AN ABSTRACT OF THE THESIS OF

ALI SHADMAN-VALAVI for the degree of DOCTOR OF PHILOSOPHY

in Electrical and Computer Engineering presented on July 25, 1977

Title: ON THE STATE AND PARAMETER ESTIMATION OF STOCHASTIC BILINEAR SYSTEMS--A SUB-OPTIMAL FILTERING APPROACH

Redacted for privacy

Abstract approved _____ P. A. Frick

Bilinear systems due to their variable structure properties offer more versatility in modelling of nonlinear processes than linear systems.

The state estimation problem for a continuous bilinear system with a continuous observation model is studied and the results are extended to the case where the observations are of a discrete nature. It is shown that the optimal filter is of infinite dimension and a sub-optimal solution based on the use of the conditional best estimate of the state in the multiplicative term, rather than the actual state, is proposed. The filter dimension is reduced to two and the mean and the variance equations are provided.

A recursive maximum likelihood procedure operating on the

proposed filter is used for the parameter identification. Both the likelihood functional and the gradient equations are provided. Computation of the gradient is dependent on computing the partial derivatives of the proposed filter equations with respect to the parameters.

Simulation of the sample functions of bilinear systems using closed form solutions is discussed and a complete solution for the scalar case is provided. Parametric conditions for obtaining closed form solutions to the vector cases are supplied.

A number of numerical examples illustrating the feasibility and performance of the proposed filter and parameter identification schemes are included. Both scalar and multivariable computational examples are considered.

On the State and Parameter Estimation of Stochastic Bilinear Systems--A Sub-optimal Filtering Approach

by

Ali Shadman-Valavi

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

June 1978

APPROVED:

Redacted for privacy

Professor of Electrical and Computer Engineering
in charge of major

Redacted for privacy

Head of Department of Electrical and Computer
Engineering

Redacted for privacy

Dean of Graduate School

Date thesis is presented July 25, 1977

Typed by A & S Bookkeeping/Typing for Ali Shadman-Valavi

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Professor Pieter A. Frick for his guidance throughout the research and preparation phases of this doctoral dissertation. I would also like to thank all my friends at Oregon State University for their constant encouragement and constructive criticism.

Last, but not least, I would like to thank my parents who never questioned my ability and provided me with their absolute support. I would like to especially thank my wife, Sharon, who has given her invaluable assistance in the preparation of this manuscript.

TABLE OF CONTENTS

Chapter		Page
I.	INTRODUCTION	1
	1.1 Background and Motivation	1
	1.2 On the Importance of Bilinear Systems	3
	1.3 Examples of Bilinear Systems	8
	1.3.1 Macro Economic Model	8
	1.3.2 Phase Tracking Problem	11
	1.4 Synopsis	11
II.	REVIEW OF NONLINEAR ESTIMATION THEORY	14
	2.1 Mathematical Framework	14
	2.2 Problem Statement	18
	2.2.1 Continuous Estimation	19
	2.2.2 Continuous-Discrete Estimation	20
	2.3 Optimal State Estimation	22
	2.4 Sub-optimal Filtering Strategies	36
	2.4.1 Continuous Case	37
	2.4.2 Continuous-Discrete Case	42
III.	STATE ESTIMATION FOR BILINEAR SYSTEMS	46
	3.1 Stochastic Models for Bilinear Systems	46
	3.2 Problem Statement	51
	3.3 The Optimal State Estimator for Bilinear Systems	52
	3.3.1 Continuous Case	53
	3.3.2 Continuous-Discrete Case	63
	3.4 Sub-optimal Filter	65
	3.4.1 Continuous Case	65
	3.4.2 Continuous-Discrete Case	70
IV.	PARAMETER ESTIMATION FOR BILINEAR SYSTEMS	74
	4.1 Introduction	74
	4.2 Problem Statement	74
	4.3 On the Choice of the Parameter Estimation Scheme	78
	4.4 Properties of Maximum Likelihood Estimators	82
	4.5 Maximum Likelihood Algorithm	84

Chapter		Page

	4.6 Maximization Procedures	92
	4.6.1 Relaxation Method	93
	4.6.2 Steepest Descent Method	93
	4.6.3 Newton-Raphson Method	94
V.	SIMULATION OF STOCHASTIC BILINEAR SYSTEMS	100
	5.1 Introduction	100
	5.2 Closed Form Solution	101
	5.2.1 Scalar Cases	102
	5.2.2 Vector Cases	105
	5.3 Importance of Data Generation	108
	5.4 Simulation Results	110
VI.	CONCLUSIONS AND SUGGESTIONS FOR FUTURE RESEARCH	125
	6.1 Summary of Results	125
	6.2 Suggestions for Future Research	126
	BIBLIOGRAPHY	128

LIST OF FIGURES

Figure		Page
1.1	Bilinear state diagram.	7
3.1	Stochastic bilinear state diagram.	50
4.1	Schematic diagram of identification procedure.	77
5.1	Flow chart of state and parameter estimation procedure.	111
5.2	State estimates for a scalar homogeneous bilinear system.	113
5.3	Parameter estimates for a scalar bilinear system.	115
5.4	State estimate for a scalar homogeneous bilinear system.	116
5.5	State estimates for an inhomogeneous scalar bilinear system.	121
5.6	State estimates for a second order bilinear system.	123
5.7	State estimates for a second order bilinear system.	124

LIST OF TABLES

Table		Page
5.1	Parameter estimates for a scalar inhomogeneous bilinear system.	120
5.2	Parameter estimates for a second order bilinear system.	122

ON THE STATE AND PARAMETER ESTIMATION OF STOCHASTIC BILINEAR SYSTEMS--A SUB-OPTIMAL FILTERING APPROACH

I. INTRODUCTION

1.1 Background and Motivation

The problem of estimating the states of a dynamical system under noisy conditions has been a subject of great interest to scientists in recent years. The application of estimation theory is quite varied and its concepts have been successfully applied to space technology, automatic control, operations research, and many other diverse fields.

Modern state estimation theory originated with the work of Wiener (W1), but the major development of the optimal recursive linear filter by Kalman and Bucy (K3-4) supplied the basis for the extensive study of the state estimation problem for linear dynamical systems. The filtering problem for nonlinear systems is much more complicated and some optimal nonlinear filtering results are available, as for example in (F4), (K7), (N1). In general, there are three alternative approaches:

i) solving a set of partial differential equations in order to determine the conditional probability density function. A closed form solution to these equations is rarely available.

ii) solving an infinite number of stochastic differential equations to determine the conditional moments of the process.

iii) solving an infinite number of stochastic differential equations to determine the cumulants or quasi-moments of the process.

It is quite evident that a practically implementable optimal nonlinear filter is rarely feasible. Faced with this dilemma researchers have suggested several sub-optimal estimation techniques (N1), (J1), (S3), (W3). The effectiveness of such approximations is obviously dependent on the type of system and nonlinearity under consideration and the performance of these sub-optimal filters has not been extensively examined.

Even though bilinear systems are almost linear, the optimal filter for them is, in general, of infinite dimension with the notable exceptions; the existence of specific structural properties (M1), (J5) or the absence of the multiplicative noise. It is the intent of this thesis to examine the filtering problem for a general class of bilinear systems without any geometric assumptions.

