by $\beta$. In [16,56,76], an implicit full discretization scheme that provides time and space discretization of the Zakai equation has been applied. The approximation takes the form

$$(I - \delta A_\beta)V^{i+1} = D^i V^i \tag{2.30}$$

for each time $t = i\delta$, where $V^i$ represent the vector of mesh points at time $i\delta$, the matrix $A_\beta$ is the approximate of $A^*$, and $D^i$ is a data dependent diagonal matrix. In [78], Picard has developed the following discretization of the Zakai equation

$$\frac{\partial \Psi}{\partial t}(t) \quad = \quad A^*\Psi(t) \quad for \ (i-1)\delta < t < i\delta \tag{2.31}$$

$$\Psi(i\delta) \quad = \quad \Psi(i\delta-)exp\left(\frac{1}{\sigma_0^2}h\Delta y(i) - \delta\frac{1}{2\sigma_0^2}h^2\right) \tag{2.32}$$

for $\Psi(0) = p_0$, $i = 1,2...N$, where $\Psi(i\delta-)$ is the solution of equation (2.31) immediately before $t = i\delta$. This type of approximations (2.31-2.32) still involves a sequence of partial differential equations (Fokker Plank equations). In practice, it is not easy to solve this kind of partial differential equations.

One can notice that these approximations are not easy to implement on a computer especially when the dimension of the diffusion process is greater than 1. In this thesis, we propose a simple but powerful technique for the approximation of the unnormalized conditional (probability) density of filtered diffusion process which satisfies Zakai equation.

# Chapter 3

# MODIFIED EXTENDED KALMAN FILTERING

## 3.1 Introduction

In this chapter, within the frame work of the model given in equations (1.4) and (1.5), we seek for algorithms calculating the minimum variance estimate of the state vector as a function of time and the accumulated measurements data. It is known from Chapter 1 that the minimum variance estimate is always the conditional mean of the state vector, regardless of its probability density function. Following the results obtained by Kushner in [56], one can see that the equation for the conditional moments, in particular, the conditional mean of the state vector and its error covariance matrix depend upon the entire conditional probability measure. To obtain practical estimation algorithms , methods of computing the conditional moments which do not depend upon knowing the conditional probability measure are needed. A method often used to achieve this goal is the Extended Kalman Filter

(EKF). In this Chapter we present a modification of the Extended Kalman Filter (EKF) which we call Modified Extended Kalman Filter (MEKF). In Section 3.3, we present Kalman and Extended Kalman Filter. In Section 3.4, we formulate the Modified Extended Kalman Filter (MEKF). In Section 3.5 , numerical examples are presented using both (EKF) and (MEKF).

## 3.2 The Filtering Problem

In this chapter we consider the filtering problem for a class of systems of mild nonlinearities and small noise intensities governed by Ito differential equations of the form

$$dX(t) = f(t, X(t))dt + \varepsilon_1 b(t)dW(t), \quad X(0) = x(0), \quad t \geq 0 \tag{3.1}$$

$$dy(t) = h(t, X(t))dt + \varepsilon_2\sigma_0(t)dV(t), \quad y(0) = 0, \quad t \geq 0, \tag{3.2}$$

where the state $X \in R^n$ and the observation $y \in R^m$ and the processes, $W$, $V$ are two mutually independent Brownian motions with dimensions $p$ and $q$ respectively, and $\varepsilon_1$ (dynamic noise intensity), and $\varepsilon_2$ (measurement noise intensity), $b$ is $n \times p$ matrix, and $\sigma_0$ is $m \times q$ matrix, $h : [0, \infty) \times R^n \rightarrow R^m$ and $f : [0, \infty) \times R^n \rightarrow R^n$ and $x(0)$ is a $\mathcal{F}_0$-measurable vector with finite second moment and independent of the Wiener processes $W$ and $V$.

Filtering problems for more general systems were considered in the literature by Kushner [54-58], Bucy[19], Liptser and Shiryayev [68], Zakai [91], Ahmed [8], Ahmed and Dabbous [19] and others. Their work was concentrated mainly on finding an equation for the conditional density (normalized or unnormalized) of the process $\{X(t), t \geq 0\}$ given the observed path $\{y(s), s \leq t\}$.

Assume that all the random processes and vectors described above are defined on a complete filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t \uparrow\subset \mathcal{F}, P)$, where $\mathcal{F}_t \uparrow\subset \mathcal{F}$ means that $\mathcal{F}_t$ is an increasing family of sub-$\sigma$-fields contained in $\mathcal{F}$ for each $t \in [0, T]$. For the stochastic system given by equation (3.1) and equation (3.2) one can use Kushner equation to derive [56] the equations for the conditional moments, in particular, $\widehat{X}(t)$ and $P(t) = [P_{ij}(t)]$. This is given by

$$d\widehat{X}(t) = \hat{f}(t)dt + \frac{1}{\varepsilon_2^2}(\widehat{Xh}(t) - \widehat{X}(t)\hat{h}(t))'(\sigma_0\sigma_0')^{-1}[dy(t) - \hat{h}(t)dt] \qquad (3.3)$$

and

$$\begin{aligned} dP_{ij}(t) &= \{(\widehat{x_i f_j} - \widehat{x_i}\widehat{f_j}) + (\widehat{f_i x_j} - \hat{f}_i\widehat{x_j} + (\varepsilon_1^2 bb')_{ij} \\ &- (\widehat{x_i h(t)} - \widehat{x_i}\widehat{h(t)})'(\varepsilon_2^2\sigma_0\sigma_0')^{-1}(\widehat{h(t)x_j} - \widehat{h(t)}\widehat{x_j})\}dt \\ &+ (\widehat{x_i x_j h}(t) - \widehat{x_i x_j}\hat{h}(t) - \widehat{x_i}\widehat{x_j h}(t) - \widehat{x_j}\widehat{x_i h}(t) + 2\widehat{x_i}\widehat{x_j}\widehat{h(t)}) \\ &\cdot\ (\varepsilon_2^2\sigma_0\sigma_0')^{-1}\left(dy(t) - \hat{h}(t)dt\right) \end{aligned} \qquad (3.4)$$

The solution of the optimal nonlinear filtering using the equations (3.3) and (3.4) can not be obtained without an excessive amount of computation to find $\hat{f}(t)$, $P(t)$, $\hat{h}(t)$. In fact (3.3) and (3.4) constitute an infinite family of moment equations, the first moment requiring the information about the second, the second requiring the information about the third and so on. In other words, this is also an infinite dimensional problem. This led to the idea of the extended Kalman filter (EKF), in which all nonlinear terms are replaced by the first two terms of the Taylor series approximations.

## 3.3 Kalman and Extended Kalman Filtering

### 3.3.1 Kalman Filtering (KF)

If $f(t, X(t)) = A(t)X(t)$ and $h(t, X(t)) = H(t)X(t)$ then we have the standard linear filtering problem. Given that $x(0)$ is Gaussian or constant, then $X(t)$ and $y(t)$ are Gaussian processes, and all the conditional distributions are normal distributions. In particular, we have the Kalman-Bucy filter

$$\begin{aligned} d\widehat{X}(t) &= A(t)\widehat{X}(t)dt + \frac{1}{\varepsilon_2^2}P(t)H'(t)(\sigma_0\sigma_0')^{-1}[dy(t) - H(t)\widehat{X}(t)dt], \\ \widehat{X}(0) &= Ex(0), \end{aligned} \tag{3.5}$$

where $P(t)$ satisfies the matrix Riccati equation

$$\frac{dP}{dt} = A(t)P(t) + P(t)A'(t) + \varepsilon_1^2 b(t)b'(t) - \frac{1}{\varepsilon_2^2}P(t)H'(t)(\sigma_0\sigma_0')^{-1}H(t)P(t); \tag{3.6}$$

with initial condition $P(0) = E\{(x(0) - Ex(0))(x(0) - Ex(0))'\}$. This is a closed form system, and $P(t)$ is independent of the observation $\mathcal{F}_t^y$. This is one of the reasons for the great success of the Kalman-Bucy filter.

### 3.3.2 Extended Kalman Filtering (EKF)

To obtain a mathematically simple but reasonably good filter, we have to find a good approximation for the $\widehat{f}(t)$, the conditional covariance $P(t)$, and $\widehat{h}(t)$ [see equations (3.3-3.4)]. By expanding $f$ and $h$ in a Taylor series about $\widehat{X}(t)$ we have

$$f(t,X(t)) = f(t,\widehat{X}(t)) + A(t,\widehat{X}(t))(X(t) - \widehat{X}(t)) + O(2), \tag{3.7}$$

$$h(t,X(t)) = h(t,\widehat{X}(t)) + H(t,\widehat{X}(t))(X(t) - \widehat{X}(t)) + O(2), \tag{3.8}$$

where $A(t,x) = f_x(t,x)$ and $H(t,x) = h_x(t,x)$.

Now taking the conditional expectations(conditioned upon $\mathcal{F}_t^y$) yields:

$$\begin{aligned} \widehat{f}(t,X(t)) &= f(t,\widehat{X}(t)) + O(2), \\ \widehat{h}(t,X(t)) &= h(t,\widehat{X}(t)) + O(2). \end{aligned} \tag{3.9}$$

Using these approximations and substituting in the equations (3.3) and (3.4) one obtains the so called extended Kalman filter equations

$$d\widehat{X}(t) = f(t,\widehat{X}(t))dt + \frac{1}{\varepsilon_2^2}P(t)H'(t,\widehat{X}(t))(\sigma_0\sigma_0')^{-1}[dy(t) - h(t,\widehat{X}(t))dt], \tag{3.10}$$

$$\begin{aligned} \frac{dP}{dt} = A(t,\widehat{X}(t))P(t) &+ P(t)A'(t,\widehat{X}(t)) + \varepsilon_1^2 bb' \\ &- \frac{1}{\varepsilon_2^2}P(t)H'(t,\widehat{X}(t))(\sigma_0\sigma_0')^{-1}H(t,\widehat{X}(t))P(t). \end{aligned} \tag{3.11}$$

This avoids solving the infinite system of equations (3.3) and (3.4). However, these equations (3.10-3.11) are approximate expressions for propagating the conditional mean of the state and its associated covariance matrix. Since the solutions of these equations do not usually coincide with the entities which they symbolically represent, the conditional expectations of the second terms in the identities (3.7) and (3.8) do not vanish. Thus the derivation of the equations (3.10-3.11) based on this argument is rather optimistic and possibly crude. However if both $f$ and $h$ are very nearly linear, the solutions of these equations may be expected to be close to the true values and the above approximation may be acceptable. In order to overcome this limitation we propose a modification of the extended Kalman Filter.

## 3.4 Modified Extended Kalman Filtering(MEKF)

Here we linearize $f$ and $h$ around quantities which are known rather than quantities which are to be estimated and further we retain the first order terms which are neglected in the EKF. Let $\overline{X}(t)$ be the solution of the following deterministic differential equation.

$$\frac{d\overline{X}(t)}{dt} = f(t, \overline{X}(t)), \quad \overline{X}(0) = Ex(0), \tag{3.12}$$

and let $X(t)$ be the true solution of the following (SDE)

$$dX(t) = f(t, X(t))dt + \varepsilon_1 b(t)dW(t), \qquad X(0) = x(0). \tag{3.13}$$

Note: $\overline{X}(t)$ is not to be confused with the expected value $E(X(t))$.

Define $\tilde{X}(t)$ so that $X(t) = \overline{X}(t) + \tilde{X}(t)$, where $\tilde{X}(t)$ denotes the fluctuation around the deterministic flow $\overline{X}(t)$. The (MEKF) equations are derived by expanding $f$ and $h$ in a Taylor series about $\overline{X}(t)$ which is the exact solution of the nonlinear problem (3.12) and is precisely known while $\widehat{X}(t)$, as used in EKF, is not. The Taylor expansion about $\overline{X}(t)$ is expected to provide a better order of approximation. This is justified by the following theorem.

**THEOREM 1.** Consider the system (3.1) and (3.2) and suppose both $f$ and $h$ are uniformly Lipschitz with Lipschitz constant $C$, and $b \in L_2([0,T], R^{n\times p})$ and $x(0)$ has finite second moment. Then

1. $E\{sup_{t\in[0,T]}\|X(t) - \overline{X}(t)\|^2\} = O(\varepsilon_1^2)$
2. $sup_{t\in[0,T]}E\|f(t,X(t)) - f(t,\overline{X}(t))\|^2 = O(\varepsilon_1^2)$
3. $sup_{t\in[0,T]}E\|h(t,X(t)) - h(t,\overline{X}(t))\|^2 = O(\varepsilon_1^2)$

Further if $x(0)$ has finite fourth order moment and $f$ has uniformly bounded second partials in $x$ then

$sup_{t\in[0,T]}E\{\|f(t,X(t)) - f(t,\overline{X}(t)) - f_x(t,\overline{X}(t))(X(t) - \overline{X}(t))\|^2\} = O(\varepsilon_1^4)$

**proof:** We give an outline of the proof. Define

$\phi(t) \equiv E\{sup_{s\in[0,T]}\|X_s - \overline{X}_s\|^2\}$

Subtracting equation (3.12) from equation (3.13), using the Lipschitz property of $f$, and the Martingale inequality [8, p. 325] one can verify that

$\phi(t) \leq 2TC^2 \int_0^t \phi(s)ds + 8\varepsilon_1^2 \int_0^T \|b(s)\|^2 ds.$

Using Gronwall inequality, it follows from this that there is a constant $K$ depending only on $C$, $T$ and the norm of $b$ so that

$\phi(T) \leq K\varepsilon_1^2$

This proves the first estimate. The second and the third estimates follow from the first and Lipschitz properties. For the proof of the last assertion we use the Lagrange formula

$f^i(t, X(t)) - f^i(t, \overline{X}(t)) - (f^i_x(t, \overline{X}(t)), X(t) - \overline{X}(t)) = \int_0^1 \alpha d\alpha \int_0^1 d\theta < f^i_{xx}(\overline{X}(t) + \theta\alpha(X(t) - \overline{X}(t)))(X(t) - \overline{X}(t)), (X(t) - \overline{X}(t)) >$.

Taking the norm, it follows from the uniform boundedness of the second partials of $f$ that there exists a positive constant $b$ such that

$\|f(t, X(t)) - f(t, \overline{X}(t)) - f_x(t, \overline{X}(t))(X(t) - \overline{X}(t))\|^2 \leq b\|X(t) - \overline{X}(t)\|^4$

It is well known that under the given assumptions, $X(t)$ has fourth order moment whenever $X(0)$ has the same [8, Corollary 7.4.5 ,p359]. Using this fact and the above inequality the last assertion can be proved by use of Ito's formula [8, Theorem 7.2.6,p337] .This outlines the proof.

Based on the above justification, and letting $A(t,x)$ and $H(t,x)$ denote the hessian matrices $f_x$ and $h_x$ respectively, we can write

$$f(t, X(t)) = f(t, \overline{X}(t)) + A(t, \overline{X}(t))\tilde{X}(t) + o(\tilde{X}(t)) \qquad (3.14)$$

$$h(t, X(t)) = h(t, \overline{X}(t)) + H(t, \overline{X}(t))\tilde{X}(t) + o(\tilde{X}(t)) \qquad (3.15)$$

Substituting these equations in the model equations (3.1-3.2) we obtain the dynamics of fluctuation ($\tilde{X}(t)$) and measurement as given by:

$$d\tilde{X}(t) = A(t, \overline{X}(t))\tilde{X}(t)dt + \varepsilon_1 b(t)dW(t), \quad \tilde{X}(0) = \tilde{x}(0), \qquad (3.16)$$

$$dy(t) = \{h(t, \overline{X}(t)) + H(t, \overline{X}(t))\tilde{X}(t)\}dt + \varepsilon_2\sigma_0(t)dV(t), \quad y(0) = 0. \qquad (3.17)$$

Thus, the filtering problem is decomposed into two sub problems. The first problem is related to the deterministic flow governed by the initial value problem (3.12) and the second problem is related to the dynamics of (stochastic) fluctuation given by equation (3.16) along with the observation equation given by (3.17).

Clearly for given $\{\overline{X}(t), t \geq 0\}$, the equations (3.16) and (3.17) are linear, and $\{\tilde{X}(t), t \geq 0\}$ and $\{y(t), t \geq 0\}$ are Gaussian processes. Defining the minimum variance unbiased estimator of $\tilde{X}(t)$ as $\widehat{\tilde{X}}(t) = E\{\tilde{X}(t)/\mathcal{F}_t^y\}$ and the error covariance matrix as $P(t) = E\{(\tilde{X}(t) - \widehat{\tilde{X}}(t))(\tilde{X}(t) - \widehat{\tilde{X}}(t))'/\mathcal{F}_t^y\}$ and using Kalman-Bucy theory for linear systems, the optimal estimator equation for the model is given by:

$$\begin{aligned} d\widehat{\tilde{X}}(t) = A(t, \overline{X}(t))\widehat{\tilde{X}}(t)dt \quad &+ \quad \frac{1}{\varepsilon_2^2}P(t)H'(t, \overline{X}(t))(\sigma_0\sigma_0')^{-1} \cdot [dy(t) \\ &- \quad \{h(t, \overline{X}(t)) + H(t, \overline{X}(t))\widehat{\tilde{X}}(t)\}dt], \qquad (3.18) \end{aligned}$$

and the covariance equation is given by :

$$\begin{aligned} \frac{dP}{dt} = A(t, \overline{X}(t))P(t) \quad &+ \quad A'(t, \overline{X}(t))P(t) + \varepsilon_1^2 bb' \\ &- \quad P(t)H'(t, \overline{X}(t))(\varepsilon_2^2\sigma_0\sigma_0')^{-1}H(t, \overline{X}(t))P(t), \ (3.19) \end{aligned}$$

with initial conditions $\widehat{\tilde{X}}(0) = 0$ and $P(0) = 0$. The error covariance matrix $P(t)$ of the process $X(t)$ is the same as that of $\tilde{X}(t)$. In the (EKF) the estimate $\widehat{X}(t)$ is obtained by solving equations (3.10-3.11), where as in (MEKF) it is given by

$\widehat{X}(t) \cong \overline{X}(t) + \widehat{\tilde{X}}(t)$ as obtained by solving equations (3.12), (3.18) and (3.19) simultanuously. The quality of performance of the two estimators (EKF) and (MEKF) is illustrated by three examples in the following Section.

## 3.5 Examples and Illustrations

For illustration of results discussed in the preceding sections, we present here three examples. In order to show the behavior of the (EKF) and (MEKF) we carried out the simulation with sampling interval $\Delta = 0.01 sec$. For comparison of performance of MEKF and EKF, we plot the integral squared error(Figs.3.3,3.7,3.9) as a function of the dynamics noise $\varepsilon_1 \rightarrow J_T(\varepsilon_1)$ for fixed measurement noise $\varepsilon_2$; and similarly for fixed dynamics noise say $\varepsilon_1$ we plot $\varepsilon_2 \rightarrow J_T(\varepsilon_2)$ (Figs.3.4,3.8,3.10).

**Example 1:** The approximation theory developed in the previous sections will be applied to the momentum dynamics of a $3-$ *axis geosynchronous satellite* given by

$$\begin{aligned} I_x\frac{dp}{dt} + (I_z - I_y)(q - \omega_0)r &= T_x \\ I_y\frac{dq}{dt} + (I_x - I_z)pr &= T_y \\ I_z\frac{dr}{dt} + (I_y - I_x)p(q - \omega_0) &= T_z \end{aligned} \tag{3.20}$$

where $T_x, T_y, T_z$ is the input torque, and $p, q, r$ are the components of the angular momentum vector of the satellite in the $x, y, z$ axes, respectively. A stochastic version of this equation can be written as

$$\begin{aligned}
dX_1(t) &= \frac{1}{I_x}\{T_x - (I_z - I_y)(X_2(t) - \omega_0)X_3(t)\}dt + \varepsilon_1 dW_1(t) \\
dX_2(t) &= \frac{1}{I_y}\{T_y - (I_x - I_z)X_1(t)X_3(t)\}dt + \varepsilon_1 dW_2(t) \\
dX_3(t) &= \frac{1}{I_z}\{T_z - (I_y - I_x)X_1(t)(X_2(t) - \omega_0)\}dt + \varepsilon_1 dW_3(t) \qquad (3.21)
\end{aligned}$$

with the measurement dynamics given by

$$\begin{aligned}
dy_1(t) &= h_1 X_1(t)dt + \varepsilon_2 dV_1(t) \\
dy_3(t) &= h_3 X_3(t)dt + \varepsilon_2 dV_3(t) \qquad (3.22)
\end{aligned}$$

where $X_1 = p$, $X_2 = q$, $X_3 = r$ are the states (attitude rates) and $y_1$, $y_2$ are the observations. For illustration we have chosen the following values of the model parameters: $I_x = 645$, $I_y = 100$, $I_z = 669 slugft^2$, $\omega_0 = 7.29 \times 10^{-5} rad/sec$, $h_1 = h_2 = 1.0$. Both (EKF) algorithm and (MEKF) algorithm (eqs.3.13,3.19,3.20) have been applied to this example. For performance evaluation, the integral-squared error,

$$J_T = \frac{1}{T}\int_0^T \|X(t) - \widehat{X}(t)\|^2 dt$$

is computed and plotted as functions of the noise intensity parameters $\varepsilon_1$ and $\varepsilon_2$. The actual state $\{X(t), t \geq 0\}$ is obtained by simulation and the estimated states $\widehat{X}(t)$, corresponding to (EKF) and (MEKF), are computed. The results, as shown in (Figs.3.1,3.2,3.3,3.4), clearly indicate that (MEKF) performs better than (EKF).

**Example 2:** A stochastic version of the dynamics of a simple pendulum with observation $y$ is given by

$$\begin{aligned} dX_1(t) &= X_2(t)dt, \quad X_1(0) = \theta_0; \\ dX_2(t) &= -\frac{g}{l} sin(X_1(t))dt + \varepsilon_1 dW(t), \quad X_2(0) = 0; \end{aligned} \tag{3.23}$$

$$dy(t) = X_2(t)dt + \varepsilon_2 dV(t), \quad y(0) = 0; \tag{3.24}$$

where $l$ is the rod length, $\theta(t)$ is the angular displacement, $g$ is the acceleration due to gravity, $X_1 = \theta$, $X_2 = \dot{\theta}$ and $W$, $V$ are independent Wiener processes.

Again we apply the (EKF) and (MEKF) algorithms for the same performance measure as in the previous example. These are computed as shown in (Figs. 3.5,3.6,3.7,3.8). Again the performance of the (MEKF) is better than that of the (EKF). If the (Figs.3.5,3.6) and are carefully scrutinized, one would find that the estimated state (dotted curves) follow the actual state (solid curves) more closely in the case of (MEKF) (Fig.3.6) than (EKF) (Fig.3.5).