The parameter estimation or identification is obviously closely related to the state estimation problem. There are several classical techniques available; the least squares (A5), stochastic approximation (S2), correlation analysis (F5), and maximum likelihood (A4) are perhaps the most celebrated of these methods. It is well known that the maximum likelihood technique provides the most efficient estimates which are asymptotically convergent to the true

value for the parameters. The parameter estimation problem for bilinear systems has not been extensively studied. Baheti (B1) studied the identification problem for bilinear systems with a single input, single output and no multiplicative noise, while Balakrishnan (B2) has addressed the same problem in a more general setting. It is essential, in our opinion, to consider the case where the noise enters in a multiplicative fashion, because as mentioned earlier, a bilinear system reduces to a linear time varying system in the absence of multiplicative noise.

It is, therefore, intended to develop an identification technique to accommodate a general class of bilinear systems employing a sub-optimal filtering approach.

In order to be both mathematically rigorous and able to use the present results in the filtering theory, the problem will be treated in an Ito calculus framework (I1-2).

1.2 On the Importance of Bilinear Systems

Consider a general nonlinear system described by the following state space model

$$\dot{x} = F(x, u, t) \qquad t \in [t_o, \infty), \tag{1.1}$$

where

$x \epsilon R^n$ is the state of the system,

$u \epsilon R^m$ is the input to the system,

F is an n function.

Expansion of (1. 1) in a Taylor series around a nominal state and control trajectory yields

$$\delta \dot{x} = \frac{\partial}{\partial x} F(x, u, t)\Big|_{\substack{x=x_n \\ u=u_n}} \delta x + \frac{\partial}{\partial u} F(x, u, t)\Big|_{\substack{x=x_n \\ u=u_n}} \delta u \tag{1.2}$$

+ higher order terms.

Clearly after a first order approximation one would be left with a linear model. Linear systems have been extensively used for modelling nonlinear systems. If one chooses to go a step further in the Taylor series expansion, the most obvious choice is a term that would include the product of the state and the control;

$$\delta \dot{x} = \frac{\partial}{\partial x} F(\cdot)\Big|_{\substack{x=x_n \\ u=u_n}} + \frac{\partial}{\partial u} F(\cdot)\Big|_{\substack{x=x_n \\ u=u_n}} \delta u + \frac{\partial}{\partial x \partial u} F(\cdot)\Big|_{\substack{x=x_n \\ u=u_n}} \delta x \delta u \tag{1.3}$$

+ higher order terms.

If one would choose to neglect the higher order terms in (1. 3)

or they would simply vanish, one would have a system where the state and the control enter linearly but the system is not linear in the state and control simultaneously. Such a system is commonly referred to as a bilinear system and can be described by the following differential equation

$$\dot{x} = Ax + Cu + B_L(x, u), \tag{1.4}$$

where

A is an nxn matrix,

C is an nxm matrix,

$B_L(\ )$ is the bilinear operator.

The bilinear term can be written in any of the following two forms:

$$B_L(x, u) = \sum_{i=1}^{m} B_i u_i x, \tag{1.5}$$

or

$$B_L(x, u) = \sum_{i=1}^{n} \overline{B}_i x_i u, \tag{1.6}$$

where

B_i is an nxn matrix,

$\bar{B}_i$ is an nxm matrix

$x = [x_1, x_2, \ldots, x_n]^T$,

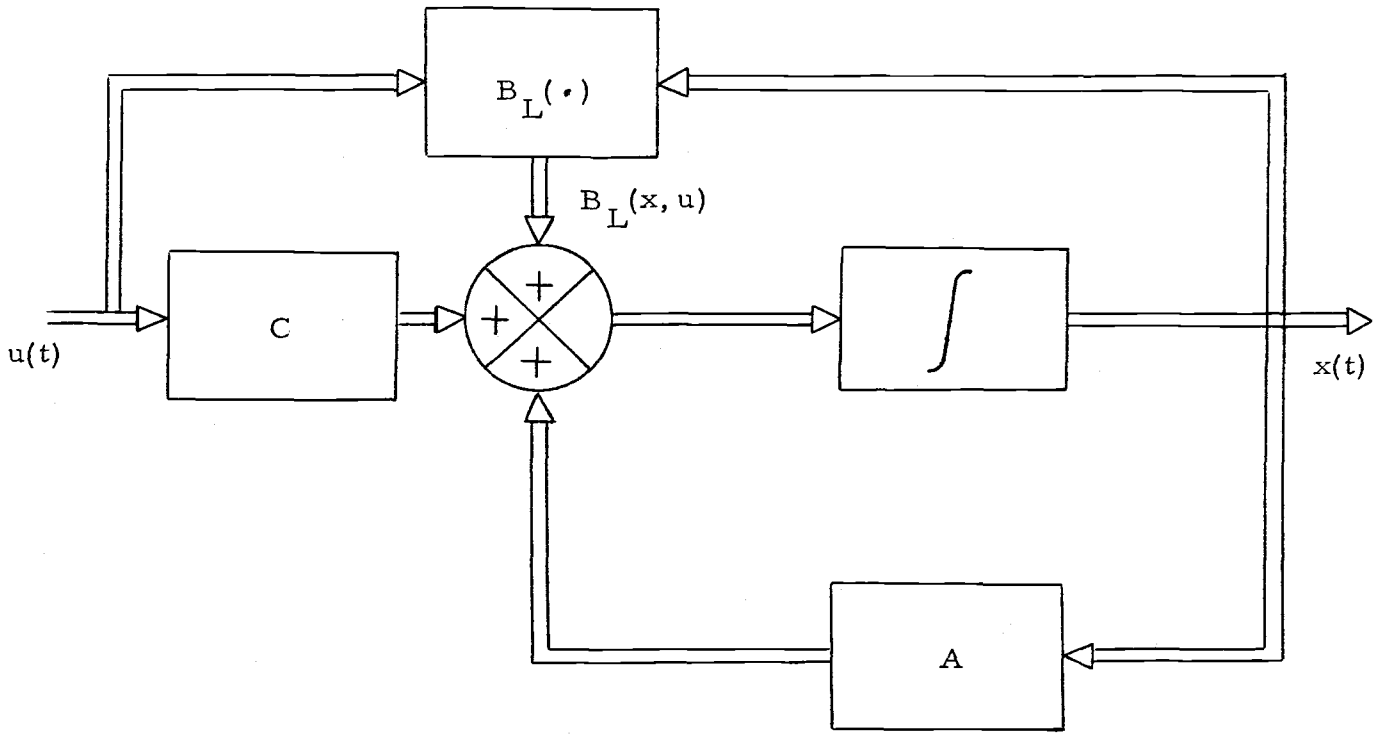
$u = [u_1, u_2, \ldots, u_m]^T$.

Equation (1.5) is the more common one and a bilinear system is generally represented by:

$$x = Ax + Cu + \sum_{i=1}^{m} B_i u_i x . \tag{1.7}$$

A schematic block diagram for a bilinear system is illustrated in Figure 1.1.

As mentioned earlier, it is quite common to approximate a nonlinear system with a linear model. There are, however, cases where such an approximation could prove to be inadequate and bilinear system models of the form (1.7) might provide a viable alternative. In addition, systems described by (1.7) can be used to model certain naturally occurring processes. A detailed review of natural bilinear systems can be found in (M5). Bilinear models have been also used in the design of a control system for nuclear reactors (M6-7).



$$B_L(x, u) = \sum_{i=1}^{m} B_i u_i x$$

Figure 1.1. Bilinear state diagram.