**Example 3:** We consider the self-excited oscillator given by the following equation:

$$\frac{d^2u}{dt^2} + (\beta(\frac{du}{dt})^2 - \alpha)\frac{du}{dt} = 0, \quad u(0) = u_0, \quad \frac{du}{dt}(0) = 0. \tag{3.25}$$

where $u$ is scalar function of time, $\alpha$, $\beta$ and $u_0$ are constants. A stochastic version of this system is given by

$$\begin{aligned} dX_1(t) &= X_2(t)dt \\ dX_2(t) &= \{-X_1(t) + \alpha X_2(t) - \beta X_2(t)^2\}dt + \varepsilon_1 dW(t), \end{aligned} \tag{3.26}$$

along with the observation process

$$dy(t) = h_2 X_2(t)dt + \varepsilon_2 dV(t). \tag{3.27}$$

where $X_1 = u$, $X_2 = \frac{du}{dt}$. Apply the (EKF) and (MEKF) algorithm for the same performance measures as in the previous two examples. Again the (MEKF) is better than that of the (EKF) as shown in (Figs. 3.9, 3.10).

## 3.6 Summary and Conclusion

We have suggested a modification of the extended Kalman filter (EKF) called a modified extended Kalman filter(MEKF). We have also presented a theoretical justification for this modification. The modified filter has performance better than the original extended Kalman filter as illustrated by several examples. However at lower levels of noise power both the filters perform equally. As the noise power increases, the (MEKF) continues to perform better than (EKF) as verified by all the three examples.

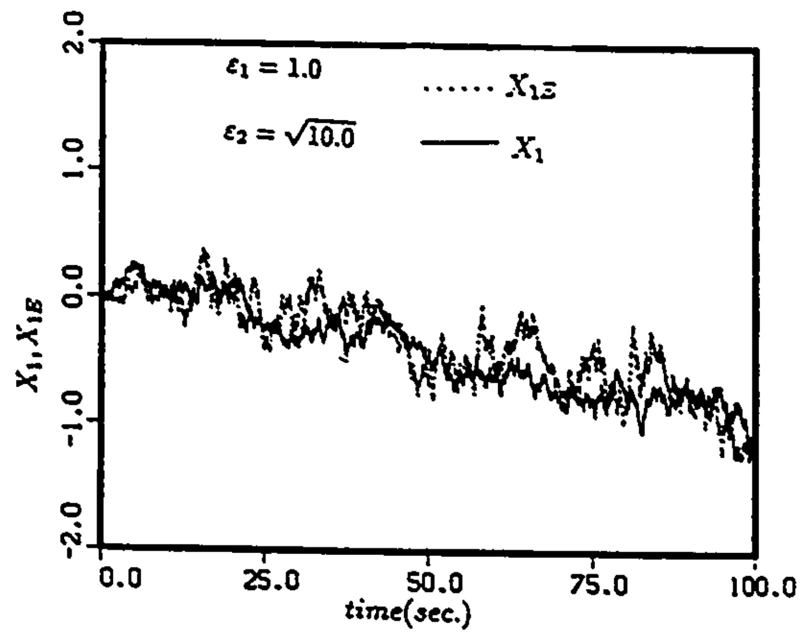


Figure 3.1: $X_{1E} - EKF$ estimated state, $X_1$- the actual state

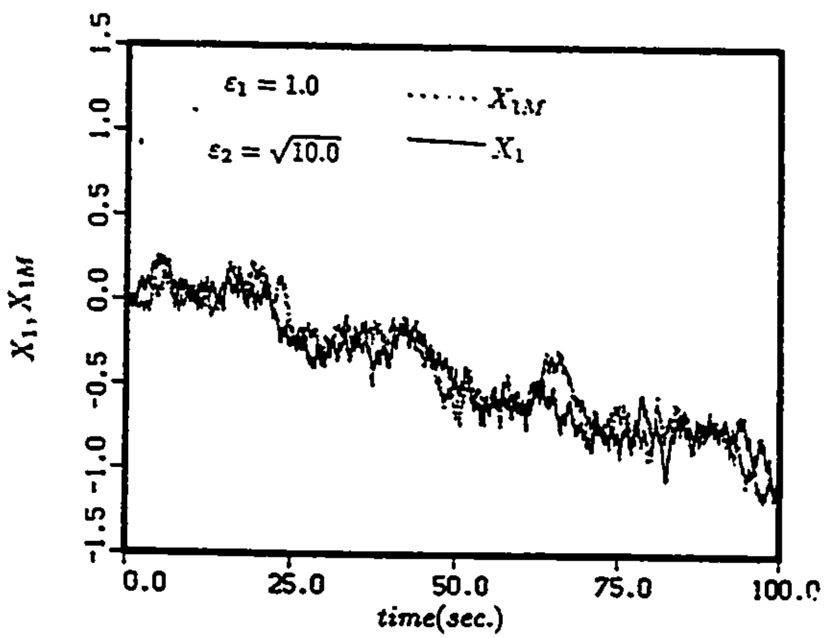


Figure 3.2: $X_{1M} - MEKF$ estimated state, $X_1$- the actual state

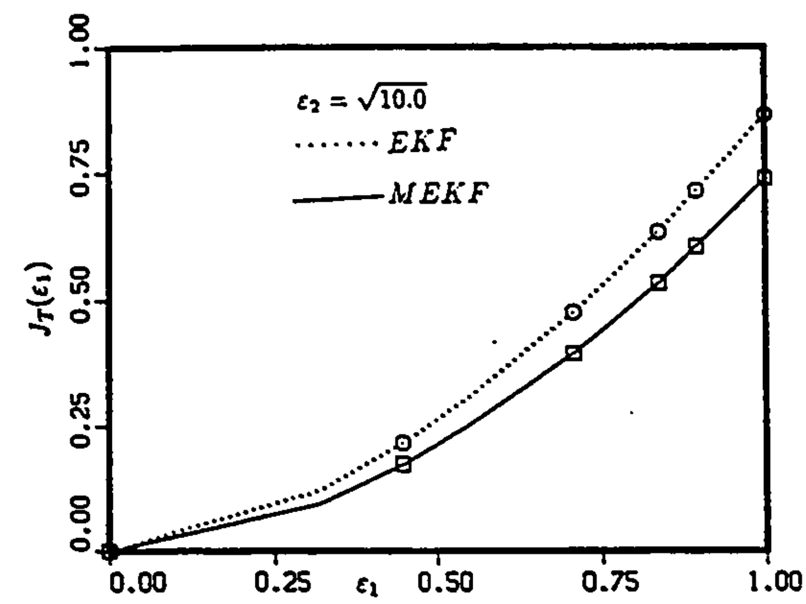


Figure 3.3: integral squared error $J_T(\varepsilon_1)$

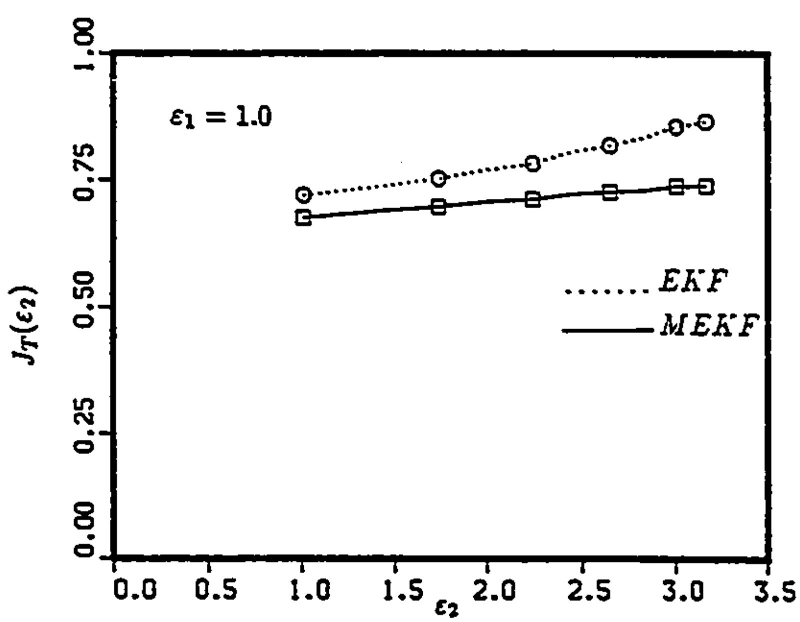


Figure 3.4: integral squared error $J_T(\varepsilon_2)$

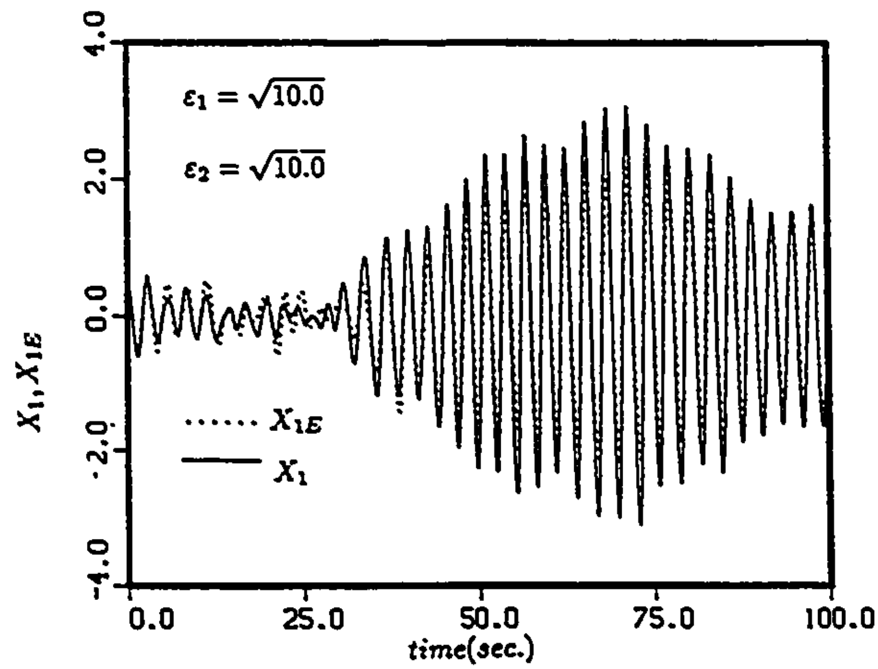


Figure 3.5: $X_{1E} - EKF$ estimated state, $X_1$- the actual state

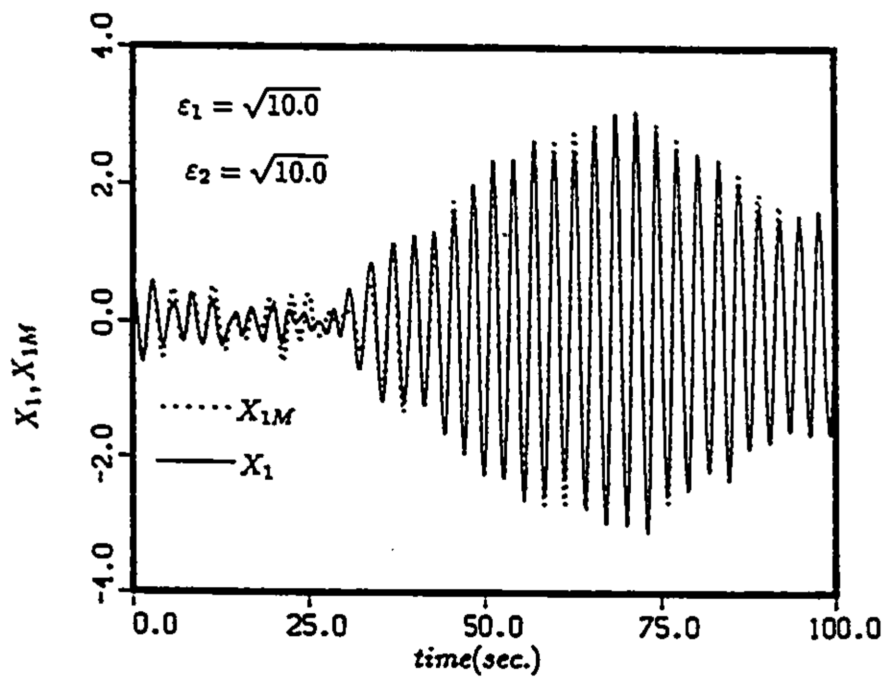


Figure 3.6: $X_{1M} - MEKF$ estimated state, $X_1$- the actual state

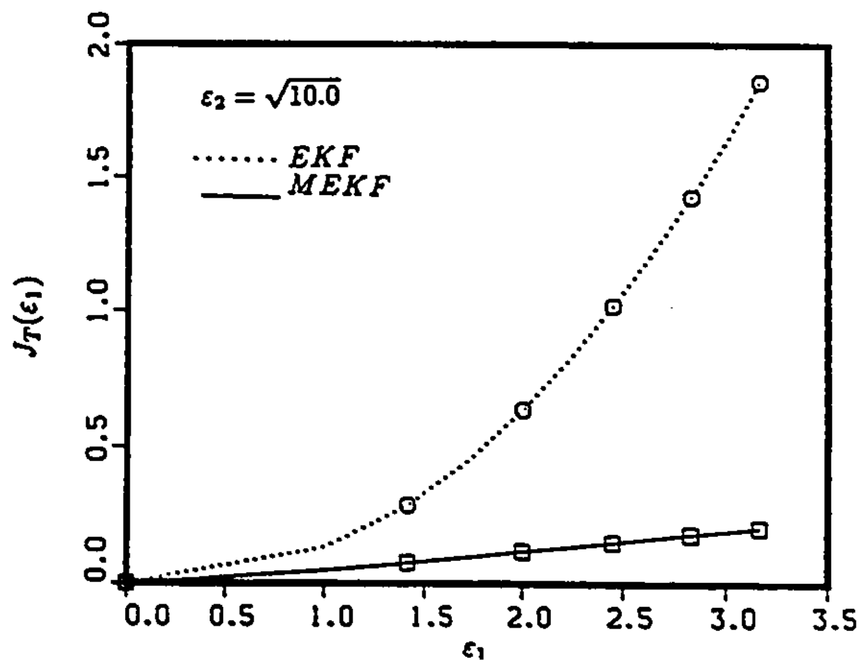


Figure 3.7: integral squared error $J_T(\varepsilon_1)$

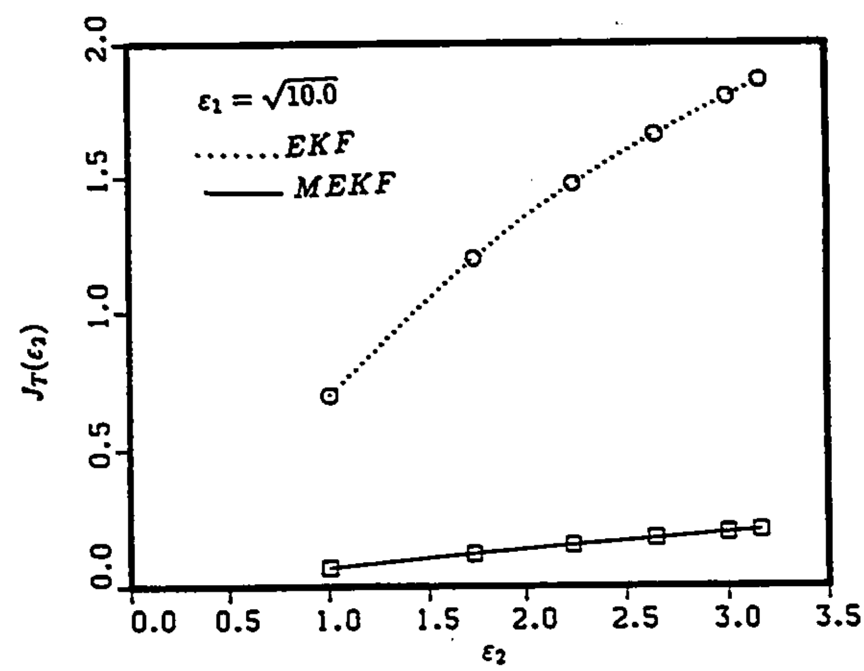


Figure 3.8: integral squared error $J_T(\varepsilon_2)$

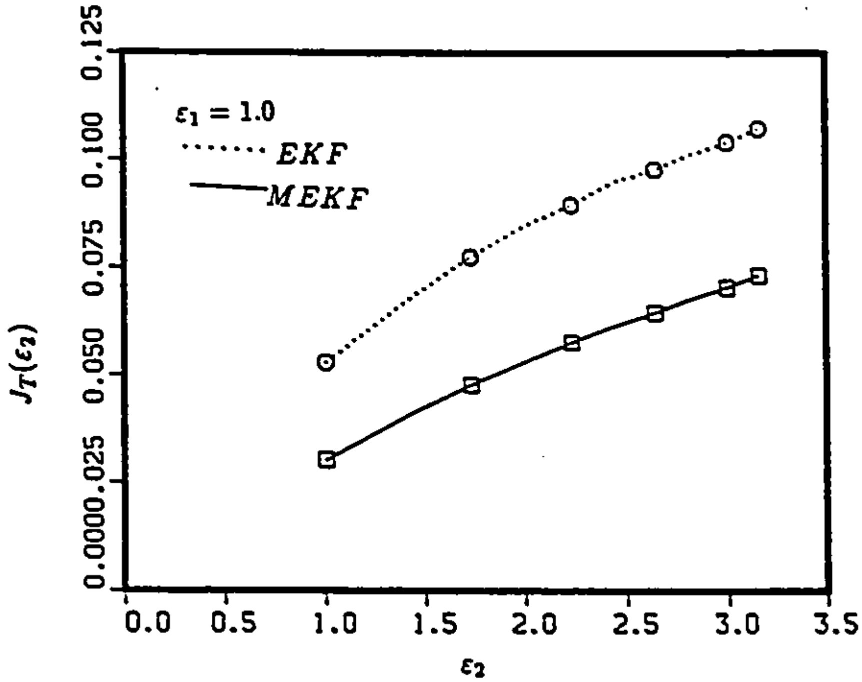


Figure 3.9: integral squared error $J_T(\varepsilon_1)$

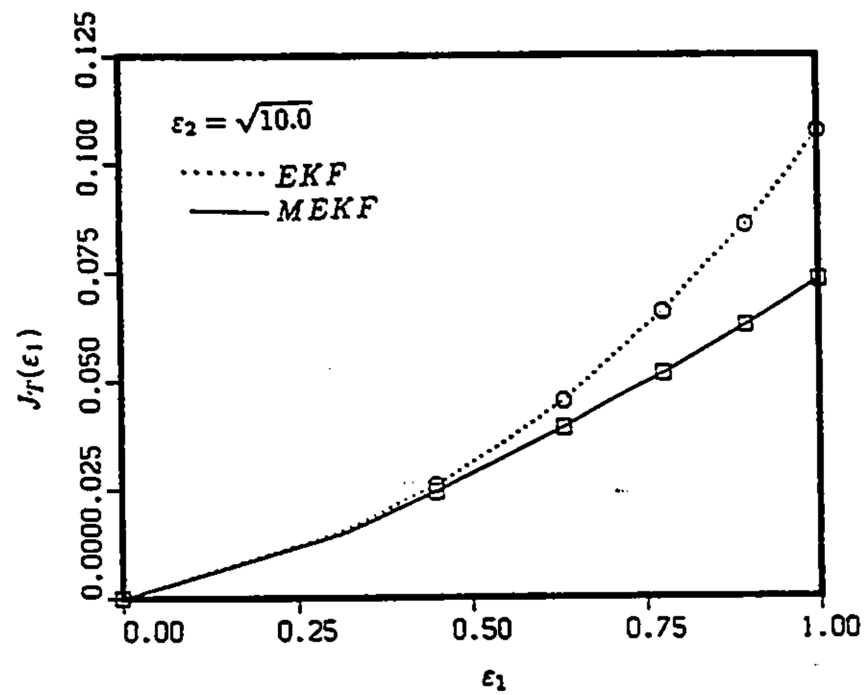


Figure 3.10: integral squared error $J_T(\varepsilon_2)$

# Chapter 4

# NONLINEAR FILTERING (ZAKAI EQUATION UNCORRELATED CASE)

## 4.1 Introduction

Over the last decade, considerable attention has been focused on nonlinear filtering problems for which the correct solution is given only by the solution of an associated stochastic partial differential equation called Zakai equation [91]. The solution of Zakai equation holds the key to the solution of partially observed stochastic control problems [18], since it provides the natural state for the equivalent fully observed problem, which the feedback controller utilizes. Similarly, Zakai equation plays a central role in the parameter identification of nonlinear stochastic systems [26]. To solve nonlinear filtering problems avoiding stochastic partial differential equations, in recent years, significant attention has been given to the so called extended Kalman filtering (EKF) which is a linearized approximation of the original problem.

It is well known that the EKF does not always perform well and in fact performs poorly if the nonlinearities are strong. Further, in any specific application, the accuracy of results obtained using EKF is not verifiable and consequently not reliable. For verification of the reliability of such approximations (EKF), it is necessary to compare these results against the exact solution of Zakai equation or a genuine approximation of it. In particular, this means that the solution of Zakai equation is to be our ideal reference and that we must be able to rely on an efficient and accurate algorithm to solve it.

Thus the problem of constructing solutions of Zakai equation is a major issue for practical applications. Since, in general, it is impossible to obtain analytical solution of the Zakai equation for nonlinear systems, it is reasonable to seek a numerical scheme that is implementable in straight forward manner and provides accurate solutions. Lie algebraic calculations give some new insights into certain nonlinear estimation problems and guidance in the search for finite dimensional estimators. Only in very special cases [45] does this approach lead to finite dimensional filters. The Hermite series has been proposed and investigated in [89]; the disadvantage of this method is that when it is truncated the resulting series approximation can take negative values and hence is not itself a valid density function. To avoid or at least reduce the nonpositivity of this approximation, it is sometimes necessary to retain a very large number of terms in the series, demanding excessive CPU time. Thus it is reasonable to seek other approximations avoiding nonpositivity and excessive computation time. The Gaussian series approximation proposed in this Chapter avoids nonpositivity and has the added advantage of providing a sequence of $C^\infty$- functions that belongs to the domain of the adjoint of the infinitesimal generator $A$ of the Markov process being filtered. This is very important since numerical computation involves operation of the basis functions by the adjoint operator $A^*$. Since Galerkin approximation is successfully used in solving deterministic partial differential equations, it seems natural to extend these methods to stochastic equations.

In this Chapter, we present a technique for constructing optimal nonlinear filters using a digital computer. The conditional density of the signal given the observations, is approximated by a finite positive (convex) combination of a sequence of linearly independent parametrized family of Gaussian densities on $R^n$. The method leads to very accurate realization of optimal nonlinear filters. The rest of the Chapter is organized as follows: In Section 4.2, a brief account of the filtering equations is given. In Section 4.3 we present Galerkin approximation for solutions of Zakai equation by solving a sequence of finite dimensional stochastic differential equations. We also discuss the crucial advantages of using Gaussian series. In Section 4.4, we present the spatial discretization and the computational method. In Section 4.5, examples (for which exact analytical solutions are available) and the corresponding simulation results are presented.