Equation (1.7) can be written in the following form:

$$\dot{x} = (A + \sum_{i=1}^{m} B_i u_i)\, x + Cu. \tag{1.8}$$

For a given set of inputs, one could look at (1.8) as a time varying linear system. The advantage of (1.8) over a linear system is that the structure of (1.8) can be easily changed to accommodate certain objectives by manipulation of the controls, whereas, a linear system does not offer such flexibility. Due to this property bilinear systems are also referred to as variable structure systems.

Principles of differential geometry have been extensively used in the study of certain bilinear systems (B7-10).

Stochastic bilinear systems in addition to their application in the modelling of nonlinear processes occur when there is uncertainty associated with the parameters of a linear system.

1.3 Examples of Bilinear Systems

In this section a few examples where bilinear systems can be used to model real world problems are provided.

1.3.1 Macro Economic Model

The following macro economic model is due to D'Alessandro (D1).

Consider the one sector growth model with a single output $y(t)$

$$y(t) = f(k(t), l(t)), \tag{1.9}$$

where

$k(t)$ is capital goods,

$l(t)$ is labor force.

The labor force is assumed to grow with a constant exponential rate n;

$$\dot{l}(t) = n\, l(t). \tag{1.10}$$

Define

$s(t)$ current total saving ratio,

v constant rate of capital depreciation.

Then, the rate of change of capital goods can be modelled by

$$\dot{k}(t) = s(t)y(t) - vk(t). \tag{1.11}$$

Now, assuming a linear production model,

$$y(t) = ak(t) + bl(t), \tag{1.12}$$

(1.11) reduces to

$$\dot{k}(t) = a\, s(t)k(t) + bs(t)l(t) - v\,k(t). \tag{1.13}$$

Define

$x_1 = k(t)$,

$x_2 = l(t)$,

$u(t) = s(t)$ = input to the system,

equations (1.10) and (1.13) can be modelled with the following state space equation

$$\dot{x} = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} (au(t) - v)\, x_1 + bu(t)\, x_2 \\ nx_2 \end{bmatrix}, \quad (1.14)$$

or equivalently,

$$\dot{x} = \begin{bmatrix} -v & o \\ o & n \end{bmatrix} x + \begin{bmatrix} a & b \\ o & o \end{bmatrix} ux . \quad (1.15)$$

The production equation can be written as

$$y(t) = ax_1 + bx_2 ,$$

or

$$y(t) = [a \quad b]\, x . \quad (1.16)$$

Equations (1.15-6) represent a bilinear system with a linear output model.

1.3.2 Phase Tracking Problem

The following example has been studied by (B12), (W2). Consider a phase tracking problem where the phase θ and the observation z are described by

$$d\theta(t) = \omega_c dt + q^{1/2}(t)\, dw(t), \quad \theta(o) = \theta_o, \tag{1.17}$$

$$dz(t) = \sin\theta(t)dt + r^{1/2}(t)\, dv(t), \tag{1.18}$$

where v and w are independent Wiener processes independent of the random initial phase θ_o. Let

$$x_1 = \sin\theta(t)\,, \qquad x_2 = \cos\theta(t). \tag{1.19}$$

Then

$$dx = \begin{bmatrix} dx_1 \\ dx_2 \end{bmatrix} = \begin{bmatrix} -q(t)/2 & \omega_c \\ \omega_c & -q(t)/2 \end{bmatrix} x + \begin{bmatrix} 0 & q^{1/2}(t) \\ -q^{1/2}(t) & 0 \end{bmatrix} x dw(t)\,, \tag{1.20}$$

$$dz(t) = [1 \quad 0] \quad x + \quad r^{1/2}(t)\, dv(t)\,. \tag{1.21}$$

Equations (1.20-1) describe a stochastic bilinear system with a linear observation model.

1.4 Synopsis

A brief summary of the contents of this thesis is presented

here. The mathematical framework and the general tools required throughout the text as well as a review of nonlinear filtering theory is provided in Chapter II. For the sake of briefness, important theorems are stated without proof, however, adequate references are supplied. Chapter III is concerned with the estimation problem for bilinear systems. A stochastic model for bilinear systems is obtained and an optimal filter for a continuous bilinear system with a continuous observation model in the presence of additive as well as multiplicative state noise, in addition to observation noise is derived. A moment truncation argument is presented to construct a sub-optimal filter. The filtering problem for similar state conditions with discrete observation is also examined. Chapter IV is devoted to the parameter estimation problem for bilinear systems and a maximum likelihood approach in conjunction with the proposed filter of Chapter III is developed. A description of the identification algorithm is given including expressions for the gradient equations. Chapter V contains the simulation results. Closed form solutions to stochastic bilinear differential equations are provided where they exist. The filtering as well as the parameter identification schemes suggested in Chapter III and IV are tested at the hand of a number of examples including scalar, two dimensional, homogeneous, and nonhomogeneous cases. The importance of data acquisition is also discussed.

Chapter VI includes a summary of the results contained in this thesis. A few concluding remarks as well as suggestions for possible future research in this area are also provided.

II. REVIEW OF NONLINEAR ESTIMATION THEORY

2.1 Mathematical Framework

Many random processes are almost Gaussian and have a flat spectrum in the frequency range under consideration. Such processes can be modelled by what is generally referred to as a white Gaussian process. A zero mean white Gaussian process u(t) has the following properties:

i) u(t) is a vector Gaussian random process,

ii) $E[u(t)] = 0$,

iii) $\mathrm{cov}[u(t_1)u^T(t_2)] = Q(t_1)\delta_D(t_1-t_2)$,

where

$E[\cdot]$ denotes the expectation operation,

$\delta_D(\ )$ is the Dirac delta function.

It is important to notice that the zero mean white Gaussian process is defined in terms of a Dirac delta function which is only rigorously defined in an integral equation

$$\int_{-\infty}^{\infty} \delta_D(s)ds = 1 .$$

A Wiener process w(t) is defined in terms of the integral

of a white Gaussian process u(t);

$$w(t) = \int_0^t u(p)dp.$$

The Wiener process possesses the appealing property that dw(t) is completely deterministic; i.e.

$$E[dw(t)] = 0 ,$$

$$E[dw(t)dw^T(t)] = Q(t)dt.$$

Throughout this thesis we will use these definitions and consider stochastic differential equations of the form

$$dx(t) = f(x, t)\, dt + G(x, t)dw(t), \tag{2.1}$$

$$t_0 \geq 0 , \qquad t \in [t_0, \infty).$$

Equation (2.1) is actually a shorthand notation for the integral equation

$$x(t) - x(t_0) = \int_{t_0}^t f(x, s)ds + \int_{t_0}^t G(x, s)dw(s) \tag{2.2}$$

$$= I_1 + I_2 ,$$

where the first integral in (2. 2) is a regular Riemann integral. The second integral, however, is not defined in the classical sense due to the fact that dw(t) is not a function of bounded variation. Several definitions for this stochastic integral (in itself a stochastic process) exist , as for example,

$$I_2 = \int_{t_0}^{t} G(x, s)dw(s) = \underset{\substack{k \to \infty \\ \delta \to 0}}{\text{L. I. M.}} \; G(x, t_i')[w(t_{i+1}) - w(t_i')] \; , \qquad (2.3)$$

where L. I. M. denotes the limit in the mean square sense.

If $t_i' = t_i$ is taken in (2. 3) the integral is referred to as an Ito integral (I1-2), while for $t_i' = (t_{i+1} + t_i)/2$ the integral is called a Stratonovich integral (S8-9). Ito's definition provides a zero mean as well as other nice mathematical properties not possessed by the Stratonovich integral, however, ordinary rules of calculus are satisfied by (2. 3) only if it is interpreted as a Stratonovich integral. If Stratonovich's definition is applied to (2. 2), it would correspond to a system corrupted by a zero mean white Gaussian noise u(t) (C1)

$$\dot{x} = f(x, t) + G(x, t)\, u(t) \; . \qquad (2.4)$$

There exists a relationship between the system representations employing the two integrals definitions as illustrated by the following theorem.

Theorem 2.1: (Wong and Zakai (W4))

Consider the following Ito differential equation:

$$dx(t) = f(x, t)dt + G(x, t)dw(t), \qquad (2.5)$$

$$t \geq t_0 .$$

There exists an equivalent Stratonovich equation to (2.5) which is described by

$$dx_i(t) = (f_i(x, t) - \frac{1}{2} \sum_{k=1}^{n} \sum_{j=1}^{m} G_{kj} \frac{\partial G_{ij}}{\partial x_k}) dt \qquad (2.6)$$

$$+ G_i(x, t)\, dw(t) ,$$

where the subscript i corresponds to the ith row.

Had (2.5) been interpreted as a Stratonovich equation the equivalent Ito equation would have been similar to (2.6) with the only exception being, the correction term would have had a positive sign.

As mentioned earlier the Ito integral and Ito differential equations do not satisfy the rules of ordinary calculus.

Lemma 2.1: (Ito Differential Rule (I1-2)

Let x(t) be the unique solution of the stochastic differential equation

$$dx(t) = f(x, t)dt + G(x, t)\, dw(t), \qquad (2.7)$$

$$t \geq t_0 .$$

Let m(x, t) be a scalar valued real function, continuously differentiable once in t and twice in x. Then, the differential dm satisfies

$$dm(x, t) = m_t(x, t)dt + m_x^T(x, t)dx(t) + \frac{1}{2}\, tr\, (GQG^T m_{xx})dt , \qquad (2.8)$$

where

$$m_t = \frac{\partial m}{\partial t} , \qquad m_x^T = [\frac{\partial m}{\partial x_1} , \ldots\ldots , \frac{\partial m}{\partial x_n}] ,$$

$$m_{xx} = \begin{bmatrix} \frac{\partial^2 m}{\partial x_1^2} & \frac{\partial^2 m}{\partial x_1 \partial x_n} \\ \frac{\partial m}{\partial x_n \partial x_1} & \frac{\partial^2 m}{\partial x_n^2} \end{bmatrix} .$$

Ito calculus offers particularly nice properties which can be exploited in estimation theory and for the remainder of this thesis all stochastic differential equations will be interpreted as Ito equations unless otherwise specified.

2.2 Problem Statement

It is the intent of this chapter to examine the state estimation

problem for a general continuous nonlinear dynamic system with a continuous or a discrete observation model.

2.2.1 Continuous Estimation

Consider the dynamical system

$$dx(t) = f(x, t)dt + G(x, t)dw(t), \qquad (2.9)$$

$$t \geq t_0, \; x(t_0) = x_0 ,$$

where

$x(t)$ is an n vector of state variables,

$f(\;)$ is an n function,

$G(\;)$ is an n x m matrix valued function,

$w(t)$ is an m vector of independent Wiener processes,

$$E[dw(t)dw^T(t)] = Q(t)dt,$$

$w(t)$ is assumed to be independent of the random initial condition x_0

$$E[x(t_0)\, dw^T(t)] = 0.$$

The states of (2.9) are not directly measurable and the following observation model is assumed:

$$dz(t) = h(x, t)dt + dv(t), \qquad (2.10)$$

where

$z(t)$ is a p vector of observations,

$h(\ )$ is a p function,

$v(t)$ is a p vector of independent Wiener processes,

$$E[dv(t)dv(t)^T] = R(t)\,dt\,.$$

It is also assumed that the observation noise is not correlated with the state noise or the initial condition

$$E[dw(t)d\,v(t)^T] = 0,$$

$$E[x(t_0)d\,v(t)^T] = 0.$$

If the observations are made over the time period $t_0 \leq t \leq \tau$, a realization Y_τ can be defined as

$$Y_\tau = (z(s): t_0 \leq s \leq \tau]. \tag{2.11}$$

The problem of estimating the states of (2.9) $x(t)$, based on Y_τ, is the continuous estimation problem. If $t = \tau$, the problem is called the continuous filtering problem; if $t > \tau$, it is called the continuous prediction problem.

2.2.2 Continuous-Discrete Estimation

Given the dynamical system in (2.9), if the observations are

made only at discrete time points $\{t_i\}$ $i = 1, 2, \ldots\ldots$, the observation model becomes

$$y(k) = h(x(k), t_k) + v(k), \qquad (2.12)$$

$$t_k \geq t_0 ,$$

where

$y(k)$ is a p vector of discrete observations,

$h(\ )$ is a p function,

$v(k)$ is a p vector of independent white Gaussian noise sequences,

$$E[v(k)v^T(j)] = R(k)\delta(k-j) ,$$

$\delta(k-j)$ is a Kronecher delta function defined as:

$$\delta(k-j) = \begin{cases} 1 & k=j \\ 0 & k \neq j \end{cases} .$$

If the observations are made over the time period $t_0 \leq t_i \leq t_1$, the manifold Y is defined as:

$$Y_1 = [y(0), y(1), \ldots\ldots, y(1)]. \qquad (2.13)$$

Estimation of the states of (2.9) $x(t_i)$, based on Y_1, is called the continuous-discrete estimation problem. If $i > 1$ the problem is

referred to as the continuous-discrete prediction, if $i = 1$, it is the continuous-discrete filtering problem.

2.3 Optimal State Estimation

The process $x(t)$, which is the solution to (2.9), is obviously a stochastic process. In order to solve the state estimation problem in the continuous case the conditional density function $p(x, t|Y_\tau)$, and in the continuous discrete case the conditional density function $p(x, t|Y_\ell)$ have to be determined.

For a linear system (f and h linear in state, and G function of time only), the density function is Gaussian and can be easily obtained. This is not, however, the case for a general nonlinear system. We have so far neglected to define what constitutes a good state estimator. Let

$$\hat{x}(t|\tau) = \text{the best conditional estimate of the}$$

study $$\tilde{x}(t|\tau) = x(t) - \hat{x}(t|\tau) .$$

An estimate $\hat{x}(\cdot|\cdot)$ is considered good if, $\tilde{x}(\cdot|\cdot)$ has a zero mean (unbiased) and a small variance. It is then quite evident that the knowledge of the mean and the variance of $\tilde{x}(\cdot|\cdot)$ is essential for any state estimation. The "best" estimate evolves from the following well known theorem.