## 4.2 The Filtering Problem

In this Chapter, we consider the class of systems governed by Ito differential equations of the form

$$\begin{aligned} dX(t) &= f(X(t))dt + b(X(t))dW(t), \qquad X(0) = x(0), \qquad t \geq 0, \\ dy(t) &= h(X(t))dt + \sigma_0(t)dV(t), \qquad y(0) = 0, \qquad t \geq 0, \end{aligned} \tag{4.1}$$

where the state $X(t) \in R^n$ is the unobserved process to be estimated, $y(t) \in R^m$ is the observation process, the processes $\{W, V\}$ are $\{R^p, R^q\}$ valued independent standard Wiener processes, $f : R^n \to R^n$ and $b : R^n \to R^{n\times p}$ and $h : R^n \to R^m$ and $\sigma_0$ is $m \times q$ matrix and $x(0)$ is an $\mathcal{F}_0-$ measurable vector independent of the Wiener processes $W$ and $V$. Let $g$ be any bounded measurable function on $R^n$ with values in $R$. The problem is to estimate $g(X(t))$ given the history of $y$ up to time $t$.

The filtering problem for the above system was considered in the literature by Kushner [54-58], Bucy [19], Liptser and Shiryayev [68], Zakai [91], Ahmed [8] and others. Their work concentrated mainly on finding an equation for the conditional density (normalized or unnormalized) of the process $\{X(t), t \geq 0\}$ given the observed path $\{y(s), s \leq t\}$, with the help of which the filtering problem is resolved. It is known that the conditional expectation gives the best (optimal) estimate in the mean square sense.

In fact the conditional density can be computed by two methods, which are basically the same. The first method is to solve Kushner equation [57], which is a nonlinear stochastic partial differential equation. The second method is to solve Zakai equation [91], which is linear stochastic partial differential equation describing the flow of the unnormalized density, which is the subject of this Chapter. Let the state and the observed processes be governed by the stochastic differential equations (4.1). Let $(\Omega, \mathcal{F}, \mathcal{F}_t \uparrow\subset \mathcal{F}, P)$ denote the basic filtered probability space with respect to which all random variables and processes referred to in this Chapter are defined. Let $\mathcal{F}_t^y$ denote the $\sigma$-field generated by $y(s)$ for $s \leq t$, and completed with respect to the basic probability measure P. In filtering the ultimate goal is to calculate $\hat{g}(t)$. It is a standard fact that $\hat{g}(t)$ is the best (unbiased minimum variance) mean square $\mathcal{F}_t^y$-measurable estimate for $g(X(t))$. For the solution of Zakai equation (1.15), it is convenient to rewrite this equation in the abstract setting so that we can exploit existing theory of differential equations on Hilbert spaces. Let $H$ denote the Hilbert space $L_2(R^n)$ and $H^1(R^n)$ denote the standard Sobolev space. Define the operator $B$ by

$$\begin{aligned} D(B) &= \{\varphi \in L_2(R^n) : \; h_i\varphi \in L_2(R^n), \; i = 1, 2, \dots m\} \\ B\varphi &= h \cdot \varphi \;\; for \; \varphi \in D(B). \end{aligned}$$

Then we can rewrite the Zakai equation as an abstract stochastic evolution equation.

$$du = A^* u dt + B(u) \cdot \Gamma_0^{-1} dy$$

$$u(0) = u_0 = p_0 \in H \ . \tag{4.2}$$

on the Hilbert space $H$.

For existence and uniqueness of solutions of equation (4.2) and their regularity properties we need the following Lemma.

**Lemma 1.** Suppose the coefficients of the operator $A$ as defined by the expression (1.14) satisfy the following assumptions:

(a1) $f_i \in C_b(R^n)$, $i = 1, 2, ..., n$.

(a2) $(bb')_{ij} \in C_b^1(R^n)$ and there exist $\alpha > 0$ such that $((bb')\zeta, \zeta) \geq \alpha |\ \zeta\ |^2$, for all $\zeta \in R^n$, $x \in R^n$.

Then,

(c1): $A^*$, the adjoint of $A$, is a bounded linear operator from $H^1$ to $H^{-1}$ (i.e., $A^* \in \mathcal{L}(H^1, H^{-1})$), and

(c2): there exist $\alpha > 0$ (same as in (a2)) and $\lambda \geq 0$ such that

$\alpha \|\ u\ \|_{H^1}^2 + < A^* u, u >_{H^{-1}, H^1} \ \leq \lambda |\ u\ |_H^2$.

**Proof:** Under the assumptions (a1) and (a2), one can verify that both $A$ and $A^* \in \mathcal{L}(H^1, H^{-1})$ and satisfy Garding's inequality as given in (c2) [37].

**Theorem 2.** Suppose the assumptions of the previous Lemma hold and there exists a constant $\eta > 0$ such that $|\ Bu\ |_{L_2(R^n; R^m)}^2 \leq \eta |\ u\ |_H^2$.

Then the Zakai equation (4.2) has a unique solution

$u \in L_\infty(I, H) \cap L_2(I, H^1)$, and further $u \in C(I, H)$, $P - a.s.$

**Proof:** Follows from well known results given by [6,30,75].

The solution of Zakai equation can be constructed by Galerkin method using any suitable set of basis functions from the Hilbert space $H = L_2(R^n)$. It is possible to choose a complete set of basis functions, like the Hermite functions in $n$-dimension, and construct the approximate solution by solving a sequence of finite dimensional stochastic differential equations obtained by the Galerkin projection. There are two shortcomings in the choice of the Hermite functions in the numerical computations:

(i) These functions take both positive and negative values, whereas the probability density function has only positive values.

(ii) In order to preserve positivity one has to retain a very large number of terms in the series demanding excessive CPU time.

In this Chapter we propose a set of basis functions which are positive for all $x \in R^n$ and $C^\infty$-smooth.

## 4.3 Galerkin Method Using Gaussian Series Approximation

### 4.3.1 The Spatial Domain for Solution of Zakai Equation

Theoretically Zakai equation is defined on the whole of $R^n$ and apparently it is required to solve it on this unbounded domain. Computationally this is not feasible. Since the initial density $p_0$ is approximately supported on a bounded subset of $R^n$, the corresponding solution of Zakai equation is also expected to be approximately supported on possibly another bounded set. Therefore we can restrict the computations on bounded sets. It is known [32] that the behavior of unnormalized conditional density is as good as that of the initial density. In other words, if $p_0(\cdot)$ is smooth so also is $\Psi(t,\cdot)$, and $\Psi(t,\cdot)$ decays to zero as $\| x \| \to \infty$, provided $p_0$ has this property. Therefore, we can select a sufficiently large domain $\Sigma \subset R^n$ where a significant part of the probability mass is concentrated. This can be approximately estimated from the approximate $\epsilon$-support of $p_0$ itself. For example, let $a > 0$ and $\Sigma_a \subset R^n$ denote the cube centered at the origin ( or centered at the position of the mean of $p_0$) having edges of size $2a$. For any given $\epsilon > 0$, we choose $a$ large enough so that

$$\Pi_0(\Sigma_a) \equiv \int_{\Sigma_a} p_0(x)dx > 1 - \epsilon. \tag{4.3}$$

If required we can double or triple the size of $\Sigma_a$ for the solution of Zakai equation and establish a grid for spatial discretization on this larger domain and impose a

Dirichlet boundary condition. In other words, we can always establish a grid in the region of the state space that supports significant part of the probability mass.

### 4.3.2 Completeness of a Special Class of Basis Functions

Now we shall introduce the basis functions we have proposed in this Chapter and discuss their properties. Let $M_s^+$ denote the class of $(n \times n)$ positive symmetric matrices. Define the family of functions, $\{w_i(x), x \in R^n, i = 1, 2, \ldots\}$, as follows

$$w_i(x) \equiv w(x, m_i, B_i) = \frac{1}{\sqrt{(2\pi)^n det B_i}} e^{-\frac{1}{2}(x-m_i)' B_i^{-1}(x-m_i)}, \quad x \in R^n, \qquad (4.4)$$

parametrized by $m_i$, $B_i$ where $m_i \in R^n$, $B_i \in M_s^+$, $m_i \neq m_j$, for $i \neq j$.

The first question that must be settled before one can use this sequence as basis functions is weather or not this can approximate any $L_2(R^n)$ function in the mean square sense. This is a question of completeness of the set $\{w_i\}$ in the class $L_2$. A system of functions , say $\Gamma \subset L_2$, is said to be closed in the class $L_2$ if $(f, \gamma)_{L_2} = 0$ for all $\gamma \in \Gamma$ implies that $f = 0$. In $L_2$ spaces completeness and closure are equivalent [ see Tricomi 85, p.90 ] . We prove completeness of the set $\{w_i\}$.

Let $K$ and $M$ denotes countable dense subset of $R^n$ and $M_s^+$ respectively. Consider the class of functions $\{w_i(\cdot)\} \equiv \{w(m_i, B_i, \cdot), \ m_i \in K, m_i \neq m_j, \ i \neq j, \ B_i \in M\}$.

**Theorem 3.** The system of functions $\{w_i\}$ is linearly independent and complete in the class $L_2$.

**Proof:** Since $m_i \neq m_j$ for $i \neq j$, a linear combination like $\sum_{1 \leq i \leq m} a_i w_i$ can never equal a function which is identically zero in $R^n$ unless all the coefficients $\{a_i\}$ vanish. This is independent of our choice of $B_i$, $B_j \in M$. Thus the set is linearly independent. Recall that any system of linearly independent vectors can be orthogonalized by Gram-Schmidt orthogonalization procedure and then also normalized if required. Hence without any loss of generality we may assume that the sequence $w_i$ is orthonormal. As noted above, for $L_2$ spaces completeness and closure are equivalent . Thus for the proof of completeness, it suffices to show that the original set $\{w_i\}$ (not necessarily orthonormal) is closed in the class $L_2$. We prove this by contradiction. Suppose there exists a $\varphi \in L_2(R^n)$ such that

$$a_i \equiv (\varphi, w_i)_{L_2} = 0 \ for \ all \ i = 1, 2, ... \tag{4.5}$$

but $\varphi \neq 0$ in the $L_2$ sense. Then there exist a set $E(\neq \emptyset) \subset R^n$ of positive Lebesgue measure such that $\varphi(x) \not\equiv 0$ for $x \in E$. Without loss of generality suppose $\varphi(x) > 0$ on $E$. Since $K$ is dense in $R^n$ we can choose an $m_i \in E \cap K$. Then for any $\epsilon > 0$, we can choose $\gamma_i = \gamma_i(\epsilon)$, sufficiently small and positive, and $B_i = \gamma_i I$ so that

$$\int_E w(m_i, B_i, x)dx = \int_E w_i(x)dx > 1 - \epsilon. \tag{4.6}$$

It is clear from this that we can choose $\epsilon > 0$ sufficiently small such that

$$(\varphi, w_i) = \int_{R^n} w_i(x)\varphi(x)dx > 0. \tag{4.7}$$

This contradicts (4.5) and hence the sequence $\{w_i\}$ is closed in the class $L_2(R^n)$ and therefore complete. This completes the proof.

In view of the above result we can use $\{w_i\}$ as a basis. Since $w_i \in C^\infty(R^n) \cap L_2(R^n)$, it is clear that $\{w_i\} \subset D(A^*)$.

### 4.3.3 Galerkin Approximation

For the purpose of numerical calculations it is more convenient to formulate Zakai equation (1.15) in terms of Stratonovich sense rather than Ito sense. Its transformation to the Stratonovich form is done by adding the Wong-Zakai correction to the drift term [87]. Thus, equation (1.15) takes the following Stratonovich form

$$d\Psi(t,x) = \{A^*\Psi(t,x) - \frac{1}{2}\Psi(t,x)h(x)\cdot\Gamma_0^{-1}h(x)\}dt + \Psi(t,x)h(x).\Gamma_0^{-1}dy(t) \tag{4.8}$$

where the term$(-\frac{1}{2}\Psi h.\Gamma_0^{-1}h)$ is the Wong-Zakai correction. It is well known [81] that the Stratonovich differentials obey the usual laws of calculus.

One can convert the family of nonorthogonal functions $\{w_i(x), x \in R^n, i = 1, 2, ...N\}$ into a family of orthogonal functions $\{v_i(x), x \in R^n, i = 1, 2, ..N\}$ using the Gram-Schmidt procedure as follows: First $v_1 = w_1$, and then each $v_i$ should be orthogonal to the preceding $v_1, v_2, ...v_{i-1}$:

$$v_i = w_i - \sum_{j=1}^{i-1} \frac{(v_j, w_i)_{L_2(R^n)} v_j}{(v_j, v_j)_{L_2(R^n)}}. \tag{4.9}$$

It is clear from the previous expression that in order to get $\{v_i(x), x \in R^n, i = 1, 2, ...., N\}$ we have to do a huge a mount of computations which demand excessive CPU time. Therefore, it is reasonable and very easy for the purpose of numerical computations, to implement the approximate solution of Zakai equation (4.8) using the non orthogonal Gaussian series as basis functions.

Hence using the Galerkin method based on the nonorthogonal Gaussian sequence one can approximate the solution of equation (4.8) in the form

$$\Psi^N(t, x) = \sum_{i=1}^{N} \psi_i^N(t)\omega_i(x), \tag{4.10}$$

where $\{\psi_i^N\}$ are the Fourier coefficients to be chosen as follows. By projecting equation (4.8) into the space spanned by $\{w_i, \ 1 \leq i \leq N\}$, and substituting expression (4.10) in it we obtain

$$\begin{aligned} \sum_{i=1}^{N} d\psi_i^N(t)(w_i, w_j) &= \{\sum_{i=1}^{N} \psi_i^N(t)(A^* w_i, w_j) - \frac{1}{2}\sum_{i=1}^{N} \psi_i^N(t)(hw_i \cdot \Gamma_0^{-1} h, w_j)\}dt \\ &+ \sum_{i=1}^{N} \psi_i^N(t)(\Gamma_0^{-1} hw_i, w_j) \cdot dy(t), \ 1 \leq j \leq N. \end{aligned} \tag{4.11}$$

This is a family of finite dimensional ($N$) stochastic differential equations for $\Upsilon^N \equiv [\psi_1^N, \psi_2^N, ....., \psi_N^N]'$ given by

$$\begin{aligned} \sum_{i=1}^{N} a_{ji} d\psi_i^N(t) = \{\sum_{i=1}^{N} b_{ji}\psi_i^N(t) &- \frac{1}{2}\sum_{k=1}^{m}\sum_{i=1}^{N} d_{ji}^k \psi_i^N(t)\}dt \\ &+ \sum_{k=1}^{m}\sum_{i=1}^{N} c_{ji}^k \psi_i^N(t)(\Gamma_0^{-1} dy(t))_k \\ \sum_{i=1}^{N} a_{ji}\psi_i^N(0) &= p_{0j}, \ j = 1, 2, ....N; \end{aligned} \tag{4.12}$$

where

$$a_{ji} = \int_{R^n} w_j w_i dx \equiv (w_j, w_i) \tag{4.13}$$

$$b_{ji} = \int_{R^n} w_j A^* w_i dx \equiv (w_j, A^* w_i) \tag{4.14}$$

$$c^k_{ji} = \int_{R^n} w_j h_k w_i dx \equiv (w_j, h_k w_i) \tag{4.15}$$

$$d^k_{ji} = \int_{R^n} h_k w_j w_i (\Gamma_0^{-1} h)_k dx \tag{4.16}$$

$$d_{ji} = \sum_{k=1}^{m} d^k_{ij} \tag{4.17}$$

$$p_{0j} = \int_{R^n} p_0 w_j dx. \tag{4.18}$$

In the matrix notation the system (4.8) is equivalent to

$$\mathcal{A}_N d\Upsilon^N = (\mathcal{B}_N - \frac{1}{2}\mathcal{D}_N)\Upsilon^N dt + \sum_{k=1}^{m} \mathcal{C}^k_N \Upsilon^N (\Gamma_0^{-1} dy)_k \,, \tag{4.19}$$

$$\mathcal{A}_N \Upsilon^N(0) = p_0^N, \tag{4.20}$$

where $\mathcal{A}_N = [a_{ji}]$, $\mathcal{B}_N = [b_{ji}]$, $\mathcal{D}_N = [d_{ji}]$, $\mathcal{C}^k_N = [c^k_{ji}]$, and $p_0^N = [p_{01}, p_{02}, \ldots\ldots, p_{0N}]'$, and $(\cdot)_k$ is the $k$-th element of the vector $(\cdot)$.

This is a finite-dimensional approximation of Zakai equation (4.8).

Using the approximation, the conditional mean $\widehat{X}(t) = E[X(t)/\mathcal{F}^y_t]$ and its associated error covariance matrix $P(t) = E\{(X(t) - \widehat{X}(t))(X(t) - \widehat{X}(t))'/\mathcal{F}^y_t\}$ can be readily calculated, and this is a crucial advantage of using Gaussian Series. These are given as follows:

$$\widehat{X}(t) \cong \frac{\sum_{i=1}^{N} \psi_i^N(t) m_i}{\sum_{i=1}^{N} \psi_i^N(t)} \tag{4.21}$$

$$P(t) \cong \frac{\sum_{i=1}^{N} \psi_i^N(t)\{B_i + (\widehat{X}(t) - m_i)(\widehat{X}(t) - m_i)'\}}{\sum_{i=1}^{N} \psi_i^N(t)}. \tag{4.22}$$

These two expressions are used in Section 4.5.

## 4.4 Spatial Discretization and Computational Algorithm

In this section we will develop an algorithm based on the proposed Galerkin Scheme using Gaussian series as presented in the previous section. Let $m_0$ be the position of the mean of $p_0$, $\Sigma_a$ denote a cube of finite size $S = 2a$ centered at the origin or centered at $m_0$ such that the inequality given in (4.3) is satisfied and suppose it is partitioned into cells of size $\beta$ so that $k$ (the number of intervals on each axis of the grid) is an integer and $S = k\beta$. The number of cells then equals $k^n$. The parameters of the Gaussian series can be selected such that position vector $m_l$ is defined as $m_l = \beta l + c$, where $l$ is an $n$-vector with integer elements, each in the range 1 to $k$, yielding $N = k^n$ total grid points which equals the number of stochastic ordinary differential equations. The entries of the vector $c$ are chosen so that the grid covers the desired region, i.e., $c_i = -a - \frac{\beta}{2}$ for $i = 1, \ldots, n$. Each position vector $m_l$ in $R^n$ is a point on the grid and it is a center of the cell $D_l$. The volume of each cell is $\alpha = \beta^n$, motivating the only restriction on $\beta$, that $\beta \neq 0$. Observe that $\bigcup_l D_l$ covers the entire region of interest ( by definition of the grid) and $D_l \cap D_s = \emptyset$ for $l \neq s$. Widening the support includes little additional probability mass, but forces the grid to cover a larger region with the same number of points. The densities then appear more sharply peaked, and as predicted the errors increase. On the other hand, reducing the support discards a significant probability mass, even as it improves the grid coverage in the central region. This means that the number of points should be large enough to provide sufficient accuracy, but as small as possible to minimize the computation time.

The covariance matrices $B_l$ are selected to be equal to $\gamma_l I$ where $\gamma_l$ is a positive scalar. Note that as $\gamma_l$ tends to zero, the Gaussian terms each approach a unit impulse function located at the position vector $m_l$.

The matrix $\mathcal{A}_N$ is independent of the observed process $y$, therefore the inverse of $\mathcal{A}_N$ has to be computed only once. The sequence $\{w_i\}$ is linearly independent and for small but positive values of $\gamma_l$ the Gaussian terms are effectively equal to zero everywhere except in a small neighborhood of $m_l$. This guarantees that the matrix $\mathcal{A}_N$ is strictly diagonally dominant [20].

The performance of the Gaussian series filter is judged by the behavior of the measurement residual. The innovation process defined by [8]

$$r(t) = \int_0^t \sigma_0^{-1}(\theta)[dy(\theta) - \hat{h}(\theta)d\theta] \tag{4.23}$$

is a standard Brownian motion if and only if $\hat{h}(t)$ is the best mean-square $\mathcal{F}_t^y$-measurable estimate for $h(X(t))$. Heuristically, there is no information left in $r(t)$, if $\hat{h}(t)$ is an optimal estimate. One can define the measurement residual as

$$\Delta r(t) \equiv r(t+\Delta t) - r(t) = \int_t^{t+\Delta t} \sigma_0^{-1}(\theta)[dy(\theta) - \hat{h}(\theta)d\theta]. \tag{4.24}$$

The measurement residual $\Delta r(t)$ has a Gaussian distribution with mean 0, and variance $\Delta t I$. For the purpose of numerical computations, one can approximate the measurement residual for small $\Delta t$ as

$$\Delta r(t) \cong \sigma_0^{-1}(t)\{\Delta y(t) - \hat{h}(t)\Delta t\}, \tag{4.25}$$

where $\Delta y(t) = y(t+\Delta t) - y(t)$. If the observed behavior of the measurement residual is inconsistent with its theoretical properties, then it must be concluded that the number of Gaussian terms used is inadequate causing the divergence and it is necessary to increase the number of terms in the Gaussian series in order to reduce approximation errors. This requires refinement of the grid size and reinitialization of the entire procedure. In all examples presented here, a sufficiently large number of terms were included to overcome the divergence. In the next section we will introduce the computational steps of the proposed technique.

### 4.4.1 Basic Computational Steps

The major steps in this algorithm may be summarized as follows:

**Step 1:** Generate the random processes $W$ and $V$.

**Step 2:** Given $x(0)$, solve the stochastic differential equation (4.1) using Runge-Kutta method to obtain the values of the observations process at discrete times

$\{y(t_i),\ i = 1, 2, ..., L\}$.

**Step 3:** Solve the system equation (4.20) to obtain $\Upsilon^N(0)$.

**Step 4:** Solve the ordinary stochastic differential equations (4.19) using Runge-Kutta method to obtain $\{\Upsilon^N(t_i),\ i = 1, 2, ..., L\}$.

**Step 5:** Check the measurement residual given by (4.25). If it is not an uncorrelated increments of Gaussian process, increase $N$ (i.e., add more terms to the series) and go to Step 3, otherwise stop.

Based on the above algorithm, we now present numerical examples to illustrate the effectiveness of the proposed filter.

## 4.5 Examples and Simulation Results

**Example 1:**(one-dimensional problem) The approximation theory developed in the pervious sections will be applied to the scalar model which was considered by Beneš [22], and given by

$$\begin{aligned} dX(t) &= f(X(t))dt + bdW(t), \\ dy(t) &= X(t)dt + \sigma_0 dV(t). \end{aligned} \tag{4.26}$$

where $f(x) = tanh(x)$, and $b$ and $\sigma_0$ are constants. Since $f(x)$ is the gradient of the scalar valued function

$$F(x) = log_e[cosh(x)]$$

it satisfies the following Beneš condition.

$$(\nabla F)^2 + \Delta F + \frac{1}{\sigma_0^2} | Hx |^2 = \frac{1}{\sigma_0^2} x^2 + 1.$$

The exact solution of Zakai equation for this case was given by Beneš in the form

$$\Psi(t,x) = c_1(t) exp( \int_0^x tanh(\tau) d\tau - \frac{(x - m(t))^2}{2P(t)}) \tag{4.27}$$

where $\Psi(t,x)$ is unnormalized density function, and $c_1(t)$ is a term independent of $x$ which disappears if $\Psi(t,x)$ is normalized. By substituting $\Psi$ in the Zakai equation one can conclude that the functions $m(t)$ and $P(t)$ satisfy the equations

$$dP(t) = (b^2 - \frac{P^2}{\sigma_0^2}) dt, \tag{4.28}$$

$$dm(t) = -\frac{P}{\sigma_0^2} dy + \frac{mP}{\sigma_0^2} dt. \tag{4.29}$$

Therefore the conditional mean of $X(t)$ is given by the following expression

$$E(X(t)/\mathcal{F}_t^y) \equiv \frac{\int_{-\infty}^{\infty} x \; exp\Big( \int_0^x tanh(\tau) d\tau - \frac{(x-m(t))^2}{2P(t)} \Big) dx}{\int_{-\infty}^{\infty} exp\Big( \int_0^x tanh(\tau) d\tau - \frac{(x-m(t))^2}{2P(t)} \Big) dx}. \tag{4.30}$$

After some elementary calculations, the following compact expression is obtained for the conditional mean:

$$E(X(t)/\mathcal{F}_t^y) = m(t) + P(t) tanh(m(t)). \tag{4.31}$$

For illustrations, we have chosen the following values of the model parameters: $P(0) = 0.001$, $m(0) = 0.0$, $b = 1.0$ and $\sigma_0 = 0.1$. The exact solution as shown in Fig.4.1 was obtained by equations (4.31),(4.29), and (4.28). The corresponding result obtained from the computational technique proposed here as given by equation (4.21) is also shown in Fig.4.1. Examining these figures it is clear that the approximation is very close to the exact solution given by Beneš.

**Example 2:**(two component-vector problem) In the following examples we consider a 2-dimensional system with two different observations processes. Let $X_1$ and $X_2$ be a two dimensional diffusion defined by

$$dX_1(t) = 2X_1(t)tanh(X_1^2(t) - X_2^2(t))dt + bdW_1(t)$$
$$dX_2(t) = -2X_2(t)tanh(X_1^2(t) - X_2^2(t))dt + bdW_2(t). \tag{4.32}$$

and the observation dynamics given by one of the following

**(O1)**

$$\begin{aligned} dy_1(t) &= X_1(t)dt + \sigma_0 dV_1(t) \\ dy_2(t) &= X_2(t)dt + \sigma_0 dV_2(t) \end{aligned} \tag{4.33}$$

**(O2)**

$$dy_1(t) = X_1(t)dt + \sigma_0 dV_1(t) \tag{4.34}$$

Since the drift vector $f(x_1, x_2) = [2x_1 tanh(x_1^2 - x_2^2), -2x_2 tanh(x_1^2 - x_2^2)]'$ is the gradient of scalar valued function

$$F(x_1, x_2) = log_e[cosh(x_1^2 - x_2^2)].$$

it satisfies the following Beneš condition [22].

$$\|\nabla F\|^2 + \Delta F + \frac{1}{\sigma_0^2}\|Hx\|^2 = x'Q_2x + d_2. \tag{4.35}$$

where the constant matrix $Q_2$ is positive definite, and $d_2$ is constant. Again an exact solution of the corresponding Zakai equation can be obtained and it is given by

$$\Psi(t,x) = c_2(t)exp\{F(x) - \frac{1}{2}(x - m(t))'\Sigma^{-1}(x - m(t))\}, \qquad (4.36)$$

where $c_2(t)$ is a term independent of $x$ which disappears if $\Psi(t,x)$ is normalized. This formula determines an unnormalized conditional density in terms of $\Sigma$ and $m$ that solve the equations

$$\frac{d\Sigma}{dt} = b^2 I - \frac{1}{\sigma_0^2}\Gamma\Sigma^2 \qquad (4.37)$$

$$dm(t) = -\frac{1}{\sigma_0^2}\Gamma\Sigma m dt + \frac{1}{\sigma_0^2}\Sigma dy \qquad (4.38)$$

where $\Sigma$ is a diagonal matrix and $\Gamma$ is also a diagonal matrix of nonnegative eigenvalues of $Q_2$ of equation (4.35). For illustrations we have chosen the following values of the model parameters: $b = 1.0$, $\sigma_0 = 0.1$, $\Sigma(0) = 0.1I$, $m_1(0) = 0.5$, and $m_2(0) = -0.5$. The exact solution as shown in (Figs.4.2,4.3) was obtained by equations (4.37),(4.38), and (4.36). The corresponding result obtained from the computational technique proposed here as given by equation (4.21) is also shown in (Figs.4.2,4.3). Examining these figures it is clear that the approximation result is very close to the exact solution given by Beneš.

**Example 3:**(three component-vector problem) To develop further confidence in our computational technique we present the following example. Consider a 3-dimensional system with three different observations processes. Let $X_1$,$X_2$ and $X_3$

be a three dimensional diffusion defined by

$$dX_1(t) = g(X_1(t), X_2(t), X_3(t))\beta_1(X_1(t), X_2(t), X_3(t))dt + bdW_1(t)$$

$$dX_2(t) = g(X_1(t), X_2(t), X_3(t))\beta_2(X_1(t), X_2(t), X_3(t))dt + bdW_2(t)$$

$$dX_3(t) = g(X_1(t), X_2(t), X_3(t))\beta_3(X_1(t), X_2(t), X_3(t))dt + bdW_3(t). \quad (4.39)$$

and the observation dynamics given by one of the following

**(O1)**

$$\begin{aligned} dy_1(t) &= X_1(t)dt + \sigma_0 dV_1(t) \\ dy_2(t) &= X_2(t)dt + \sigma_0 dV_2(t) \\ dy_3(t) &= X_3(t)dt + \sigma_0 dV_3(t) \end{aligned} \quad (4.40)$$

**(O2)**

$$\begin{aligned} dy_1(t) &= X_1(t)dt + \sigma_0 dV_1(t) \\ dy_2(t) &= X_2(t)dt + \sigma_0 dV_2(t) \end{aligned} \quad (4.41)$$

**(O3)**

$$dy_1(t) = X_1(t)dt + \sigma_0 dV_1(t) \quad (4.42)$$

where

$$\begin{aligned} g(x_1, x_2, x_3) &\equiv tanh(2x_1^2 - x_2^2 - x_3^2 + x_1x_2 + x_1x_3 + x_2x_3), \\ \beta_1(x_1, x_2, x_3) &\equiv 4x_1 + x_2 + x_3, \end{aligned}$$

$$\beta_2(x_1, x_2, x_3) \equiv x_1 - 2x_2 + x_3,$$

$$\beta_3(x_1, x_2, x_3) \equiv x_1 + x_2 - 2x_3.$$

Since the drift vector $f(x_1, x_2, x_3) = [g\beta_1, g\beta_2, g\beta_3]'$ is the gradient of scalar valued function

$$F(x_1, x_2, x_3) = log_e[cosh(2x_1^2 - x_2^2 - x_3^2 + x_1x_2 + x_1x_3 + x_2x_3)]$$

it satisfies the following Beneš condition [22],

$$\|\nabla F\|^2 + \Delta F + \frac{1}{\sigma_0^2}\|Hx\|^2 = x'Q_3x + q_3'x + d_3, \qquad (4.43)$$

where the constant matrix $Q_3$ is positive definite, and $q_3$, $d_3$ are constants. Again an exact solution of the corresponding Zakai equation can be obtained and it is given by

$$\Psi(t, x) = c_3(t)exp\{F(x) - \frac{1}{2}(x - m(t))'\Sigma^{-1}(x - m(t))\}. \qquad (4.44)$$

where $c_3(t)$ is a term independent of $x$ which disappears if $\Psi(t, x)$ is normalized. This formula determines an unnormalized conditional density in terms of $\Sigma$ and $m$ that solve the equations

$$\frac{d\Sigma}{dt} = b^2I - \frac{1}{\sigma_0^2}\Gamma\Sigma^2 \qquad (4.45)$$

$$dm(t) = -\frac{1}{\sigma_0^2}\Gamma\Sigma m dt - \frac{1}{\sigma_0^2}\Sigma q_3 dt + \frac{1}{\sigma_0^2}\Sigma dy \qquad (4.46)$$

where $\Sigma$ is a diagonal matrix and $\Gamma$ is also a diagonal matrix of nonnegative eigenvalues of $Q_3$ of equation (4.43). For illustrations we have chosen the following values of the model parameters: $b = 1.0$, $\sigma_0 = 0.015$, $\Sigma(0) = 0.1I$, $m_1(0) = 0.2$, $m_2(0) = 0.1$, and $m_3(0) = -0.1$. The exact solution as shown in (Figs.4.4,4.5,4.6) was obtained by equations (4.45),(4.46), and (4.44). The corresponding result obtained from the computational technique proposed here as given by equation (4.21) is also shown in (Figs.4.4,4.5,4.6). Examining these figures it is clear that the approximation result is very close to the exact solution given by Beneš.

## 4.6 Summary and Conclusion

The Galerkin method using Gaussian series approximation has been introduced and proposed as the means whereby the solution of Zakai equation can be implemented in a straightforward manner. As long as the measurement residual is consistent with its theoretical properties as stated in the previous section, one can be satisfied that the proposed method performs well. If an inconsistency occurs, it may be necessary to decrease the value of the parameter $\gamma_l$ associated with the covariance matrix $B_l = \gamma_l I$, and add more terms to the series. On VM/SP Conversational Monitor System (CMS) Computers, the computation time needed to find the inverse of the matrix $\mathcal{A}_N$ in example 1 is 2 min CPU time, and in example 2 is 13 min CPU time. The computation time per estimate in example 1 is 0.7 second CPU time, and in example 2 is 4.3 second CPU time. The proposed technique offers the computational advantage that the matrix $\mathcal{A}_N$ can be computed off-line and stored in the computer. In developing the computer code, no special attention was given to minimize the CPU time.

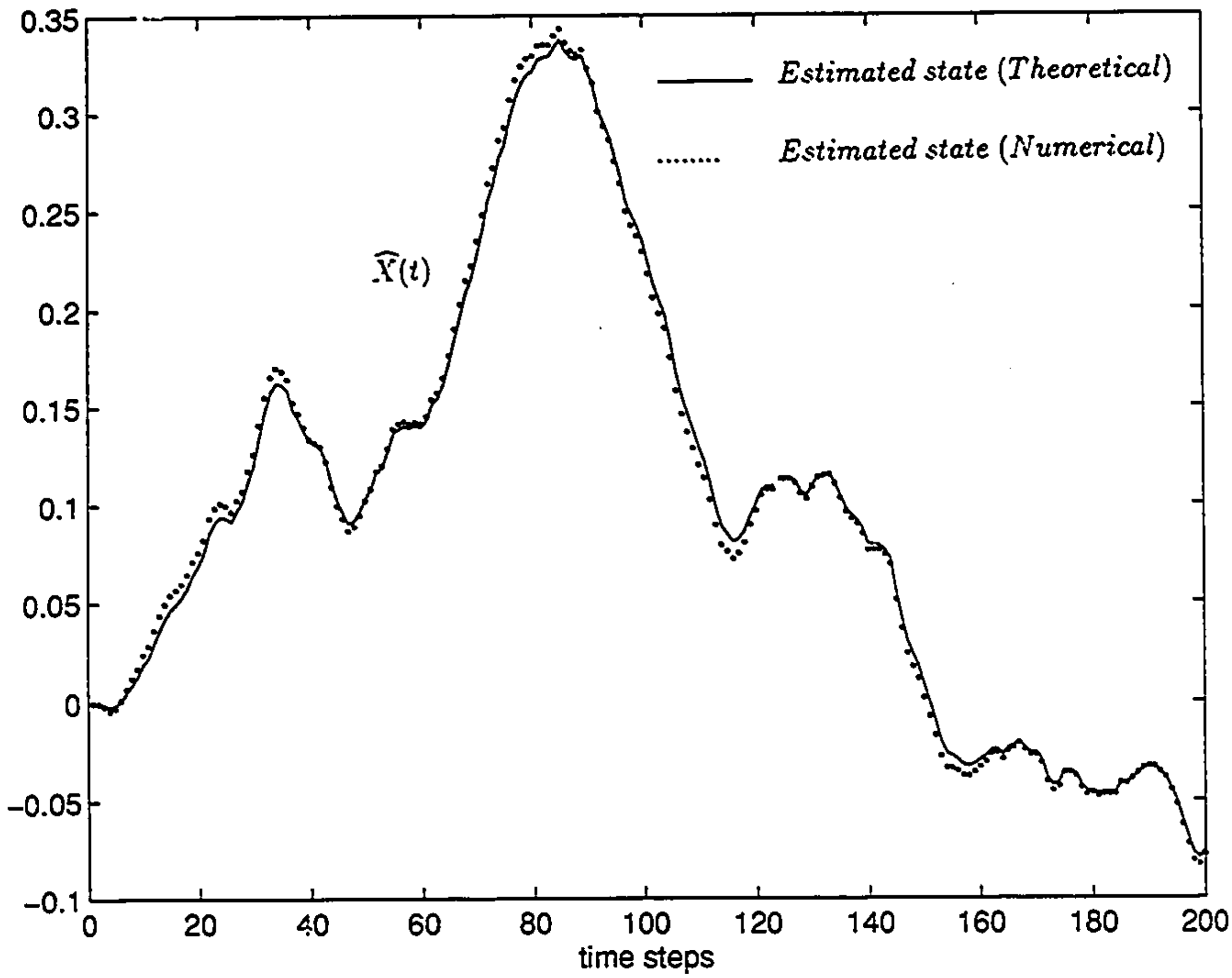


Figure 4.1: (Example 1) Estimated state using the exact solution given by Benes and the approximate solution obtained by the proposed scheme.

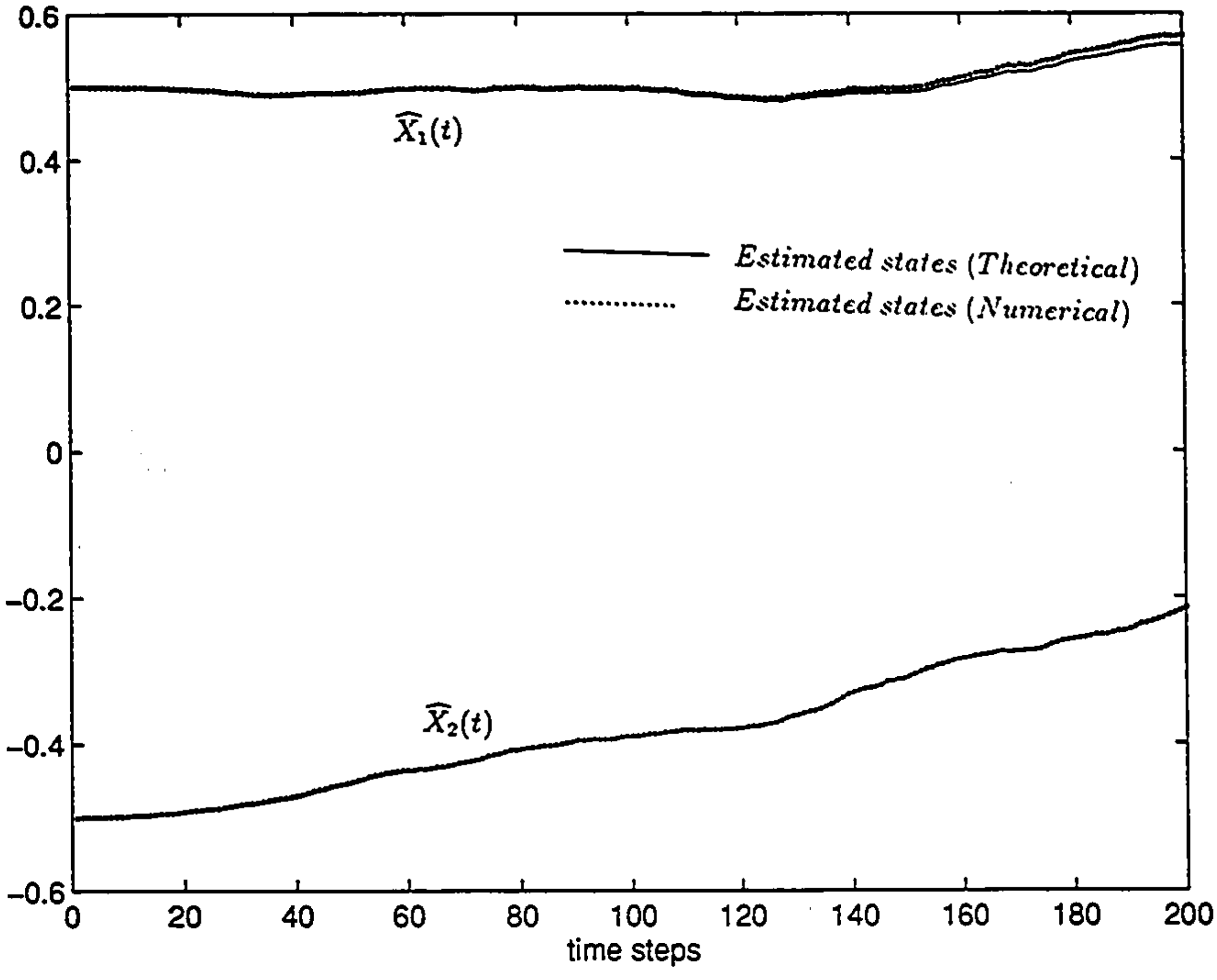


Figure 4.2: (Example 2,O1) Estimated state using the exact solution given by Benes and the approximate solution obtained by the proposed scheme.

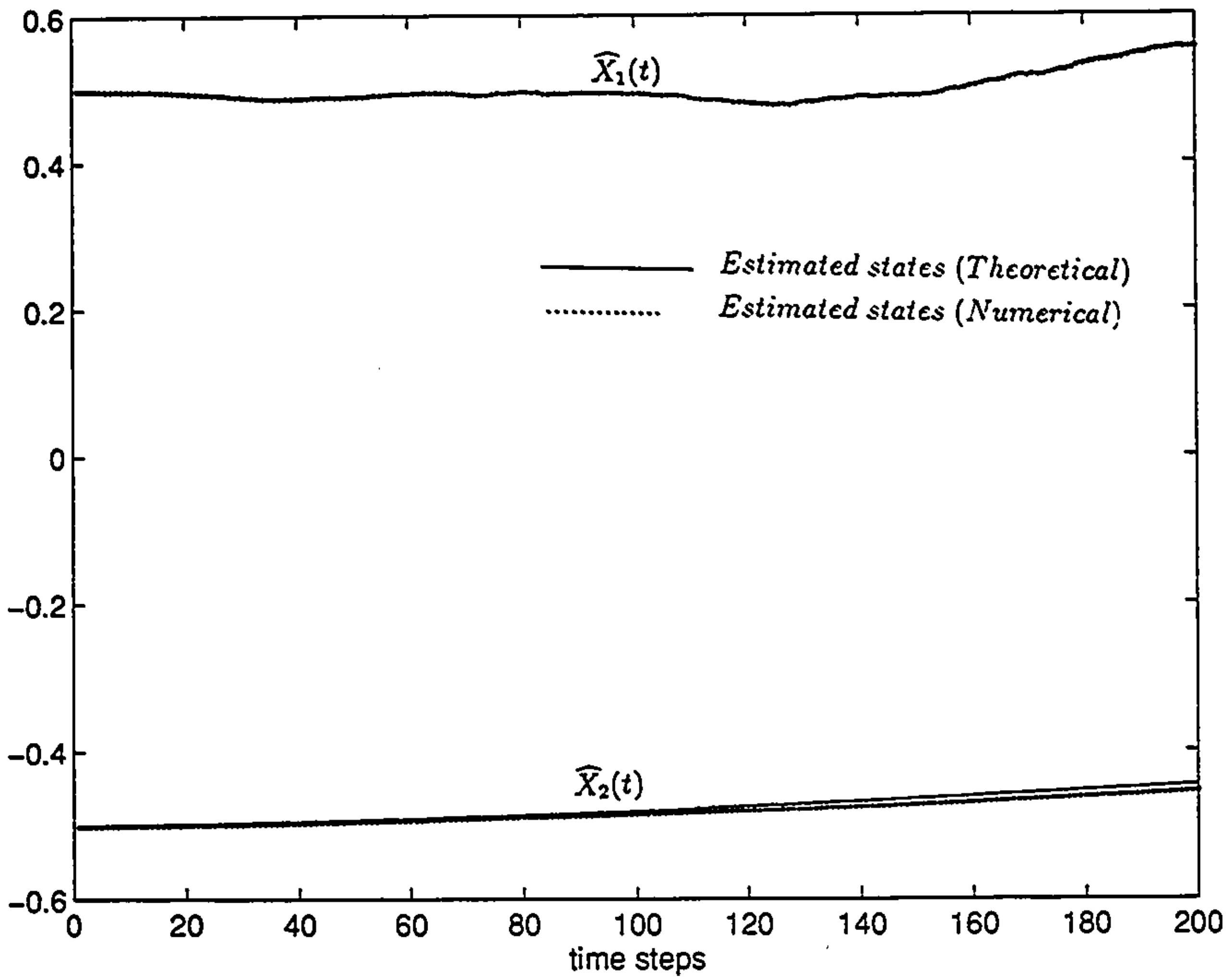


Figure 4.3: (Example 2,O2) Estimated state using the exact solution given by Benes and the approximate solution obtained by the proposed scheme.