Theorem 2.2:

Let $p(x, t|Y)$ have mean $\hat{x}(t|\cdot) = E[x(t)|Y]$. Suppose $p(x, t|Y.)$ is symmetric about its mean and unimodal (one peak), and let $L(\tilde{x})$ be a class of convex cost functions, then, the optimal estimate of $x(t)$ in the sense of minimizing $L(\tilde{x})$, is the conditional mean $\hat{x}(t|\cdot)$.

This theorem is based on a similar theorem due to Sherman (S5-6) who provides a proof for the unconditional density function $p(x, t)$. Extension of the proof to the conditional case is obvious.

Throughout our discussion we will essentially have no information about the shape of the conditional density function $p(x, t|Y.)$ and the conditional mean $\hat{x}(t|\cdot)$ will be regarded as the best estimate. The following theorem provides a set of partial differential equations for the evaluation of the density function $p(x, t)$.

Theorem 2.3:

Consider the dynamic system described by (2.9). In the absence of any observations, the process $x(t)$ is defined by the density function $p(x, t)$ which satisfies the following partial differential equation:

$$\frac{\partial}{\partial t} p = - \sum_{i=1}^{n} \frac{\partial (pf_i)}{\partial x_i} + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 (GQG^T)_{ij}}{\partial x_i \partial x_j} , \quad (2.14)$$

which is generally known as the Fokker-Planck, or Kolmogorov's

forward equation. Let $\mathcal{L}(\ )$ be an operator defined as:

$$\mathcal{L}(\cdot) = - \sum_{i=1}^{n} \frac{\partial(\cdot f_i)}{\partial x_i} + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2(\cdot GQG^T)_{ij}}{\partial x_i \partial x_j} , \tag{2.15}$$

then, (2.14) can be written as

$$dp = \mathcal{L}(p)\, dt . \tag{2.16}$$

Proof of this theorem can be found in any standard text on estimation theory (J1), (S1).

A solution to (2.16) is rarely available, but it provides an important relationship which can be used to determine the equations for the evolution of the moments of the process x(t).

Lemma 2.2:

Consider the dynamic system of (2.9). Let m(x) be a scalar, twice continuously differentiable function of the state vector x(t) and define

$$\hat{m}(x) = E[m(x)] = \int m(x)p(x,t)dx. \tag{2.17}$$

Then, $\hat{m}(x)$ will obey the following differential equation:

$$dm(x) = E[m_x^T f]\, dt + \frac{1}{2} \operatorname{tr} E[GQG^T m_{xx}]\, dt , \tag{2.18}$$

where m_x and m_{xx} are defined in Lemma 2.1.

Proof: Since $x(t)$ satisfies (2.9) and $m(x)$ is a scalar function of $x(t)$, application of Ito's differential rule (Lemma 2.1) and taking expectations on both sides yields (2.18).

Theorem 2.4:

For the dynamic system of (2.9), in the absence of observations, the mean and variance of the process $x(t)$ satisfy

$$\frac{d}{dt}\hat{x}(t) = E[\hat{f}(x,t)] = \hat{f}(x,t), \tag{2.19}$$

$$\frac{d}{dt}P(t) = \widehat{x(t)f}^T - \hat{x}(t)\hat{f}^T + \widehat{fx}^T(t) - \hat{f}\hat{x}^T(t) + \widehat{GQG}^T , \tag{2.20}$$

where $\wedge$ denotes the expectation operation.

Proof: By respectively setting $m(x) = x$ for the mean, $m(x) = x_i x_j$ for the variance, and recognizing the fact that

$$P(t) = E[(x(t) - \hat{x}(t))(x(t) - \hat{x}(t))^T]$$

$$= E[x(t)x^T(t)] - \hat{x}(t)\hat{x}^T(t),$$

(2.19-20) follow directly from Lemma 2.2.

Let us now incorporate an observation process given by

$$dz(t) = h(x,t)\,dt + dv(t) . \tag{2.21}$$

The solution to the state estimation problem is now equivalent

to determination of the conditional probability density function $p(x, t|Y_\tau)$, and the following theorem due to Kushner (K5), (K9) and later Bucy (B11) provides an equivalent to the Kolmogorov's forward equation for the case with observations.

Theorem 2.5:

Consider the dynamic system of (2.9) and the observation model (2.21). The conditional probability density function $p(x, t|Y_\tau)$ satisfies the equation:

$$dp = \mathcal{L}(p)dt + (h(t) - \hat{h}(t))^T R^{-1}(t)\,(dz(t) - \hat{h}(t)dt)p, \qquad (2.22)$$

$$t \geq t_0, \quad p(t_0) = p_0 ,$$

where

$\mathcal{L}(\cdot)$ is the Kolmogorov's forward operator,

$R(t)$ is the covariance of the observation noise,

$$\hat{h}(t) = E[h(x, t)|Y_\tau] = \int h(x, t)p(x, t|Y_\tau)dx \ .$$

It is worth noting that (2.22) in the case of valueless or no observations ($R^{-1}(t) = 0$), reduces to Kolmogorov's forward equation, as expected. It was mentioned that the solution to the Kolmogorov forward equation is rarely available. A solution to (2.22) is obviously more difficult to achieve, and (2.22) in its present form is of little practical use. However, (2.22) can again be used to determine the equations

for the evolution of the moments of the process x(t) in the presence of an observation process.

Lemma 2.3:

Given the dynamic system of (2.9) and the observation model (2.21), let m(x) be a twice continuously differentiable scalar function of x(t), then, $\hat{m}(x) = E[m(x)|Y_\tau]$ satisfies the following differential equation:

$$d\hat{m}(x) = E[m_x^T f + \frac{1}{2} \operatorname{tr}(GQG^T m_{xx})]\, dt + (\widehat{m(x)h} - \hat{m}(x)\hat{h})^T R^{-1}(t)[dz(t) - \hat{h}dt], \quad t \geq t_0, \tag{2.23}$$

where m_x and m_{xx} are defined in Lemma 2.1.

Proof: This lemma was first proven by Kushner (K7) and provides a tool for writing the equations for the conditional moments of any order. The proof given here is that of Jazwinski (J1). By definition,

$$\hat{m}(x) = \int m(x)p(x, t|Y_\tau)dx,$$

thus,

$$d\hat{m}(x) = \int m(x)dpdx,$$

$$d\hat{m}(x) = \int m(x) \quad (p)dxdt + \int m(x)(h-\hat{h})^T R^{-1}(t)(dz(t)-\hat{h}dt)pdx$$

$$= I_1 + I_2 . \tag{2.24}$$

But, due to Lemma 2.2,

$$I_1 = E[m_x^T f + \frac{1}{2} \operatorname{tr} (GQG^T m_{xx})] \, dt. \tag{2.25}$$

The second integral can be easily evaluated

$$I_2 = (\widehat{m(x)h} - \hat{m}(x)\hat{h})^T R^{-1}(t)(dz(t) - \hat{h}dt). \tag{2.26}$$

Combination of (2.24-26) results in (2.23).

Lemma 2.3 can be employed in determining the equations for the evolution of the mean and covariance for the process x(t).