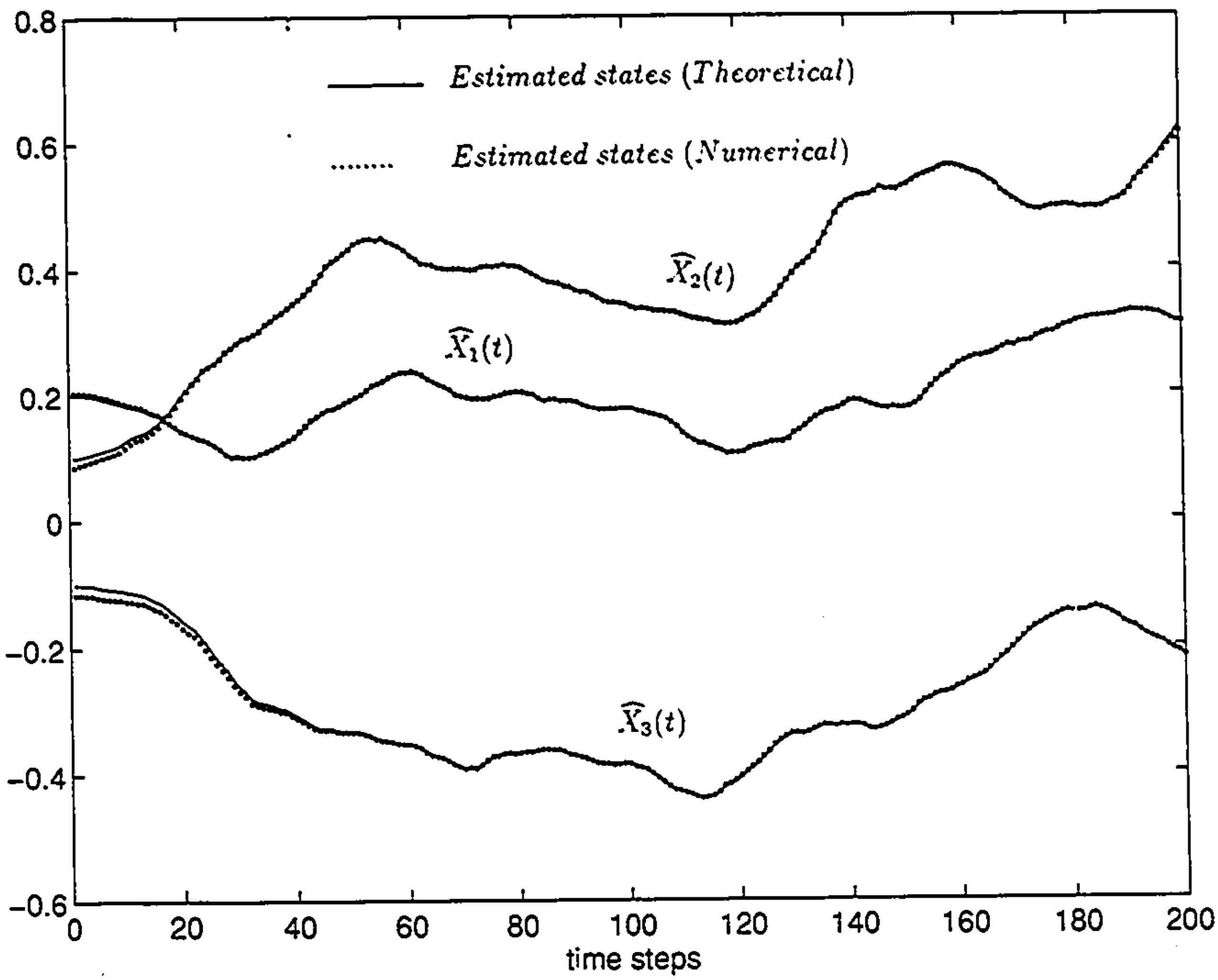


Figure 4.4: (Example 3,O1) Estimated state using the exact solution given by Benes and the approximate solution obtained by the proposed scheme.

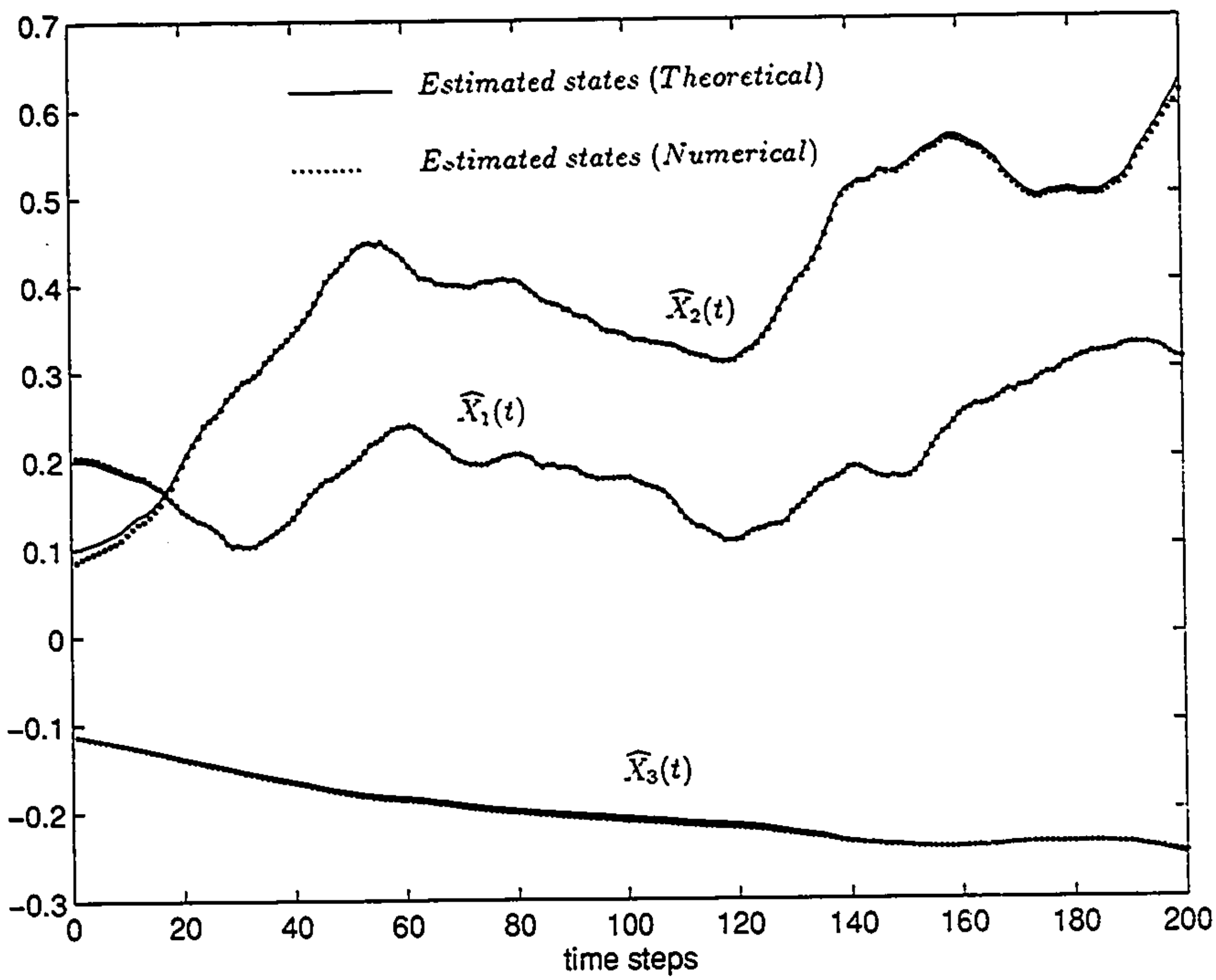


Figure 4.5: (Example 3,O2) Estimated state using the exact solution given by Benes and the approximate solution obtained by the proposed scheme.

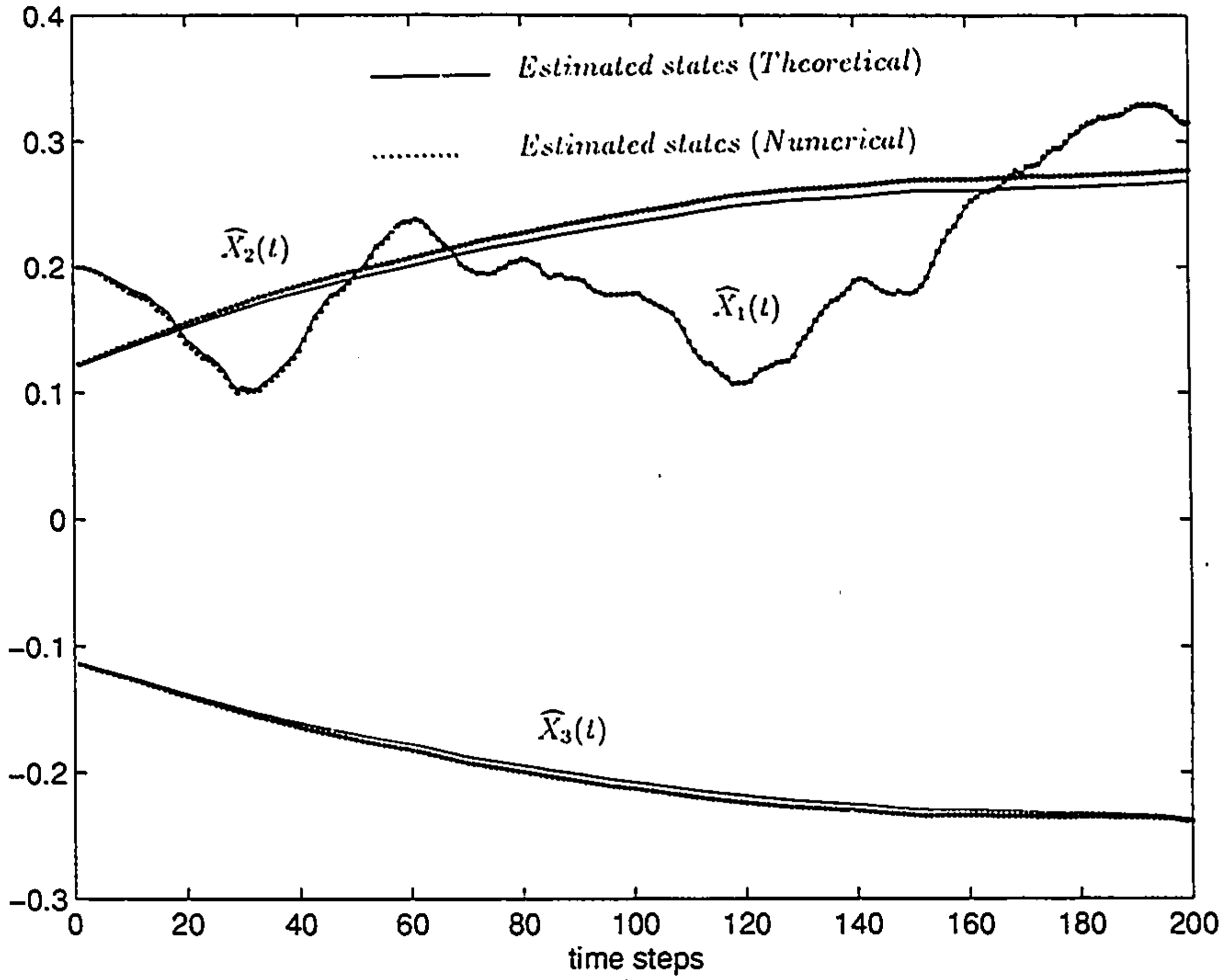


Figure 4.6: (Example 3,O3) Estimated state using the exact solution given by Benes and the approximate solution obtained by the proposed scheme.

# Chapter 5

# NONLINEAR FILTERING (ZAKAI EQUATION CORRELATED CASE)

## 5.1 Introduction

In many filtering problems of practical interest, the diffusion processes observed in correlated noises (the observation noise and the system noise are correlated). In this Chapter, the solution of Zakai equation arising from diffusion processes observed in correlated noises is to be our ideal reference and that we must be able to rely on an efficient and accurate algorithm to solve it. One can write the diffusion processes observed in correlated noises by the following stochastic differential equations

$$\begin{aligned} dX(t) &= f(X(t))dt + b(X(t))dW(t), \quad X(0) = x(0), \\ dy(t) &= h(X(t))dt + \sigma_0(t)dV(t) + \rho(t)dW(t), \quad y(0) = 0, \end{aligned} \tag{5.1}$$

where the state $X(t) \in R^n$ is the unobserved process to be estimated, $y(t) \in R^m$ is the observation process, the processes $W$ and $V$ are independent standard Wiener processes with values in $R^p$ and $R^q$, respectively, $f : R^n \to R^n$ and $b : R^n \to R^{n\times p}$ and $h : R^n \to R^m$ and $\sigma_0$ is $m \times q$ matrix and $\rho$ is $m \times p$ matrix and $x(0)$ is an $\mathcal{F}_0$ measurable vector independent of the Wiener processes $W$ and $V$, with probability density $p_0(\cdot)$. The only new term is the deterministic matrix $\rho$ in the observation process, creating a correlation between the system noise and the observation noise. Throughout this Chapter, it is assumed that the coefficients, $f$, $b$, and $h$, are globally Lipschitz continuous functions, so that the stochastic differential system (5.1) has a unique strong solution. One can verfiy that the Zakai equation arises from the previous model is given by [32,38,74,75,80]

$$\begin{aligned} d\Psi(t,x) &= A^{*}\Psi(t,x)dt + \Psi(t,x)h(x)\cdot\Gamma_0^{-1}dy(t) \\ &- \Psi(t,x)C(t,x)\cdot\Gamma_0^{-1}dy(t) - \rho(t)b'(x)\nabla\Psi(t,x)\cdot\Gamma_0^{-1}dy(t), \\ \Psi(0,x) &= p_0(x), \end{aligned} \tag{5.2}$$

where the vector $C = [div\ c_1, div\ c_2, ..., div\ c_m]'$, $c_j = b\rho_j'$, where $\rho_j$ is the $j$th row of $\rho$. Since by using Galerkin technique the solution of Zakai equation (arises from diffusion processes observed in uncorrelated noises) is successfully approximated by means of a sequence of nonstandard basis functions given by a prarameterized family of Gaussian densities, it seems natural to extend this technique to Zakai equation arises from diffusion processes observed in correlated noises.

## 5.2 Galerkin Approximation

Constructing the solution of Zakai equation using the orthonormal Gaussian series as a basis demands excessive CPU time. Therefore, it is reasonable and very easy for the purpose of numerical calculations to implement the approximate solution of Zakai equation (equation (5.2)) using the nonorthogonal Gaussian series as a basis functions. Hence, using the Galerkin method based on the nonorthogonal Gaussian sequence, one can approximate the solution of equation (5.2) in the form

$$\Psi^N(t,x) = \sum_{i=1}^{N} \psi_i^N(t) w_i(x), \tag{5.3}$$

where $\{\psi_i^N\}$ are the fourier coefficients to be chosen as follows, $\{w_i\}$ are the nonorthogonal sequence as given by equation (4.4). For approximate solutions, equation (5.3) is usually a finite summation. Substituting equation (5.3) into equation (5.2), we get

$$\begin{aligned} \sum_{i=1}^{N} d\psi_i^N(t) w_i &= \sum_{i=1}^{N} \psi_i^N(t) A^* w_i dt + \sum_{i=1}^{N} \psi_i^N(t) w_i h \cdot \Gamma_0^{-1} dy(t) \\ &- \sum_{i=1}^{N} \psi_i^N(t) w_i C \cdot \Gamma_0^{-1} dy(t) - \sum_{i=1}^{N} \psi_i^N(t) \rho b' \nabla w_i \cdot \Gamma_0^{-1} dy(t), \quad (5.4) \\ \sum_{i=1}^{N} \psi_i^N(t) w_i &= p_0. \quad (5.5) \end{aligned}$$

For the purpose of numerical calculations it is more convenient to formulate equation (5.4) in terms of Stratonovich sense rather than Ito sense. Its transformation to the Stratonovich form is done by adding the Wong-Zakai correction to the drift term [87]. Thus, equation (5.4) takes the following Stratonovich form

$$\begin{aligned}\sum_{i=1}^{N} d\psi_i^N(t) w_i &= \sum_{i=1}^{N} \psi_i^N(t) A^* w_i dt + \sum_{i=1}^{N} \psi_i^N(t) w_i h \cdot \Gamma_0^{-1} dy(t) \\ &- \sum_{i=1}^{N} \psi_i^N(t) w_i C \cdot \Gamma_0^{-1} dy(t) - \sum_{i=1}^{N} \psi_i^N(t) w_i \rho b' \alpha_i \cdot \Gamma_0^{-1} dy(t) \\ &- \frac{1}{2} \sum_{i=1}^{N} \psi_i^N(t) w_i (h - C - \rho b' \alpha_i) \cdot \Gamma_0^{-1} (h - C - \rho b' \alpha_i) dt, \quad (5.6)\end{aligned}$$

where $\alpha_i = [\frac{\partial J_i}{\partial x_1}, \frac{\partial J_i}{\partial x_2}, ..., \frac{\partial J_i}{\partial x_n}]'$, $J_i = \frac{-1}{2}(x - m_i)' B_i^{-1} (x - m_i)$, where again $\{m_i\}$ and $\{B_i\}$ are the parameters of the Gaussian series. One can choose $\{m_i\}$ and $\{B_i\}$ by using the same approach given in Section 4.4 of Chapter 4. Now define a set of weighting or test functions $\{v_i\}$ in the range of $A^*$ (for Galerkin case $v_i = w_i$) and take the inner product of equation (5.6) with $w_j$, we obtain

$$\begin{aligned}\sum_{i=1}^{N} d\psi_i^N(t)(w_i, w_j) &= \sum_{i=1}^{N} \psi_i^N(t) \Big( (A^* w_i, w_j) dt \\ &- \frac{1}{2}((h - C - \rho b' \alpha_i) \cdot \Gamma_0^{-1} (h - C - \rho b' \alpha_i) w_i, w_j) dt \\ &+ ((h - C - \rho b' \alpha_i) w_i, w_j) \cdot \Gamma_0^{-1} dy(t) \Big), \quad 1 \le j \le N, \quad (5.7)\end{aligned}$$

This is a family of finite dimensional (N) stochastic differential equations for $\Upsilon \equiv [\psi_1^N, \psi_2^N, ..., \psi_N^N]'$ given by

$$\begin{aligned}\sum_{i=1}^{N} a_{ji} d\psi_i^N(t) = \Big( \sum_{i=1}^{N} b_{ji} \psi_i^N(t) &- \frac{1}{2} \sum_{k=1}^{m} \sum_{i=1}^{N} d_{ji}^k \psi_i^N(t) \Big) dt \\ &+ \sum_{k=1}^{m} \sum_{i=1}^{N} g_{ji}^k \psi_i^N(t) (\Gamma_0^{-1} dy(t))_k \\ \sum_{i=1}^{N} a_{ji} \psi_i^N(0) &= p_{0j}, \; j = 1, 2, ..., N, \quad (5.8)\end{aligned}$$

where

$$a_{ji} = \int_{R^n} w_j w_i dx \equiv (w_j, w_i). \tag{5.9}$$

$$b_{ji} = \int_{R^n} w_j A^* w_i dx \equiv (w_j, A^* w_i). \tag{5.10}$$

$$g_{ji}^k = \int_{R^n} (h - C - \rho b' \alpha_i)_k w_i w_j dx. \tag{5.11}$$

$$d_{ji}^k = \int_{R^n} (h - C - \rho b' \alpha_i)_k w_i w_j (\Gamma_0^{-1} (h - C - \rho b' \alpha_i))_k dx. \tag{5.12}$$

$$d_{ji} = \sum_{k=1}^{m} d_{ji}^k. \tag{5.13}$$

$$p_{0j} = \int_{R^n} p_0 w_j dx. \tag{5.14}$$

In the matrix notation system (5.8) is equivalent to

$$\mathcal{A}_N d\Upsilon^N = (\mathcal{B}_N - \frac{1}{2}\mathcal{D}_N)\Upsilon^N dt + \sum_{k=1}^{m} \mathcal{G}_N^k \Upsilon^N (\Gamma_0^{-1} dy)_k . \tag{5.15}$$

$$\mathcal{A}_N \Upsilon^N(0) = p_0^N. \tag{5.16}$$

where $\mathcal{A}_N = [a_{ji}]$, $\mathcal{B}_N = [b_{ji}]$, $\mathcal{D}_N = [d_{ji}]$, $\mathcal{G}_N^k = [g_{ji}^k]$, and $p_0^N = [p_{01}, p_{02}, \ldots\ldots, p_{0N}]'$, and $(\cdot)_k$ is the $k$th element of the vector $(\cdot)$.

This is a finite-dimensional approximation of Zakai equation (5.2). The performance of the proposed technique is judged by the behavior of the measurement residual. The innovation process in this case is defined by

$$r(t) = \int_0^t [dy(\theta) - \hat{h}(\theta) d\theta] \tag{5.17}$$

is a Brownian motion if and only if $\hat{h}(t)$ is the best mean square $\mathcal{F}_t^y$-measurable estimate for $h(X(t))$. One can define the measurement residual as

$$\Delta r(t) \equiv r(t + \Delta t) - r(t) = \int_t^{t+\Delta t} [dy(\theta) - \hat{h}(\theta)d\theta]. \tag{5.18}$$

The measurement residual $\Delta r(t)$ has a Gaussian distribution with mean 0, and variance $(\sigma_0(t)\sigma_0'(t) + \rho(t)\rho'(t))\Delta t$. For the purpose of numerical computations, one can approximate the measurement residual for small $\Delta t$ as

$$\Delta r(t) \cong \Delta y(t) - \hat{h}(t)\Delta t, \tag{5.19}$$

where $\Delta y(t) = y(t + \Delta t) - y(t)$. In the light of the algorithm which given in the previous Chapter, one can write easily the computational steps of the proposed technique which solves the Zakai equation arises from the diffusion processes observed in correlated noises.

### 5.2.1 Basic Computational Steps

The major steps in this algorithm may be summarized as follows:

**Step 1:** Generate the random processes $W$ and $V$.

**Step 2:** Given $x(0)$, solve the stochastic differential equation (5.1) using Runge-Kutta method to obtain the values of the observations process at discrete times $\{y(t_i),\ i = 1, 2, ..., L\}$.

**Step 3:** Solve the system equation (5.16) to obtain $\Upsilon^N(0)$.

**Step 4:** Solve the ordinary stochastic differential equations (5.15) using Runge-

Kutta method to obtain $\{\Upsilon^N(t_i),\ i = 1, 2, \ldots L\}$.

**Step 5:** Check the measurement residual given by (5.19). If it is not an uncorrelated increments of Gaussian process, increase $N$ (i.e., add more terms to the series) and go to Step 3, otherwise stop.

Based on the above algorithm, we present next some numerical examples for which the exact analytical solutions are available to illustrate the effectiveness of the proposed technique.

## 5.3 Examples and Simulation Results

Consider the following correlated filtering problem

$$dX(t) = A(t)X(t)dt + b(t)dW(t), \quad X(0) = x(0). \tag{5.20}$$

where $A$ is $n \times n$ matrix, $b$ is a $n \times p$ matrix, and $W$ is an $p$-dimensional Wiener process. Suppose that an observed $m$-dimensional process $y$ is described by the following Ito equation

$$dy(t) = H(t)X(t)dt + \sigma_0(t)dV(t) + \rho(t)dW(t), \quad y(0) = 0. \tag{5.21}$$

where $H$ is an $m \times n$ matrix, $\sigma_0$ is a $m \times q$ matrix, $\rho$ is $m \times p$ matrix, and $V$ is a $m$-dimensional Wiener process. If the random variable $x(0)$ is normally distributed or constant, then $X(t)$ and $y(t)$ are Gaussian processes. Therefore, all conditional distributions are normal distributions. In other words, the exact analytical solution is available. The exact analytical optimal estimator for the system (5.20-5.21) can be computed by using one of the following methods.