<u>Theorem 2.6:</u>

For the dynamical system described by (2.9) and the observation model (2.21), the conditional mean and covariance of the process x(t) satisfy the stochastic differential equations

$$d\hat{x}(t) = \hat{f}(\ )\, dt + (\widehat{xh}^T(\ ) - \hat{x}\hat{h}^T(\ ))\, R^{-1}(t)(dz(t) - \hat{h}(\ )dt), \tag{2.27}$$

$$
\begin{aligned}
(dP(t))_{ij} = [& \widehat{x_i f_j} - \hat{x}_i \hat{f}_j + \widehat{f_i x_j} - \hat{f}_i \hat{x}_j + (\widehat{GQG^T})_{ij} \\
& - (\widehat{x_i h} - \hat{x}_i \hat{h})^T R^{-1}(t)\, (\widehat{hx_j} - \hat{h}\hat{x}_j)]\, dt \\
& + (\widehat{x_i x_j h} - \widehat{x_i x_j} \hat{h} - \hat{x}_i \widehat{x_j h} - \hat{x}_j \widehat{x_i h} + 2\hat{x}_i \hat{x}_j \hat{h})^T \\
& X \left\{ R^{-1}(t)\, (dz(t) - \hat{h}dt) \right\},
\end{aligned}
\tag{2.28}
$$

where

x_i	is the ith. element of the state vector $x(t)$,
f_i	is the ith. row of n function f,
$(\ )_{ij}$	is the ij element of an n x n matrix.

Proof: To find the mean $\hat{x}(t)$ set

$$m(x) = x(t) ,$$

then, (2. 27) follows directly from (2. 23). Determination of the covariance is more complicated. By definition

$$P(t) = \widehat{xx}^T - \hat{x}\hat{x}^T ,$$

thus,

$$dP(t) = d(\widehat{xx}^T) - d(\hat{x}\hat{x}^T).$$

Since x satisfies (2. 27), $d(\hat{x}\hat{x}^T)$ can be determined by applying Ito's differential rule to (2. 27) and $d\widehat{xx}^T$ can be derived using

$m(x) = x_i x_j$ and applying Lemma 2.3.

As can be readily seen, (2.27-8) both involve conditional expectation operations and this generally means that the mean and covariance equations are dependent on the higher order moments as well. The exception being, of course, the linear case and the following theorem due to Kalman-Bucy (K4) provides the mean and covariance for a linear system.

Theorem 2.7:

Given the following linear system

$$dx(t) = F(t)xdt + G(t)dw(t) , \tag{2.29}$$

$$dz(t) = H(t)xdt + dv(t), \tag{2.30}$$

the minimum variance filter is given by

$$d\hat{x}(t) = F(t)\hat{x}dt + P(t)H^T(t)R^{-1}(t)(dz(t) - H(t)\hat{x}dt), \tag{2.31}$$

$$\frac{d}{dt}P(t) = F(t)P(t) + P(t)F^T(t) + G(t)QG^T(t) - P(t)H^T(t)R^{-1}(t)H(t)P(t). \tag{2.32}$$

The proof is a special case of Theorem 2.6 where due to the Gaussian property of the resulting conditional density function $p(x, t|Y_\tau)$, the mean and covariance are the only independent conditional moments. The covariance equation (2.32) is no longer a

stochastic differential equation and is decoupled from the equation for the mean.

All the theorems stated thus far have dealt with the continuous case. If the observations are made only at discrete time points a continuous-discrete estimation problem results. This type of estimation was introduced by Jazwinski (J2), and a complete treatment of the subject can be found in (J1). There are, however, a few items of direct interest in our discussion.

Theorem 2.8:

For the dynamic system of (2.9) with the observation model (2.12), if the conditional density function $p(x, t|Y_\ell)$ exists and is once continuously differentiable with respect to time, and twice with respect to the state vector $x(t)$, let $h(.)$ be continuous in both arguments and bounded for each t_i wp. 1. Then, between the observations, $p(x, t|Y_\ell)$ satisfies Kolmogorov's forward equation

$$dp(x, t|Y_\ell) = \mathcal{L}(p)\, dt, \qquad (2.33)$$

$$t_\ell \leq t < t_{\ell+1} ,$$

$$p(x, t_0) = p(x_0),$$

$$p(x, t_\ell|Y_\ell) = \text{initial condition.}$$

At an observation instant (t_1), the conditional density function satisfies the difference equation

$$p(x, t_\ell | Y_\ell) = \frac{p(y(\ell)|x(\ell))\, p(x, t_\ell | Y_{\ell-1})}{\int p(y(\ell)|s)\, p(s, t_\ell | Y_{\ell-1}) ds} \tag{2.34}$$

where

$$p(y(\ell)|x(\ell)) = \left(\frac{1}{2\pi}\right)^{-p/2} |R(\ell)|^{-1/2} \, EXP\left(-\frac{1}{2}(y(\ell)-h(\cdot))^T R^{-1}(\ell)\right.$$

$$\left.(y(\ell) - h(\cdot))\right). \tag{2.35}$$

Proof (J1): Since no information is acquired between the observations, (2.33) is a direct consequence of Theorem 2.3. In order to prove the second part of the theorem, recall

$$p(x, t | Y_\ell) = p(x, t | y(\ell), Y_{\ell-1}) ,$$

applying Baye's rule

$$p(x, t | Y_\ell) = \frac{p(y(\ell)|x, Y_{\ell-1})\, p(x, t_\ell | Y_{\ell-1})}{p(y(\ell)|Y_{\ell-1})} , \tag{2.36}$$

since the observation noise is assumed to be a white Gaussian sequence, then,

$$p(y(\ell)|x, Y_{\ell-1}) = p(y(\ell)|x) , \tag{2.37}$$

in addition,

$$y(\ell) = h(x, t) + v(\ell) ,$$

thus,

$$v(\ell) = y(\ell) - h(x, t),$$

or equivalently,

$$p(y(\ell) \mid x) = p_v(y(\ell) - h(x, t)) \tag{2.38}$$

where

$$p_v(s) = (\frac{1}{2\pi})^{p/2} \; |R(\ell)|^{-1/2} \mathrm{EXP}(-\frac{1}{2} s^T R^{-1}(t) s). \tag{2.39}$$

Again, employing basic rules of probability

$$p(y(\ell) \mid Y_{\ell-1}) = \int p(y(\ell), \; s \mid Y_{\ell-1}) \, ds \tag{2.40}$$

where

$$p(y(\ell), \; s \mid Y_{\ell-1}) = p(y(\ell) \mid s) \; p(s, t \mid Y_{\ell-1}). \tag{2.41}$$

Combining (2.36-41) results in (2.34-5) and the proof is completed.

Theorem 2.8 provides a set of differential and difference equations for the evolution of the conditional density function $p(x, t \mid Y_\ell)$. As was the case in the continuous problem, a solution to (2.33-5) is not generally available. However, Theorem 2.8 provides a mechanism through which the moments of the process x(t) can be determined.