**The Equivalent Method.** By This method [70] the filtering problem given by equations (5.20-5.21) is converted into an equivalent problem with no correlation between the process and measurement noises. To do this, add zero term to the right hand side of equation (5.20), in the form

$$\begin{aligned} dX(t) &= A(t)X(t)dt + b(t)dW(t) \\ &+ D(t)[dy(t) - H(t)X(t)dt - \sigma_0(t)dV(t) - \rho(t)dW(t)], \end{aligned} \tag{5.22}$$

where the observed process $y$ considered here as an additional deterministic input. we want to choose the matrix $D$ such that the noise for the new system equation is independent of the measurement noise, i.e.,

$$E[((b - D\rho)W(t) - D\sigma_0 V(t))(\sigma_0 V(t) + \rho W(t))'] = 0. \tag{5.23}$$

Multiplying through, we see that the value for $D$ that solves this equation is

$$D = b\rho' R^{-1}. \tag{5.24}$$

where $R = [\rho\rho' + \sigma_0\sigma_0']$. Then, equation (5.22) and equation (5.21) represent the standard linear filtering problem. In particular, using Kalman-Bucy theory for linear systems, we get

$$\begin{aligned} d\widehat{X}(t) &= A(t)\widehat{X}(t)dt + K_m(t)\Big(dy(t) - H(t)\widehat{X}(t)dt\Big). \\ \widehat{X}(0) &= Ex(0). \end{aligned} \tag{5.25}$$

where $K_m$ is the modified Kalman gain given by

$$K_m = (PH' + b\rho')R^{-1}. \tag{5.26}$$

The associated error covariance equation is given by

$$\frac{dP}{dt} = AP + PA' + bb' - K_m R K_m'. \tag{5.27}$$

with initial condition $P(0) = E((x(0) - Ex(0))(x(0) - Ex(0)')$.

<u>**Linear Quadratic Regulator (LQR) Method.**</u> Reference [Ahmed, Li 10] takes us through the necessary steps to find the exact analytical optimal estimator for the system (5.20-5.21). One demands that the estimator be linear and driven by the observed process $y$ in the form

$$\begin{aligned} d\widehat{X}(t) &= \phi_1(t)\widehat{X}(t)dt + \phi_2(t)dy(t) \\ &= \phi_1(t)\widehat{X}(t)dt + \phi_2(t)H\widehat{X}(t)dt + \phi_2(t)\sigma_0 dV + \phi_2(t)\rho dW \end{aligned} \tag{5.28}$$

where $\phi_1$ and $\phi_2$ are to be chosen so as to obtain an unbiased $(E\widehat{X}(t) = EX(t))$ and minimum variance filter. Then comparing the above equation with the state equation (5.20) one has

$$\begin{aligned} d(X(t) - \widehat{X}(t)) &= (A - \phi_2 H)(X(t) - \widehat{X}(t))dt \\ &+ (A - \phi_2 H - \phi_1)\widehat{X}(t)dt + (b - \phi_2\rho)dW - \phi_2\sigma_0 dV. \end{aligned} \tag{5.29}$$

Hence for the unbiased estimate $\phi_1$ must equal $A - \phi_2 H$ leading to the estimator equation:

$$d\widehat{X}(t) = (A - \phi_2 H)\widehat{X}(t)dt + \phi_2 dy(t)$$

$$= (A - \phi_2 H)\widehat{X}(t)dt + \phi_2 H X(t)dt + \phi_2\sigma_0 dV + \phi_2\rho dW. \quad (5.30)$$

Subtracting this from the state equation one obtains

$$d(X(t) - \widehat{X}(t)) = (A - \phi_2 H)(X(t) - \widehat{X}(t))dt + (b - \phi_2\rho)dW - \phi_2\sigma_0 dV, \quad (5.31)$$

and the problem reduces to find $\phi_2$ that minimize $E(X(t) - \widehat{X}(t), \zeta)^2$ for every $\zeta \in R^n$. This is equivalent to the optimal control on the space of matrix- functions $\{\phi_2\}$.

$$\begin{aligned} \frac{dP}{dt} = P(A - \phi_2 H) \quad &+ \quad (A - \phi_2 H)'P + bb' \\ &- \quad b\rho'\phi_2' - \phi_2\rho b' + \phi_2(\rho\rho' + \sigma_0\sigma_0')\phi_2'. \\ J(\phi_2) \quad &= \quad \int_0^T tr(PS)dt = min. \end{aligned} \quad (5.32)$$

For any positive definite symmetric matrix $S$. Again, by [theorem 8, 10], the following must hold in order for $\phi_2^0$ to be optimal:

(i)

$$\begin{aligned} \frac{dL^0}{dt} &= (A - \phi_2^0 H)'L^0 + L^0(A - \phi_2^0 H) = S. \\ L^0(T) &= 0; \end{aligned} \quad (5.33)$$

(ii)

$$\begin{aligned} \frac{dP^0}{dt} &= P^0(A - \phi_2^0 H) + (A - \phi_2^0 H)'P^0 + bb' \\ &- b\rho'(\phi_2')^0 - \phi_2^0\rho b' + \phi_2^0(\rho\rho' + \sigma_0\sigma_0')(\phi_2')^0 \\ P^0(0) &= P(0); \end{aligned} \quad (5.34)$$

and (iii)

$$\int_0^T tr\{(\phi_2 - \phi_2^0)[HP^0 + (b\rho')' - (\rho\rho' + \sigma_0\sigma_0')(\phi_2^0)']L^0\}dt \geq 0 \qquad (5.35)$$

for all $\phi_2 \in \mathcal{E} \subset L_\infty([0,T], \mathcal{L}(R^m, R^n))$. By lifting the constraints we can obtain the correlated Kalman-Bucy filter equations. Indeed, suppose $\mathcal{E} = L_\infty([0,T], \mathcal{L}(R^m, R^n))$, then it follows from the expression (5.35) that the term within the middle bracket must vanish giving

$$HP^0 + (b\rho')' = (\rho\rho' + \sigma_0\sigma_0')(\phi_2^0)'. \qquad (5.36)$$

If $(\rho\rho' + \sigma_0\sigma_0')$ is invertible, then $\phi_2^0$ becomes

$$\phi_2^0 = (P^0H' + b\rho')(\rho\rho' + \sigma_0\sigma_0')^{-1}. \qquad (5.37)$$

This is the same result we had earlier for the correlated.

**Example 1:**(one-dimensional problem) The approximation theory developed in the previous sections will be applied to the following scalar model

$$\begin{aligned} dX(t) &= FX(t)dt + bdW(t), \quad X(0) = x(0), \\ dy(t) &= X(t)dt + \sigma_0 dV(t) + \rho dW(t), \end{aligned} \qquad (5.38)$$

where $F$, $b$, $\sigma_0$, and $\rho$ are constants and the random variable $x(0)$ is normally distributed with mean $m_0$ and variance $P_0$. For illustrations we have chosen the following values of the model parameters: $m_0 = -0.2$, $P_0 = 0.1$, $F = 1.0$, $b = 1.0$, $\sigma_0 = 0.1$, and $\rho = 0.01$. The exact solution as shown in Fig.5.1 was obtained by equations (5.25) and (5.26) and (5.27). The corresponding result of the computational technique proposed here was obtained by equation (5.15), (5.3) and (4.20) is also shown in Fig.5.1. Examining these figures it is clear that the approximation is very close to the exact solution, providing confidence in our procedure.

**Example 2:**(two component-vector problem) To develop further confidence in our computational technique we present the following example. Let $X_1$ and $X_2$ be a two dimensional diffusion defined by

$$\begin{aligned} dX_1(t) &= X_2(t)dt + bdW_1(t) \\ dX_2(t) &= -2(X_1(t) + X_2(t))dt + bdW_2(t), \end{aligned} \tag{5.39}$$

and the observation dynamics given by

$$dy(t) = X_2(t)dt + \sigma_0 dV(t) + \rho_1 dW_1(t) + \rho_2 dW_2(t), \tag{5.40}$$

where the processes $W_1$, $W_2$, and $V$ are independent standard Brownian motions, $b$, $\sigma_0$, $\rho_1$, and $\rho_2$ are constants, and the random $x(0)$ is normally distributed with mean $m_0$ and variance $P_0$, and independent of the processes $W_1$, $W_2$, and $V$.

For illustrations we chosen the following values of the model parameters: $b = 1.0$, $\sigma_0 = 0.1$, $\rho_1 = 0.1$, $\rho_2 = 0.1$, $m_0 = [0.2 \;\; 0.2]'$, and $P_0 = diag[0.01]$. The exact solution as shown in Fig.5.2 was obtained by equations (5.25), (5.26) and (5.27). The corresponding result of the computational technique proposed here was obtained by equations (5.15), (5.3) and (4.20) is also shown in Fig.5.2. Examining theses figures it is clear that the approximation result is very close to the exact solution.

## 5.4 Summary and Conclusion

The Galerkin method using Gaussian series approximation has been introduced and proposed as means whereby solution of Zakai equation which arises from the correlated filtering model can be implemented in a straightforward manner. Finally, we have concluded from the results of Chapter 4 and Chapter 5, that the Galerkin method using a parametrized family of Gaussian densities as a basis functions represents a computable and accurate approximation for the unnormalized conditional (probability) density of a diffusion processes (correlated and uncorrelated) observed in continuous time.

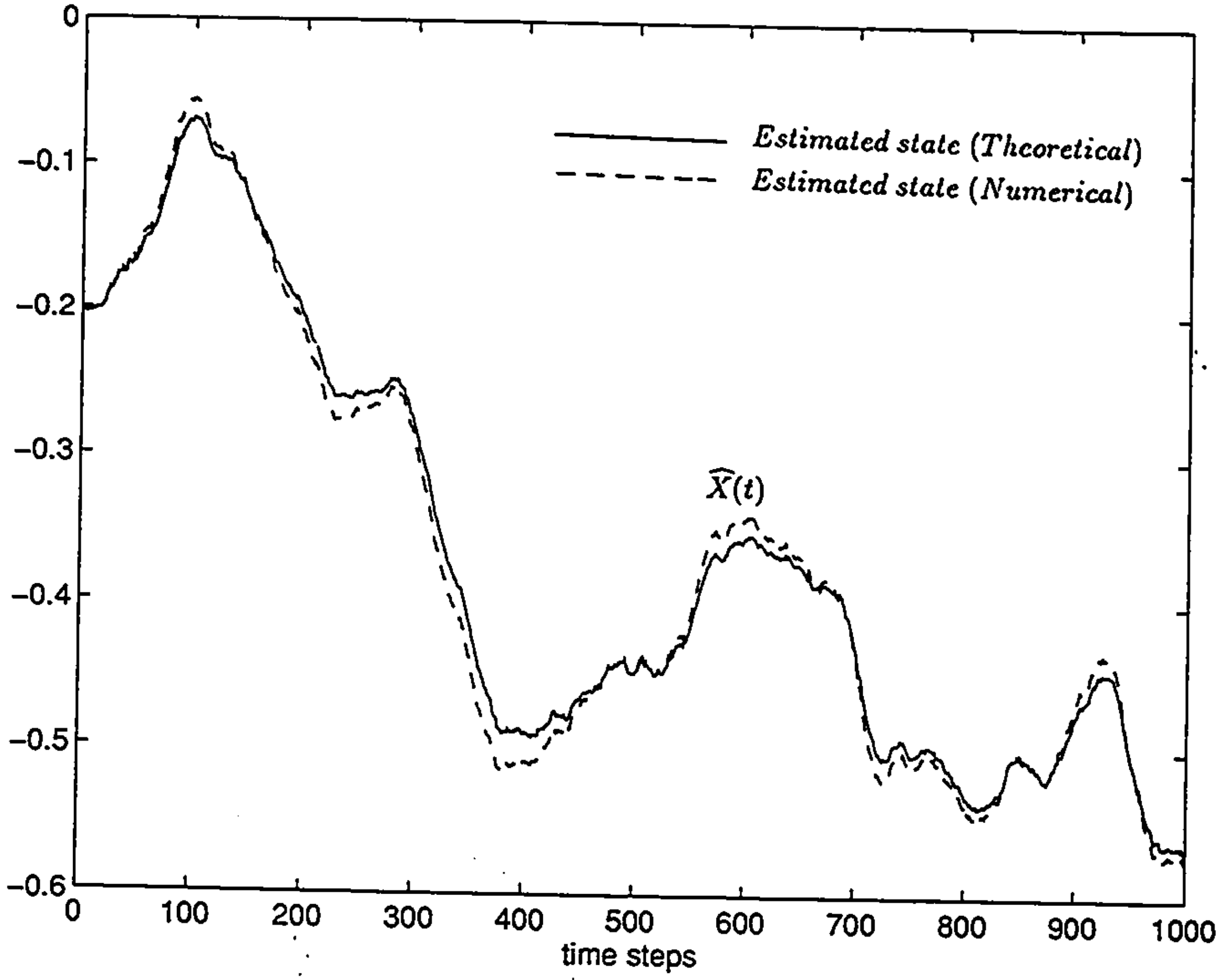


Figure 5.1: (Example 1) Estimated state using the exact solution given by equations (5.25-5.27) and the approximate solution obtained by the proposed scheme.

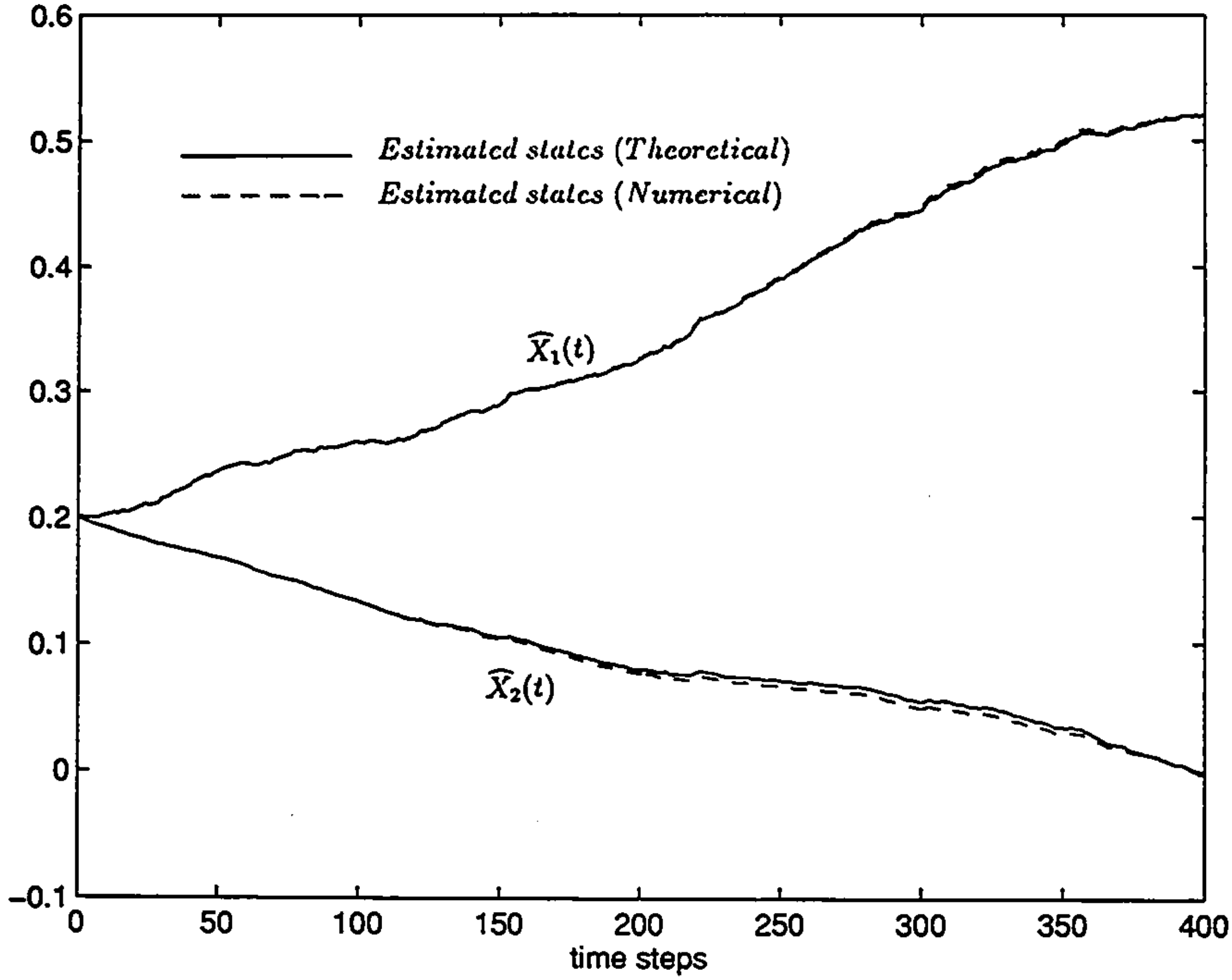


Figure 5.2: (Example 2) Estimated state using the exact solution given by equations (5.25-5.27) and the approximate solution obtained by the proposed scheme.

# Chapter 6

# IDENTIFICATION OF LINEAR STOCHASTIC SYSTEMS BASED ON PARTIAL INFORMATION

## 6.1 Introduction

An important and essential aspect of modelling any physical systems is the identification of the parameters in the model equation. These model equations are usually inferred on the basis of fundamental physical laws and some idealizing assumptions but contains certain parameters which are completely unknown because of the lack of precise understanding of the system, or partly known because of poor measurement data. The analyst must determine these parameters on the basis of the available field data.

In this Chapter we consider the identification problem for a system of partially observed linear stochastic differential equations. Utilizing the results of Chapters 1 and 2, we formulate this identification problem as a deterministic control problem. We prove the equivalence of the two problems. In Section 6.2, we present the identification problem and the necessary assumptions which have been used in the sequel. In Section 6.3, we formulate the identification problem as a deterministic control problem. In Section 6.4, we present the measurement of the autocovariance function of the process $\epsilon$. In Section 6.5, we developed iterative procedure for computing the estimated parameters along with some numerical examples to illustrate the results of this Chapter.

## 6.2 Identification Problem (IP)

To introduce the identification problem we shall need some basic notations. For each pair of integers $n, m \in N$, let $M(n \times m)$ denote the space of $n \times m$ matrices with entries all real and let $M^+(m \times m)$, a subset of $M(m \times m)$, denote the class of all positive definite matrices. Define

$$M_0(m \times q) \equiv \{\sigma_0 \in M(m \times q) : \sigma_0\sigma_0' \in M^+(m \times m)\},$$

and

$$\Sigma \equiv M(n \times n) \times M(n \times p) \times M(m \times n) \times M_0(m \times q).$$

We shall denote our identification problem as (IP) which is described as follows: We are given a class of linear stochastic systems governed by

$$dX(t) \;=\; AX(t)dt + bdW(t), \quad X(0) = X(0)$$

$$dy(t) \;=\; HX(t)dt + \sigma_0 dW_0(t), \quad y(0) = 0, \tag{6.1}$$

where $X$ is an $R^n$-valued signal process, and $y$ is an $R^m$-valued observation process. The processes $\{W, W_0\}$ are $\{R^p, R^q\}$ valued independent standard Wiener processes. In general, each $\pi = \{A, b, H, \sigma_0\} \in \Sigma$, determines a distinct linear stochastic system of the form (6.1).

The (IP) is to estimate the unknown parameters $\pi = \{A, b, H, \sigma_0\}$, based on the observation $\{y(t), 0 \le t \le T\}$, and the knowledge of the mean $\overline{X}(0)$ and covariance $P(0) \equiv E\{(X(0) - \overline{X}(0))(X(0) - \overline{X}(0))'\}$. Let $\pi^0 \in \Sigma$ denote the true system parameters. Our objective is to develop a method including an algorithm for identification of the true parameter. We formulate this problem as a deterministic control problem, and use a simulated annealing algorithm to estimate the unknown parameters.

## 6.3 Formulation of the Identification Problem as a Deterministic Control Problem

In this section, we shall show that the (IP) is equivalent to an optimal control problem. This is given in the following theorem.

**Theorem 1:** Consider the (IP) as stated above. This problem is equivalent to the following optimal control problem:

estimate $\pi = \{A, b, H, \sigma_0\} \in \Sigma$ that minimizes the objective functional

$J(\pi, T) = \int_0^T tr\{(\tilde{K}_\pi(t) + K_\pi(t) - P_\pi(t))(\tilde{K}_\pi(t) + K_\pi(t) - P_\pi(t))'\}dt,$

subject to the dynamic constraints

$$
\begin{aligned}
de(t,\pi) &= (A - K_\pi H'R^{-1}H)e(t,\pi)dt + K_\pi H'R^{-1}[dy(t) - H\overline{X}(t,\pi)dt],\ e(0,\pi) = 0,\\
\dot{K}_\pi(t) &= AK_\pi + K_\pi A' + bb' - K_\pi H'R^{-1}HK_\pi,\ K_\pi(0) = K(0) = P(0),\\
\dot{\overline{X}}(t,\pi) &= A\overline{X}(t,\pi),\ \overline{X}(0,\pi) = EX(0),\\
\dot{P}_\pi(t) &= AP_\pi + P_\pi A' + bb',\ P(0,\pi) = P(0), \qquad (6.2)
\end{aligned}
$$

where $R \equiv \sigma_0\sigma_0{}'$ and $\tilde{K}_\pi(t) = E(e(t,\pi)e(t,\pi)')$.

**Proof:** Let $\pi \in \Sigma$ constitute the system given by (6.1). Then by Kalman-Bucy filter theory, the estimator is given by $\widehat{X}(t,\pi) = E(X(t,\pi)/\mathcal{F}_t^y)$ which satisfies the following stochastic differential equation (SDE):

$$
\begin{aligned}
d\widehat{X}(t,\pi) &= A\widehat{X}(t,\pi)dt + K_\pi(t)H'R^{-1}d\nu(t,\pi)\quad \widehat{X}(0,\pi) = \overline{X}(0),\\
\nu(t,\pi) &= y(t) - \int_0^t H\widehat{X}(s,\pi)ds, \qquad (6.3)
\end{aligned}
$$

where $K_\pi$ is the state estimation error covariance and it satisfies the following matrix Riccati differential equation

$$\dot{K}_\pi(t) = AK_\pi + K_\pi A' + bb' - K_\pi H'R^{-1}HK_\pi,\ K_\pi(0) = K(0) = P(0). \qquad (6.4)$$

Here, $\mathcal{F}_t^y$ is the smallest $\sigma$-algebra generated by $\{y(t), t \in [0,T]\}$, and $\nu(t,\pi)$ is a Wiener process. The latter is a so-called innovations process, with

$$E\{\nu(t,\pi)\nu'(s,\pi)\} = R\ min(t,s). \qquad (6.5)$$

The mean of $X = \{X(t,\pi),\ t \geq 0\}$ given by $\overline{X}(t,\pi) = E(X(t,\pi))$, satisfies the following deterministic differential equation

$$\dot{\overline{X}}(t,\pi) = A\overline{X}(t,\pi),\ \overline{X}(0,\pi) = EX(0). \tag{6.6}$$

Defining $e(t,\pi) = \widehat{X}(t,\pi) - \overline{X}(t,\pi)$, we have from equations (6.3) and (6.6) that $e$ satisfies the following (SDE):

$$de(t,\pi) = (A - K_\pi H'R^{-1}H)e(t,\pi)dt + K_\pi H'R^{-1}[dy - H\overline{X}(t,\pi)dt],\ e(0,\pi) = 0. \tag{6.7}$$

In terms of the innovations process, one can write system (6.7) as:

$$de(t,\pi) = Ae(t,\pi)dt + K_\pi H'R^{-1}d\nu(t,\pi),\ e(0,\pi) = 0. \tag{6.8}$$

Further, the process $e = \{e(t,\pi),\ t \geq 0\}$ and the error covariance matrix $K_\pi$ are related through the equation

$$(K_\pi(t)\eta,\eta) = (P_\pi(t)\eta,\eta) - E(e(t,\pi),\eta)^2,\ for\ all\ \eta \in R^n, \tag{6.9}$$

where $P_\pi$ is the covariance of the process $X = \{X(t,\pi),\ t \geq 0\}$ and it satisfies the following differential equation:

$$\dot{P}_\pi(t) = AP_\pi(t) + P_\pi(t)A' + bb',\ \ P_\pi(0) = P(0). \tag{6.10}$$

This is justified as follows: by definition, for each $\eta \in R^n$, we have

$$(K_\pi(t)\eta,\eta) \ \equiv \ E(X(t,\pi) - \widehat{X}(t,\pi),\eta)^2$$

$$
\begin{aligned}
&= E(X(t,\pi) - \overline{X}(t,\pi) + \overline{X}(t,\pi) - \widehat{X}(t,\pi),\eta)^2 \\
&= E(X(t,\pi) - \overline{X}(t,\pi),\eta)^2 + E(\overline{X}(t,\pi) - \widehat{X}(t,\pi),\eta)^2 \\
&+ 2E((X(t,\pi) - \overline{X}(t,\pi),\eta)(\overline{X}(t,\pi) - \widehat{X}(t,\pi),\eta)). \qquad (6.11)
\end{aligned}
$$

Since $((\overline{X}(t,\pi) - \widehat{X}(t,\pi)$ is $\mathcal{F}_t^y$-measurable, we have

$$
E\{(X(t,\pi)-\overline{X}(t,\pi),\eta)(\overline{X}(t,\pi)-\widehat{X}(t,\pi),\eta)\} = -E(\overline{X}(t,\pi) - \widehat{X}(t,\pi),\eta)^2,\ t \in [0,T]. \qquad (6.12)
$$

Using this in the third term of the preceding equation we obtain that

$$
\begin{aligned}
(K_\pi(t)\eta,\eta) &= (P_\pi(t)\eta,\eta) - E(e(t,\pi),\eta)^2 \\
&= (P_\pi(t)\eta,\eta) - (\widetilde{K}_\pi(t)\eta,\eta), \qquad (6.13)
\end{aligned}
$$

for each $t \in [0,T]$. This validates equation (6.9). For the identification of the system parameters, equations (6.8) and (6.13) are most crucial. Suppose the process $\{y^0(t),\ t \in [0,T]\}$, as observed from laboratory(field) measurements, corresponds to the true system parameter, say $\pi^0$. If one uses the same observed process to excite the model system (6.8) with the arbitrary choice of the parameter $\pi$, it is clear that one can not expect equality (6.13) to hold. On the other hand, (6.13) must hold if the trial parameter $\pi$ coincides with the true parameter $\pi^0$. Hence it is logical to adjust the parameter $\pi$ to have this equality satisfied. This can be achieved by choosing for the cost function the functional given by

$$
J(\pi,T) = \int_0^T tr\{(\widetilde{K}_\pi^0(t) + K_\pi(t) - P_\pi(t))(\widetilde{K}_\pi^0(t) + K_\pi(t) - P_\pi(t))'\}dt, \qquad (6.14)
$$

where $K_\pi$ and $P_\pi$ are solutions of equations (6.4) and (6.10), and $\widetilde{K}_\pi^0$ is the covariance of the process $e_0(t,\pi) = e(t,\pi,y^0)$ given by the solution of equation (6.7) driven by the observed process $y^0$. This functional is to be minimized on $\Sigma$ subject to dynamic equations (6.2) as proposed in the theorem. This proves that the (IP) is equivalent to the optimal control problem as stated. This completes the proof.

**REMARK 1.**(Uniqueness) Let $\Sigma_0$ be a subset of $\Sigma$. Define $m_0 \equiv inf\{J(\pi,T), \pi \in \Sigma_0\}$. Given that the actual physical system is governed by a linear Ito equation, in general we may expect that $m_0 = 0$. In any case, let $M \equiv \{\pi \in \Sigma_0 : J(\pi,T) = m_0\}$ denotes the set of points in $\Sigma_0$ at which the infimum is attained. It is easy to verify that this set is closed. If the set $M$ is singleton, the system is uniquely defined. In general, for (IP's) which are basically inverse problems, we may not expect uniqueness since the same natural behavior may be realized by many different parameters.

**REMARK 2.** (Weighted cost functional) The cost functional $J(\pi,T)$, given by equation (6.14), can be generalized by introducing a positive semidefinite weighting matrix (valued function) $\Gamma(t)$ in the cost integrand giving

$J(\pi,T) \equiv \int_0^T tr\{L_\pi \Gamma(t) L'_\pi\}, \quad L_\pi \equiv \widetilde{K}_\pi^0 + K_\pi - P_\pi.$

By choosing $\Gamma$ suitably one can assign weights as required for any specific problem.

## 6.4 Measurement of Autocovariance Function of the Zero Mean Estimator Process

In the real world, we can never measure the actual covariance $\widetilde{K}_{\pi}^{0} = E(e_0(t,\pi)e_0'(t,\pi))$ because we can never have all sample functions of the process $\{e\}$. One obtains only a sample path $\{y^0(t), t \in [0,T]\}$ corresponding to the true system parameters, $\pi^0$. Thus, our only recourse is to determine time average based on observation of one sample path of finite length. The time interval is taken large enough so that the ensemble average equals the time average. This is possible if the process is ergodic. In the following discussion we will establish sufficient conditions for ergodicity of the process $\{e\}$ which will be then presented in Proposition 1.

We extend the Brownian motions $W$ and $W_0$ over the entire real line by standard techniques that is we introduce two independent Brownian motions $\widetilde{W}$ and $\widetilde{W}_0$ which are also independent of $W$ and $W_0$ and define

$$W(t) = \begin{cases} W(t), & t \geq 0 \\ \widetilde{W}(-t), & t < 0 \end{cases} \tag{6.15}$$

$$W_0(t) = \begin{cases} W_0(t), & t \geq 0 \\ \widetilde{W}_0(-t), & t < 0. \end{cases} \tag{6.16}$$

Therefore, for $t_0 > 0$, system (6.1) and equation (6.8) can be rewritten as

$$\begin{aligned} dX(t) &= AX(t)dt + bdW(t), \quad X(-t_0) = X(0) \\ dy(t) &= HX(t)dt + \sigma_0 dW_0(t), \quad y(-t_0) = 0, \end{aligned} \tag{6.17}$$

and

$$de(t,\pi) = Ae(t,\pi)dt + K_\pi H'R^{-1}d\nu(t,\pi),\ e(-t_0,\pi) = 0. \qquad (6.18)$$

Suppose the following conditions hold:

**Condition I:** For every $\pi = \{A, b, H, \sigma_0\} \in \Sigma_0$, $A$ is a stable matrix, i.e., all eigenvalues of A have negative real parts.

**Condition II:** For every $\pi = \{A, b, H, \sigma_0\} \in \Sigma_0$, the pair $(A, H)$ is completely observable, that is, the rank $[H', A'H', \ldots\ldots, (A')^{n-1}H']$=n.

Condition I implies that the initial condition of the state has no effect on the asymptotic behavior of the system. Conditions I and II imply that $\lim_{t\to\infty} K_\pi(t)$ exists and is unique. We denote this limit by $K_\pi^0$, which satisfies the algebraic Riccati equation

$$AK_\pi^0 + K_\pi^0 A' + bb' - K_\pi^0 H'R^{-1}HK_\pi^0 = 0. \qquad (6.19)$$

Furthermore, the matrix $A - K_\pi^0 H'R^{-1}H$ is stable [63, Theorem 4.11, pp.367]. Using the steady state version of Kalman-Bucy filter (6.3), that is, using $K_\pi^0$ instead of $K_\pi(t)$ in equation (6.3), one can write $\widetilde{K}_\pi(t)$ as

$$\begin{aligned}\widetilde{K}_\pi(t) &\equiv E(\widehat{X}(t,\pi)\widehat{X}'(t,\pi)) - \overline{X}(t,\pi)\overline{X}'(t,\pi)\\ &= V_1(t) - GV(t)G' - \overline{X}(t,\pi)\overline{X}'(t,\pi). \qquad (6.20)\end{aligned}$$

where the matrices $G$, $V_1$ and $V$ are given as follows:

The matrix $G$ is a $n \times 2n$ with elements $g_{i,i} = 1$, $g_{i,i+n} = -1$, for $1 \le i \le n$, and 0 everywhere else. The matrix $V_1(t) = E(X(t,\pi)X'(t,\pi))$ and it satisfies the matrix differential equation

$$\frac{dV_1(t)}{dt} = AV_1(t) + V_1(t)A' + bb', \quad V_1(-t_0) = V_1(0). \tag{6.21}$$

The matrix $V(t) = \begin{bmatrix} E(X(t,\pi)X'(t,\pi)) & E(X(t,\pi)\widehat{X}'(t,\pi)) \\ E(\widehat{X}(t,\pi)X'(t,\pi)) & E(\widehat{X}(t,\pi)\widehat{X}'(t,\pi)) \end{bmatrix}$ and it satisfies the matrix differential equation

$$\frac{dV(t)}{dt} = \mathcal{A}_\pi V(t) + V(t)\mathcal{A}'_\pi + C_\pi C'_\pi, \quad V(-t_0) = V(0). \tag{6.22}$$

where

$$\mathcal{A}_\pi = \begin{bmatrix} A & 0 \\ K^0_\pi H'R^{-1}H & A - K^0_\pi H'R^{-1}H \end{bmatrix}, \quad C_\pi = \begin{bmatrix} b & 0 \\ 0 & K^0_\pi H'R^{-1}\sigma_0 \end{bmatrix}.$$

Under the conditions $I$ and $II$, the matrices $\{A, \mathcal{A}_\pi\}$ are both stable for every $\pi \in \Sigma_0$, and therefore equations (6.21) and (6.22) have steady state solutions $V_1^0$ and $V^0$, respectively. They are given by the solution of the following algebraic Lyapunov equations

$$AV_1^0 + V_1^0 A' + bb' = 0 \tag{6.23}$$

and

$$\mathcal{A}_\pi V^0 + V^0\mathcal{A}'_\pi + C_\pi C'_\pi = 0. \tag{6.24}$$

respectively. We shall show that the process $e(t,\pi)$, given by equation (6.8), is ergodic. This is presented in the following Proposition.

**Proposition 1:** Suppose that Conditions I and II are satisfied, and the processes $W$ and $W_0$ are the extended Brownian motions as in (6.15) and (6.16). Then for each $\pi \in \Sigma_0$, the process $\{e(t,\pi),\ t \in R\}$ is stationary and ergodic.

**Proof:** It is clear from equation (6.8) that the random process $e(\cdot,\pi)$ is a zero mean Gaussian process. It is stationary if we can show that the corresponding autocovariance matrix $R(s,t)$ is dependent only on the time difference. For this purpose define

$$
\begin{aligned}
R(s,t) &\equiv E(e(s,\pi)e'(t,\pi)) \\
&\equiv E(\widehat{X}(s,\pi)\widehat{X}'(t,\pi)) - \overline{X}(s,\pi)\overline{X}'(t,\pi) \\
&= I_1(s,t) - G\ I_2(s,t)\ G' - \overline{X}(s,\pi)\overline{X}'(t,\pi),
\end{aligned} \tag{6.25}
$$

where the matrices $I_1$ and $I_2$ are given by

$$
\begin{aligned}
I_1(s,t) &= E\int_{-t_0}^{s}\int_{-t_0}^{t} e^{A(s-\theta)} b dW(\theta) dW'(\gamma) b'\ e^{A'(t-\gamma)} \\
&+ e^{A(s+t_0)} V_1(-t_0) e^{A'(t+t_0)}
\end{aligned} \tag{6.26}
$$

and

$$
\begin{aligned}
I_2(s,t) &= E\int_{-t_0}^{s}\int_{-t_0}^{t} e^{A_\pi(s-\theta)} C_\pi d\beta(\theta) d\beta'(\gamma) C_\pi'\ e^{A_\pi'(t-\gamma)} \\
&+ e^{A_\pi(s+t_0)} V(-t_0) e^{A_\pi'(t+t_0)}
\end{aligned} \tag{6.27}
$$

for $\beta = [W, W_0]'$. Setting $s - t \equiv \tau$, after some elementary calculations, we have

$$
I_1(s,t) = \begin{cases} e^{A\tau}\ V_1(t), & \tau \geq 0 \\ V_1(s) e^{-A'\tau}, & \tau \leq 0 \end{cases} \tag{6.28}
$$

$$
I_2(s,t) = \begin{cases} e^{A_\pi\tau}\ V(t), & \tau \geq 0 \\ V(s) e^{-A_\pi'\tau}, & \tau \leq 0 \end{cases} \tag{6.29}
$$

Then, from equations (6.3),(6.25),(6.28) and (6.29), we have

$$R(s,t) = \begin{cases} e^{A\tau}\, V_1(t) - G\, e^{\mathcal{A}_\pi \tau}\, V(t)G' - e^{A\tau}\overline{X}(t,\pi)\overline{X}'(t,\pi), & \tau \geq 0 \\ V_1(s)\, e^{-A'\tau} - G\, V(s)\, e^{-\mathcal{A}_\pi' \tau}G' - e^{A\tau}\overline{X}(t,\pi)\overline{X}'(t,\pi), & \tau \leq 0 \end{cases} \tag{6.30}$$

Since $A$ and $\mathcal{A}_\pi$ are stable matrices, letting $t_0 \to +\infty$, we obtain

$$R(\tau) = \begin{cases} e^{A\tau}\, V_1^0 - G\, e^{\mathcal{A}_\pi \tau}\, V^0\, G', & \tau \geq 0 \\ V_1^0\, e^{-A'\tau} - G\, V^0\, e^{-\mathcal{A}_\pi' \tau}\, G', & \tau \leq 0 \end{cases} \tag{6.31}$$

The latter proves that the process $\{e(t,\pi), t \in R\}$ is stationary. It is well known that the zero mean stationary Gaussian process is ergodic if the corresponding autocovariance matrix $R(\tau)$ satisfies [69, theorem 7.6.1,pp.484]

$$\int_{-\infty}^{\infty} \|R(\tau)\| d\tau < \infty. \tag{6.32}$$

It is clear from (6.31) that $R(\tau) = R'(-\tau)$ and hence

$$\int_{-\infty}^{\infty} \|R(\tau)\| d\tau = 2\int_{0}^{\infty} \|R(\tau)\| d\tau. \tag{6.33}$$

For any $\tau > 0$, we have

$$\|R(\tau)\| \leq \|e^{A\tau}\| \|V_1^0\| + \|G\|^2 \|e^{\mathcal{A}_\pi \tau}\| \|V^0\|. \tag{6.34}$$

Since, for every $\pi \in \Sigma_0$, $A$ and $\mathcal{A}_\pi$ are stable matrices, it holds true that

$$\|e^{A\tau}\| \leq e^{\lambda_1 \tau}, \quad \|e^{\mathcal{A}_\pi \tau}\| \leq e^{\lambda_2 \tau},$$

where $\lambda_1 < 0$, $\lambda_2 < 0$ are the real parts of the largest eigenvalues of the matrices $A$ and $\mathcal{A}_\pi$, respectively. Hence, it follows that

$$\int_{0}^{\infty} \|R(\tau)\| d\tau < \infty, \tag{6.35}$$

and therefore (6.32) holds, proving the ergodicity of the process $\{\epsilon\}$.

Therefore, under conditions $I$ and $II$ and by taking the observation time $T$ large enough, one can approximate the ensemble average of $\epsilon(t,\pi)\epsilon'(t,\pi)$ by its time average. This is what has been done in estimating the unknown parameters in our simulation experiments as given in Section 6.6.

If the stability and observability conditions are not satisfied, one must use Monte-Carlo techniques to produce an ensemble average.

## 6.5 Numerical Algorithm

In this Chapter we applied the method of simulated annealing to determine the optimal parameters that minimize the cost function. The method of simulated annealing is an iterative improvement technique that is suitable for large scale minimization problems. The method avoids being trapped in local minima by using stochastic approach for making moves, based on Metropolis optimization algorithm to minimize the cost function [71]. It works by analogy to the physical annealing of molten material. In the physical situation, the material is cooled slowly, allowing it to coalesce into the lowest possible energy state giving the strongest physical structure. If a liquid metal is cooled quickly, it may end up in a polycrystalline state having a higher energy.

The main idea behind this algorithm is while being at a high temperature, $\tau_a$ called the annealing temperature, where most moves are accepted, then slowly reduce the temperature, while reducing the cost function until only good moves are accepted. The pseudo-code of the algorithm is presented as follows:

**Step 1:** Generate an initial scheduling order randomly and set the temperature at high level.

**Step 2:** Randomly pick one of the elements of $\pi = \{c_1, c_2, \dots\dots\dots, c_n\} \in \Sigma_0$ .A picked parameter moves as

$$c_i = c_i + \alpha U_{c_i} \tag{6.36}$$

where $\alpha$ is the maximal allowed displacement, which for the sake of this argument is arbitrary. $U_{c_i}$ is random number uniformly distributed in the interval $[-1, +1]$, and $U_{c_i}$ is independent of $U_{c_j}$ for $i \neq j$.

**Step 3:** Calculate the change in the cost function, $\Delta J$, which is caused by the move of $c_i$ into $c_i + \alpha U_{c_i}$.

**Step 4:** If $\Delta J < 0$ (i.e., the move would bring the system to a state of lower energy) we allow the move.

**Step 5:** If $\Delta J > 0$ we allow the move with probability $exp(-\Delta J/\tau_a)$; i.e., we take a random number $U$ uniformly distributed between 0 and 1, and if $U < exp(-\Delta J/\tau_a)$, we allow the move. If $U > exp(-\Delta J/\tau_a)$, we return it to its old value.

**Step 6:** Go to step 2 until the cost function stabilizes.

**Step 7:** If $\tau_a = 0$, then stop; otherwise reduce the temperature, and repeat steps 2-6.

## 6.6 Examples and Illustrations

In this Section we will present a two-dimensional example illustrating our results. We assume that the observation data $\{y^0(t), t \in [0, T]\}$ for the real system is generated by the true parameters

$\pi^o = \{A^o, b^o, H^o, \sigma_0^o\}$

where

$$A^o = \begin{bmatrix} -2.0 & 2.0 \\ 0.5 & -2.0 \end{bmatrix}, \quad b^o = \begin{bmatrix} 1.0 & 0.1 \\ 0.1 & 1.0 \end{bmatrix}, \quad H^o = [0.0 \quad 1.0], \quad \sigma_0^o = [1.0].$$

The basic procedure used to obtain the best estimate of the unknown parameters using the algorithm as proposed in section 6.4 is as follows: Let $\pi_c$ be the initial choice for the true parameter $\pi^0$. Using the algorithm with this choice of $\pi_c$ and starting the annealing temperature at $\tau_a$ we arrive at $\pi_{\tau_a}$ by decreasing $\tau_a$ step by step (slowly) to zero. The distance between the computed parameter $\pi_{\tau_a}$ (using the algorithm) and the true parameter $\pi^0$ is denoted by $d(\pi^0, \pi_{\tau_a})$. The simulation was carried out with sampling interval $\delta = 0.01 sec.$, and the observation time $T \in [0, 120] sec.$, and the weighting matrix $\Gamma(t) = 100I$.

**Example:** In general, the (system) dynamic noise and measurement noise are modelled as Wiener processes but the noise power and hence the system and measurement noise covariance matrices may be unknown. In the Ito equation, the martingale terms may then be modelled as $bdW$ and $\sigma_0 dW_0$ where $W$ and $W_0$ are standard Wiener processes and $b$ and $\sigma_0$ are constant but unknown matrices. We assume also that the matrices $A$ and $H$ are unknown. The problem is to determine $A$, $b$, $H$, and $\sigma_0$ based on the observation data $\{y^0(t), t \in [0, T]\}$. The end results are given in table 6.1 which are quite close to the true values. Fig.6.1 shows the estimation error as a function, $\tau_a \rightarrow d(\pi^0, \pi_{\tau_a})$, of the starting annealing temperature $\tau_a$. For fixed observation time $T$, three curves are plotted for three different initial choices $\pi_c$ for the true parameter $\pi^0$. It is clear from this graph that the larger the discrepancy is between the true value and the initial choice the larger is the starting annealing temperature required to reach the true values.

Figure 6.2 shows the estimation error as a function of the observation time $T$, $T \rightarrow d(\pi^0, \pi_T)$, where $\pi_T$ is the estimated value of $\pi^0$ based on the observation $\{y^0(t), 0 \leq t \leq T\}$ till time $T$. As expected, it is a nonincreasing function of $T$ and tends to a limit (saturation) as $T$ becomes larger and $\pi_T$ comes closer to $\pi^0$. The best starting annealing temperature required to obtain the estimate $\pi_T$, in this example, was found to be 25. In other words, the choice of a starting annealing temperature beyond 25 doesn't improve the estimate; it only consumes more CPU time.