Lemma 2.4:

Hypothesis of Theorem 2.7. Let m(x) be a twice continuously differentiable scalar function of x(t). Then, between the observations $\hat{m}(x)$ satisfies

$$\frac{d}{dt}\,\hat{m}(x) = m_x^T f + \frac{1}{2}\,\mathrm{tr}\,(\widehat{GQG^T m_{xx}})\,, \tag{2.42}$$

$$t_\ell \leq t < t_{\ell+1}\,.$$

At an observation at t_ℓ,

$$\hat{m}(x\,|\,\ell) = \hat{m}^+(x) = \frac{E[\,m(x)p(y(\ell)\,|\,x)]}{E[\,p(y(\ell)\,|\,x)]}\,, \tag{2.43}$$

where $\hat{m}^+(x)$ is the expected value of m(x) after the observation.

Proof: Equation (2.42) follows directly from Lemma 2.2. By multiplying (2.34) by m(x) and taking expectation on both sides, (2.43) is derived.

Results of Lemma 2.4 can be used to determine the mean and the covariance equations for the continuous discrete problem.

Theorem 2.9:

Hypothesis of Theorem 2.8. Between observations, the conditional mean and covariance satisfy

$$\frac{d}{dt}\hat{x}(t) = \hat{f}(x, t) , \tag{2.44}$$

$$\frac{d}{dt}P(t) = \widehat{x(t)f}^T - \hat{x}(t)\hat{f}^T + \widehat{fx}^T(t) - \hat{f}\hat{x}^T(t) + \widehat{GQG}^T,$$

$$t_\ell \leq + < t_{\ell+1} , \tag{2.45}$$

and $\hat{x}(\ell|\ell)$, $p(\ell|\ell)$ serve as initial conditions.

At an observation instance t_ℓ,

$$\hat{x}(\ell|\ell) = \frac{E[x(\ell)p(y(\ell)|x)]}{E[p(y(\ell)|x)]} , \tag{2.46}$$

$$p(\ell|\ell) = \frac{E[x(\ell)x^T(\ell)p(y(\ell)|x)]}{E[p(y(\ell)|x)]} - \hat{x}(\ell|\ell)\hat{x}^T(\ell|\ell). \tag{2.47}$$

Proof: Equations (2.44-5) are due to Theorem 2.4 and (2.46-7) are simple extensions of Lemma 2.4.

Solutions to (2.44-7) are not generally available. However, if the dynamics of the system are linear, the mean and covariance equations can be obtained from the following theorem.

<u>Theorem 2.10:</u>

For a dynamic system described by

$$dx(t) = F(t)xdt + G(t)dw(t) , \tag{2.48}$$

$$y(k) = H(k)\, x(k) + v(k) , \tag{2.49}$$

the optimal minimum variance filter satisfies the following equations.

$$\frac{d}{dt}\hat{x}(t|k) = F(t)\hat{x}(t|k) , \tag{2.50}$$

$$\frac{d}{dt} P(t|k) = F(t)P(t|k) + P(t|k)F^T(t) + G(t)QG^T(t), \tag{2.51}$$

$$t_k \leq t < t_{k+1} ,$$

and at an observation at t_k

$$\hat{x}(k|k) = \hat{x}(k|k-1) + K(k)[y(k) - H(k)\hat{x}(k|k-1)] , \tag{2.52}$$

$$P(k|k) = P(k|k-1) - K(k)H(k)P(k|k-1) , \tag{2.53}$$

where K(k) is the Kalman gain and is given by

$$K(k) = P(k|k-1)H^T(k)[H(k)P(k|k-1)H^T(k) + R(k)]^{-1} . \tag{2.54}$$

Proof: The proof of (2.50-1) is a simple exercise in the use of Theorem 2.9. Derivation of (2.52-4) depends on the normal property of $p(x, t|Y_k)$ for a linear system. Equations (2.50-1) correspond to the prediction problem while (2.52-4) are the filtering equations identical to the discrete Kalman filter (K3-4).

2.4 Sub-optimal Filtering Strategies

We have been able to derive a set of differential or difference equations for the mean and covariance of the optimal minimum

variance filter. These equations involve expectation operations which make them dependent on the higher order moments. To obtain the optimal filter, in general, an infinite number of coupled moment equations have to be solved. Such a filter is not computationally feasible and a type of approximation is in order.

The problem of sub-optimal filtering has been addressed by many authors, as for example in (K6), (S7), (S10-12). Most of the techniques proposed are quite complicated and offer very little, if any, improvement over the sub-optimal filters that will be considered here.

2.4.1 Continuous Case

The optimal minimum variance filter for the linear system described by (2.29-30) is the famous Kalman-Bucy filter (K4). This filter was derived in Theorem 2.7 and the mean and the covariance equations are given by (2.31-2). The most common approach to the derivation of a sub-optimal filter is to consider the linearized version of the system equations. Suppose a nonlinear dynamical system is described by

$$dx = f(x, t)dt + G(t)dw(t), \tag{2.55}$$

and the observation model satisfies

$$dz(t) = h(x, t)dt + dv(t) , \tag{2.56}$$

where G( ) is only a function of time.

Now assume there exists a nominal trajectory x_n satisfying

$$\dot{x}_n = f(x_n, t), \quad t \geq t_0 , \tag{2.57}$$

$$x_n(t_0) = x(t_0) = x_0 ,$$

and a reference observation vector z_n. Define

$$\begin{aligned} \delta x(t) &= x(t) - x_n(t) , \\ \delta z(t) &= z(t) - z_n(t) , \end{aligned} \tag{2.58}$$

then, $\delta x(t)$ and $\delta z(t)$ satisfy the following equations

$$\begin{aligned} \delta \dot{x} &= F(t)\delta x + G(t)\, w'(t) , \\ \delta z(t) &= H(t)\, \delta x + v'(t) , \end{aligned} \tag{2.59}$$

where $v'(t)$ and $w'(t)$ are white Gaussian processes and

$$F(t) = \left.\frac{\partial f(x, t)}{\partial x}\right|_{x=x_n} , \quad H(t) = \left.\frac{\partial h(x, t)}{\partial x}\right|_{x=x_n} . \tag{2.60}$$

Equation (2.59) represents a linear system for which an optimal filter can be obtained. The approximate mean value for the process x(t) will satisfy the following equation:

$$\hat{x}(t) = x_n(t) + \delta\hat{x}(t) , \tag{2.61}$$

and this filter is commonly referred to as the continuous extended or linearized Kalman filter. If the function f in (2.9) is expanded in a Taylor series around its mean

$$f_i(x, t) = f_i(x, t)\Big|_{x=\hat{x}} + \frac{\partial f_i(n, t)}{\partial x}\Big|_{x=\hat{x}} (x-\hat{x}) + \frac{1}{2} \sum_{j=1}^{n} \sum_{k=1}^{n} (x_i - \hat{x}_i)(x_j - \hat{x}_j)$$

$$X \quad \frac{\partial^2 f(x, t)}{\partial x_j \partial x_k}\Big|_{x=\hat{x}} , \tag{2.62}$$

and expectation is taken on both sides of (2.62), neglecting higher than the second order terms and remembering that the best estimate is unbiased ($E[x-\hat{x}] = 0$), we have

$$\hat{f}_i(x, t) = f_i(\hat{x}, t) + \frac{1}{2} \sum_{j=1}^{n} \sum_{k=1}^{n} P_{jk} \frac{\partial^2 f_i(x, t)}{\partial x_j \partial x_k}\Bigg|_{x=\hat{x}} . \tag{2.63}$$

A similar expression can be derived for $\hat{h}_i(x, t)$. However, to make any further simplifications, certain assumptions normally have to be made about the structure of the conditional density function $p(x, t|Y.)$.