**REMARK 3.** Since it well known that the probability law of any Ito process is

Table 6.1:

| | Starting value | Estimated value | Actual value |
|---|---|---|---|
| $a_{11}$ | -1.0 | -1.994406 | -2.0 |
| $a_{12}$ | 1.0 | 1.995625 | 2.0 |
| $a_{21}$ | 2.0 | 0.504037 | 0.5 |
| $a_{22}$ | -3.0 | -2.064103 | -2.0 |
| $s_{11}$ | 0.1 | 1.010031 | 1.01 |
| $s_{12}$ | 0.5 | 0.2000456 | 0.2 |
| $s_{22}$ | 0.1 | 1.010027 | 1.01 |
| $h_{11}$ | -0.7 | -0.000276 | 0.0 |
| $h_{12}$ | 2.0 | 1.009440 | 1.0 |
| ${\sigma_0}^2$ | 0.1 | 0.992090 | 1.0 |

determine by $bb'$ rather than $b$ itself, it is not possible to uniquely identify $b$ or $\sigma_0$. Therefore the results shown in the table are those for $bb'$ and $\sigma_0\sigma_0'$. In the table, $s_{ij}$ are the entries of the matrix $S = bb'$.

## 6.7 Summary and Conclusion

We have presented a formulation of the identification problem for partially observed linear stochastic systems as a deterministic control problem. For this purpose, an appropriate and also natural objective functional has been introduced for the first time in the literature. Using this method we successfully identified all the unknown system parameter $\pi$ simultaneously, as shown in section 6.6.

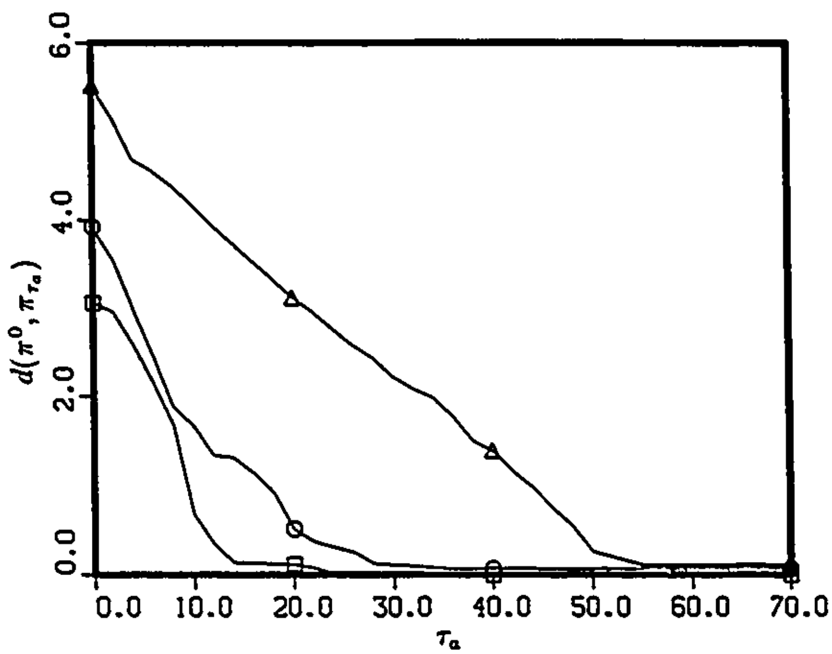


Figure 6.1: The distance between the computed parameter $\pi_{\tau_a}$ and the true parameter $\pi^0$, $d(\pi^0, \pi_{\tau_a})$, as a function of the starting annealing temperature $\tau_a$, for three different initial choices $\pi_c$.

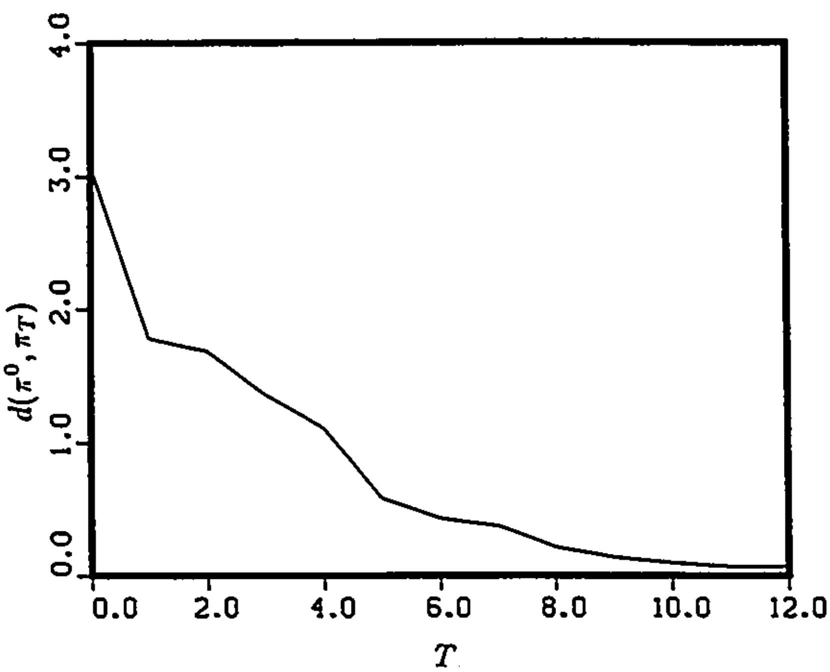


Figure 6.2: The distance between the computed parameter $\pi_T$ and the true parameter $\pi^0$, $d(\pi^0, \pi_T)$, as a function of observation time $T$

# Chapter 7

# CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH

## 7.1 Conclusions

In this thesis filtering and identification problems of stochastic systems have been introduced. In Chapters 4 and 5, we have developed a simple but powerful numerical method for the approximation of the unnormalized conditional (probability) density of filtered diffusion process which satisfies Zakai equation (arises from diffusion processes observed in correlated or uncorrelated noises)and solve the nonlinear filtering. Using Galerkin technique the solution of Zakai equation is approximated by means of a sequence of nonstandard basis functions given by the parameterized family of Gaussian densities. The Gaussian series approximation avoids nonpositivity and has the added advantage of providing a $C^{\infty}$- functions that belongs to the domain of the adjoint of the infinitesimal generator **A** of the markov process being filtered. This technique produced a solution for the Zakai equation in straight forward manner and was readily implementable on a digital computer. The proposed technique has been successfully applied to the one dimensional problems and also to the two and three component- vector problems as shown in Chapter 4 and 5.

In Chapter 6, we consider an identification problem for a system of partially observed linear stochastic differential equations. We present a result whereby one can determine all the system parameters including the covariance matrices of the noise processes. Finally, We formulate the original identification problem as a deterministic control problem and by using the method of simulated annealing, we developed a computational algorithm for identifying the unknown parameters from the available observation.

## 7.2 Suggestions for Further Research

as a continuation of this thesis, further research could be conducted along several directions.

The numerical comparison between Kushner and Zakai equations which arise from diffusion processes observed in correlated (or uncorrelated) noises is an interesting problem to investigate.

Another interesting area is the extension of the results obtained in Chapter 4 and Chapter 5 to the case where the diffusion processes are reflected by the boundary (Numann boundary condition).

Real time implementation of the techniques proposed in Chapters 4 and 5 for analog measurements, would be very useful.

The extension of the results obtained for the identification problems of linear partially observed systems, under time varying parameters $\pi(t)$, is also of importance. Finally, a computer software for parameter identification of nonlinear partially observable stochastic systems, would be very useful.

# Bibliography


[1] Ahmed, N.U., Radaideh, S.M., Modified Extended Kalman Filtering, IEEE Transactions on Automatic control,pp.1322-1326, 1994.

[2] Ahmed, N.U., Radaideh, S.M., Identification of Linear Stochastic Systems Based on Partial Information, Journal of Applied Mathematics and Stochastic Analysis 8, No.3, p.349- 260., 1995.

[3] Ahmed, N.U., Optimal Control of a class of Strongly Nonlinear parabolic systems, J.Math. analysis Appl., vol. 61, p.188-207, 1977.

[4] Ahmed, N.U., Optimal Control of Stochastic Systems, Probabilistic Analysis and Related Topics, p.2-68, Academic Press, New York, 1979.

[5] Ahmed, N.U., Stochastic Control on Hilbert Space for Linear Evolution equation with Random Operator Valued Coefficients, SAIM J. Contr. Optim., vol. 19, No. 3, p.401-430, 1981.

[6] Ahmed, N.U. and Teo, K.L., Optimal Control of Distributed Parameter Systems, Elsevier North Holland, New York, Oxford, 1981.

[7] Ahmed, N.U. and Dabbous, T.E. Nonlinear Filtering of Systems Governed by Ito Differential Equations with Jump Parameters, J. Math. analysis Appl., vol. 115, No.1, p.76-92, 1986.

[8] Ahmed, N.U., Elements of Finite Dimensional Systems and Control Theory, Longman Scientific and Technical, U.K.;Co-publisher: John Wiley, New York, 1988.

[9] Ahmed, N.U., Dabbous, T.E. and Wong, H.W., Gradient Method for Computing Optimal Controls for Stochastic Differential Equations, Stochastic Analysis and Applications, 5(2), 121-150, 1987.

[10] Ahmed.N.U., Li, P., Quadratic Regulator Theory and Linear Filtering Under System Constraints, IMA Journal of Mathematical Control and Information, 8(8), pp.93-107,1991.

[11] Ahmed,N.U., Georganas, N.D., Comments on Invariance in Linear Estimators, Journal of Math. Analysis and Applications, Vol.44, No.1, 1973.

[12] Arnold, L., stochastic differential equations: Theory and applications, John Wiley and Sons, New york, 1974.

[13] Anderson, B.D.O, Moore, J.B., Optimal Filtering, Prentice- Hall, Inc., Englewood Cliffs, N.J., 1979.

[14] Balakrishnan, A.V.,Radon Nikodym Derivatives of a Class of Weak Distributions on Hilbert Spaces, J. Appl. Math. Optim., vol.3, no.2/3 p.209-225, 1977.

[15] Baras, J.S., Blankenship, G.L. and Mitter, S.K., Nonlinear Filtering of Diffusion Processes, Proc. IFAC Congr., Koyoto, Japan Aug. 1981.

[16] Baras, J.S., Vigna, A.L., Real Time Sequential Detection for Diffusion Signals, Proceeding of the 26th Conference on Decision and Control, Los Angeles, CA., 1987.

[17] Banks, J., Carson, J.S., Discrete-Event System Simulation, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1984.

[18] Bensoussan, A., Stochastic Control of Partially Observable systems, Cambridge University Press, 1992.

[19] Bucy, R.S., Nonlinear Filtering Theory, IEEE Aut. Contr., 10, p.198-212, 1965.

[20] Burden, R.L., Faires, J. D., Reynolds, A.C., Numerical Analysis, Prindle, Weber and Schmidt, Boston, 1981.

[21] Bensoussan, A., Maximum Principle and Dynamical Programing Approaches to the Optimal Control of Partially Observed Diffusion, stochastics 9, pp.169-222, 1983

[22] Benes, V.E., Exact Finite Dimensional Filter for Certain Diffusion with Nonlinear Drifts, Stochastics, vol. 6, p.65-92, 1981.

[23] Bagchi, A., Continuous Time Systems Identifications with Unknown Noise Covariance, Automatica, Vol.11, pp.533-536, 1975.

[24] Clark, J.M.C., The Design of Robust Approximation to the Stochastic Differential Equation of Nonlinear Filtering, Communication Systems and Random Process Theory, pp.721-735, 1981.

[25] Dabbous, T.E. and Ahmed, N.U., Nonlinear Filtering of Diffusion Processes with Discontinuous Observations, Stoch. Analys. Appl., vol.2, No.1, p.87-106,1984.

[26] Dabbous,T.E. and Ahmed, N.U.,Parameter Identification for Partially Observed Diffusion, J. OF Optim. Theory and Appl., vol. 75, No.1, 1992.

[27] Davis, M.H. and Marcus, S.I., An Introduction to Nonlinear Filtering, Stochastic Systems: The Mathematics of Filtering and Identification and Applications(NATO Advanced Study Institute Series), Dordrecht, Reidel, p.53-75, 1981.

[28] Davis, M.H., New Approach to Filtering for Nonlinear systems, IEE Proc. Vol.128, No. 5, pp.166-172, 1981

[29] Di Masi, G.B. and Runggaldier, W.J., An Approximation to Optimal Nonlinear Filtering with Discontinuous Observations Stochastic Systems: The Mathematics of Filtering and Identification and Applications(NATO Advanced Study Institute Series), Dordrecht, Reidel, p.583-590, 1980.

[30] Da Prato, G., Zabczyk, J., Stochastic Equations in Infinite Dimensions, Cambridge University Press, Cambridge, New York, 1992.

[31] Elliot, R.J.,Stochastic Calculus and Applications, Springer- Verlag, Heidelburg Berlin, New York, 1982.

[32] Elliot, R.J., Glowinski, R., Approximations to solutions of Zakai Filtering equation, Stochastic Analysis and Applications, 7(2), PP.145-168,1989.

[33] Fleming, W.H., Measure Valued Processes in Control of Partially Observable Stochastic Systems, Appl. Math. Optim., 6, p.271-285, 1980.

[34] Fleming, W.H. and Mitter, S.K., Optimal Control and Pathwise Nonlinear Filtering of Nondegenerate Diffusions, presented at the 20th IEEE conf. Decision Contr., San Diego, CA. 1981.

[35] Fleming, W.J. and Pardoux, E., Optimal Control for Partially Observed Diffusions, SIAM J. Optim., vol. 20, No.2, p.261-285, 1982.

[36] Fleming,W.H., Zhang, Q., Nonlinear Filtering with Small Observation Noise: Piecewise Monotone Observations, stochastic analysis, Academic press, Inc., pp.153-168, 1991.

[37] Friedman, A., Partial Differential Equations of Parabolic Type, Prentice-Hall, Englewood Cliffs, New York, 1964.

[38] Florchinger, P., Gland, L.F., Time Discretization of the Zakai Equation for Diffusion Processes Observed in Correlated Noise. Stochastics and stochastics Reports, Vol.35, pp.233-256,1991.

[39] Gland L.F., Time Discretization of Nonlinear Filtering Equations. Proceedings 28th CDC . Tampa, Florida, pp.2601-2606,1989.

[40] Gland,F.L., Systematic Numerical Experiments in Nonlinear Filtering With Automatic Fortran Code Generation. Proceedings of 25th CDC. Greece, pp.638-642, 1986.

[41] Gland, L.F., Nonlinear Filtering Equation and Problem of Parameter Estimation, Stochastic Systems: The Mathematics of Filtering and Identification and Applications(NATO Advanced Study Institute Series). Dordrecht, Reidel, p.613-620, 1980.

[42] Gihman, I.I. and Skorokhod, A.V., The Theory of Stochastic Processes III, Springer-Verlag, Heidelberg Berlin, New York, 1979.

[43] Gupta, N.K., Mehra, R.K., Computational Aspects of Maximum Likelihood Estimation and Reduction in Sensitivity Function Calculations, IEEE Trans. on Auto. Contr. Vol.19, No.6, 1974.

[44] Gardiner, C.W., Hand Book of Stochastic Methods, Springer- Verlag, New York, 1983.

[45] Hazewinkel, M., On Lie Algebras and Finite Dimensional Filtering, Stochastics, vol. 7, p. 29-62, 1982.

[46] Jazwinski, A.H., Nonlinear Filtering with Discrete Observations, AIAA 3rd Aerospace Sciences Meeting, New York, paper No. 66-39,1966.

[47] Jazwinski, A.H., Adaptive Filtering, Automatica, vol.5, p.475-485, 1969.

[48] Jazwinski.A.H., Stochastic processes and Filtering Theory. Academic Press. New York. London.1970.

[49] Kallianpur. G., Stochastic Filtering Theory. Springer-Verlag. Heidelberg Berlin. New York. 1980.

[50] Kallianpur. G. and Karandikar. R.L., A finitely Additive White Noise Approach to Nonlinear Filtering. Appl. Math. Optim., 10. p.159-185.1983.

[51] Kalman. R.E. and Bucy. R.S., New Results in Linear Filtering and Prediction Theory. Trans. ASME., ser. D: J. Basic Eng., 83. p.95-108.1961

[52] Kunita. H., Asymptotic Behavior of the Nonlinear Filtering Errors of Markov Processes. J. Multivariate Analys., 1. p.365-393. 1971.

[53] Kunita. H., The Stability and Approximation Problems in Nonlinear Filtering Theory. stochastic analysis. Academic press. Inc., pp.311-329, 1991.

[54] Kushner. H.J., On the Dynamical Equations of Conditional Probability Density Functions with Application to Optimal Stochastic Control Theory, J. Math. Analys. Appl., 8, p.332-344, 1964.

[55] Kushner, H.J., On the Differential Equations Satisfied by Conditional Probability Densities of Markov Processes, SIAM J. Contr.. 2, p.106-119. 1964.

[56] Kushner. H.J., Dynamical Equations for Optimal Nonlinear Filtering. J. Diff. Equations. 3, p.179-190, 1967.

[57] Kushner. H.J., Nonlinear Filtering: The Exact Dynamical Equations Satisfied by the Conditional Modes, IEEE Trans. Aut. Contr.. 12, p.262-267. 1967.

[58] Kushner. H.J. Stochastic Stability and Control. Academic Press. New York. London, 1967

[59] Kushner. H.J.. A Robust Discrete State Approximation to the Optimal Nonlinear Filtering for a Diffusion, Stochastics. vol.3, p.75-83. 1979.

[60] Kushner. H.J., Probability Methods for Approximations in Stochastic Control and for Elliptic Equations, Academic Press, 1977.

[61] Kushner. H.J., Nonlinear Filtering for Singularly Perturbed Systems. stochastic analysis, Academic press, Inc., pp.347-369, 1991.

[62] Korezlioglu,H., Mazziotto, G., Approximations of Nonlinear Filter by Periodic Sampling and Quantization, Analysis and Optimization os Systems(Nice 1984),pp.553-567, Springer-Verlag, 1984.

[63] Kwakernaak, H., Sivan, R., Linear Optimal Control Systems. John Wiley and Sons, Inc.,1972.

[64] Katzur, R., Bobrovsky, B.Z., Schuss, Z., Asymptotic Analysis of the Optimal Filtering Problem for One Dimensional Diffusions Measured in a Low Noise Channel, Part I, SIAM J. Appl. Math. Vol.44, No.3, pp.594-604, 1984.

[65] Katzur, R., Bobrovsky, B.Z., Schuss, Z., Asymptotic Analysis of the Optimal Filtering Problem for One Dimensional Diffusions Measured in a Low Noise Channel, Part II, SIAM J. Appl. Math. Vol.44, No.6, pp.1176-1191, 1984.

[66] Kloeden, P.E. and Platen, E., Numerical Solution of Stochastic Differential Equations, Springer-Verlag, Heidelberg Berlin, New York, 1992.

[67] Liang,D.F., Exact and Approximation State Estimation Techniques for Nonlinear Dynamical Systems, A advances in Contr. and Dynamic Systems, Academic Press, New York, London,vol.19, p.1-71, 1983

[68] Liptser, R.S. and Shiryayev, A.N., Studies of Random Processes I and II Springer-Verlag, Heidelberg Berlin, New York, 1978.

[69] Larson. H.J.. Shubert. B.O. Probabilistic models in Engineering Sciences. Vol.I.II. John Wiley and Sons. Inc.. 1979.

[70] Lewis. F.L.. Optimal Estimation with an Introduction to Stochastic Control Theory. John Wiley and Sons. Inc.. 1986.

[71] Metropolis. H.J.. Rosenbluth. A.W.. Rosenbluth. M.N.. Teller. A.H.. Teller. E., Equation of state Calculations by fast Computing Machines. The Journal of Chemical Physics. Vol.21. No.6. 1953.

[72] Maruyama. G.. Continuous Markov Processes and Stochastic Equations. Rend. Cir. Mat. Palermo. Vol.4. pp.48-90.1955.

[73] Milstein. G.N.. Approximate Integration of Stochastic Differential Equations. Theory Prob. Appl., 19. 1974.

[74] Pardoux. E., Nonlinear Filtering Prediction and Smoothing. Stochastic Systems: The Mathematics of Filtering and Identification and Applications(NATO Advanced Study Institute Series), Dordrecht. Reidel. p.529-577. 1980.

[75] Pardoux. E., Stochastic Partial Differential Equations and Filtering of Diffusion Processes, Stochastics 3. No.2. pp.127-167. 1979.

[76] Pardoux. E., Talay. D.. Discretization and Simulation of Stochastic Differential Equations, Acta Applicandae Mathematicae 3. pp.23-47.1985.

[77] Pardoux.E., Roubaud.M.C., Finite Dimensional Approximate Filters in the Case of High Signal-to-Noise Ratio, stochastic analysis. Academic press, Inc., pp.433-447. 1991.

[78] Picard. J., An Estimate of the Error in Time Discretization of Nonlinear Filtering Problems. Theory and Applications of Nonlinear Control Systems, North-Holland, pp.401-412, 1986.

[79] Radaideh. S.M.. Ahmed. N.U.. A Powerful Numerical Technique Solving Zakai Equation for Nonlinear Filtering. Submitted.

[80] Rozovskii.B.L.. A Simple Proof of Uniqueness for Kushner and Zakai Equations. stochastic analysis. Academic press. Inc.. pp.449-458. 1991.

[81] Stratonovich. R.L.. Conditional Markov Processes. Theory of Prob. Appl.. 5. p.156-178. 1960.

[82] Schuss. Z.. Theory and Applications of Stochastic Differential Equations. John Wiley. New York. 1980.

[83] Tungait. J.K.. Continuous- Time System Identification on Compact Parameter sets. IEEE trans. on information Theory. vol. IT-31. p. 652-659. 1985.

[84] Teo. K.L., Liu. Y. and Goh. C.J., Nonlinearly Constrained Discrete Time Optimal Control Problems. Appl. Math. Comp.. vol. 38. p.227-248. 1990.

[85] Tricomi. F.G.. Integral Equations. Interscience Publishers Inc.. New York. 1957.

[86] Vaca. M.V. and Snyder. D.L., Estimation and Decision for Observations Derived from Martingales: Part II. IEEE Trans. on Inform. Theory. vol. IT-22. No.6. p.691-707. 1976.

[87] Wong. E. and Zakai. M.. On the Convergence of Ordinary Integral to Stochastic Integral. Ann. Mat. 36. 1965.

[88] Wong.W.S.. Nonlinear Filtering Approximations by Ovcyannikov-Treves Operators. The 20th IEEE conf. on decision and Control. 1981. p.311-315.

[89] Willman. W.W.. Edgeworth Expansions in State Perturbation Estimation. IEEE Trans. on Auto. Contr.. Vol.26. No.2. pp.493-498. 1981.

[90] Zeitouni, O., Bobrovsky, B.Z., On the Reference Probability Approach to the Equations of Non-Linear Filtering, Stochastics, Vol.19 pp.133-149, 1986.

[91] Zakai, M., On the Optimal Filtering of Diffusion Processes, Z. Wahrsch., Verw. Geb. 11, 230-243, 1969.