If the third and higher order moments are assumed to be negligible, the resulting sub-optimal filter is called a truncated second-order filter (B4),(J3-4), (S4). This filter is useful if the conditional density function $p(x, t|Y)$ is symmetrical and concentrated around its mean. The other possibility is to assume that $p(x, t|Y)$ has Gaussian properties. Since the mean is zero (unbiased), all odd moments vanish, and the fourth central moment can be expressed in terms of the covariance $P(t)$

$$E[(x_i-\hat{x}_i)(x_j-\hat{x}_j)(x_k-\hat{x}_k)(x_\ell-\hat{x}_\ell)] = P_{jk}P_{i\ell}+P_{j\ell}P_{ik}+P_{k\ell}P_{ij} \quad (2.64)$$

and since the nonlinearities are only carried up to second order terms, the higher than fourth central moments do not play any role in the derivation of filter equations. This filter is known as the Gaussian second-order filter (S4). The two filters mentioned are the most common approximations to the nonlinear optimal filter and both reduce to the Kalman-Bucy filter if the dynamics and observations are linear. In order to present these two filters in a compact form the following notation is adopted (J1)

$$(P\partial^2 f)_i = \sum_{j=1}^{n}\sum_{k=1}^{n} P_{jk} \left.\frac{\partial^2 f_i(x,t)}{\partial x_j \partial x_k}\right|_{x=\hat{x}} ,$$

$$(P^2\partial^2 h)_{ijk} = \sum_{q=1}^{n}\sum_{r=1}^{n} \left.\frac{\partial^2 h_k(x,t)}{\partial x_q \partial x_r}\right|_{x=\hat{x}} P_{iq}P_{jr} ,$$

$$(P\partial^2 f) = [(P\partial^2 f)_1, \ldots\ldots, (P\partial^2 f)_n]^T ,$$

$$(P^2\partial^2 h) = [(P^2\partial^2 h)_1, \ldots\ldots, (P^2\partial^2 h)_p]^T ,$$

if g is a p-vector

$$(P^2\partial^2 h): g = \left[\sum_{k=1}^{p} (P^2\partial^2 h)_{ijk} g_k \right] \quad (n \times n) .$$

With the above notation in mind, the truncated second-order filter is given by

$$\begin{aligned} dx(t) = {} & [f(\hat{x}, t) + \frac{1}{2}(P\partial^2 f)]\, dt \\ & + PH^T R^{-1}(t)[dz(t) - (h(\hat{x}, t) + \frac{1}{2}(P\partial^2 h))dt] , \end{aligned} \tag{2.64}$$

$$\begin{aligned} dP = {} & [FP + PF^T + \widehat{GQG}^T - PH^T R^{-1}(t)\, HP]\, dt \\ & - \frac{1}{2} P \{(P\partial^2 h)^T R^{-1}(t)[dz(t) - (h(\hat{x}, t) + \frac{1}{2}(P\partial^2 h))dt]\} , \end{aligned} \tag{2.65}$$

where F and H are defined by (2.60) for $x_n = \hat{x}$.

The Gaussian second-order filter has (2.64) for the mean and the covariance is given by

$$\begin{aligned} dP = {} & [FP + PF^T + \widehat{GQG}^T - PH^T R^{-1}(t)\, HP]\, dt \\ & + (P^2\partial^2 h): \{R^{-1}(t)[dz(t) - (h(\hat{x}, t) + \frac{1}{2}(P\partial^2 h))dt]\} . \end{aligned} \tag{2.66}$$

It should be noted that for a linear observation model both filters will have identical equations for the mean and covariance, however, $\widehat{GQG}^T$ will have, in general, different results for the two filters. Also, the mean and covariance equations are no longer decoupled as in the linear case and the two equations have to be solved simultaneously.

2.4.2 Continuous-Discrete Case

In the case where the observations are of a discrete nature modelled by (2.12), and $G(\cdot)$ is only a function of time, linearizing f and h around a nominal trajectory (2.60) - for small perturbations - the system equations reduce to a linear model. A similar argument to the one presented in the previous section can be used to determine an approximation to the mean $\hat{x}(t)$. Such a filter is referred to as the extended Kalman-Bucy filter for the continuous-discrete case. The sub-optimal filter equations can be derived under similar conditions on the density function $p(x, t \mid Y_\ell)$. The filter equations between observations for the truncated second-order or Gaussian second-order can be easily derived from the corresponding continuous filter equations by setting $R^{-1}(t) = 0$. Thus, between observations the mean and covariance satisfy the following equations

$$\frac{d}{dt}\hat{x}(t) = f(\hat{x}, t) + \frac{1}{2}(P\partial^2 f) , \qquad (2.67)$$

$$\frac{d}{dt} P(t) = FP(t) + P(t)F^T + \widehat{GQG}^T . \tag{2.68}$$

Even though the mean and covariance equations are identical for both filters, $\widehat{GQG}^T$ will, in general, be different for the two approximations. There is still a need to devise an approximation to the solutions of the difference equations (2.46-7). By definition

$$\hat{x}(k|k) = E[x(k)|Y_k] = E[x(k)|y(k), Y_{k-1}] . \tag{2.69}$$

Equation (2.69) suggests that $x(k|k)$ is dependent on the observation $y(k)$. It would then seem logical to expand the mean and covariance in a power series of the following type:

$$\hat{x}(k|k) = \sum_{i=1}^{N} a_i (y(k) - \hat{y}(k))^i , \tag{2.70}$$

$$P(k|k) = \sum_{i=1}^{M} b_i (y(k) - \hat{y}(k))^i , \tag{2.71}$$

where $(y(k) - \hat{y}(k))$ corresponds to the useful information contained in the new observation.

For a general vector equation, computation of the coefficients in (2.70-1) becomes quite involved. For any practical purpose, the expansions can only be carried out to two terms, in which case,

$$\hat{x}(k|k) = a + B(y(k) - \hat{y}(k)) , \tag{2.72}$$

$$[P(k|k)]_{ij} = C_{ij} + \sum_{\ell=1}^{p} D_{ij\ell}(y(k) - \hat{y}(k)). \tag{2.73}$$

The coefficients in (2.72-3) have been derived in (J1) and are as follows

$$a = \hat{x}(k|k-1), \tag{2.74}$$

$$B = (\widehat{xh}^T - \hat{x}\hat{h}^T)\{[(h-\hat{h})(h-\hat{h})^T]\hat{} + R(k)\}^{-1}, \tag{2.75}$$

$$C = P(k|k-1) - (\widehat{xh}^T - \hat{x}\hat{h}^T)\{[(h-\hat{h})(h-\hat{h})^T]\hat{} + R(k)\}^{-1}$$

$$X\ (\widehat{hx}^T - \hat{h}\hat{x}^T), \tag{2.76}$$

$$D_{ij\ell} = \sum_{q=1}^{p} \{[(x_i-\hat{x}_i)(x_j-\hat{x}_j)h_q]\hat{} - \sum_{r=1}^{p} B_{jr}[(h_r-\hat{h}_r)(x-\hat{x}_i)h_q]\hat{}$$

$$- \sum_{r=1}^{p} B_{ir}[(h_r-\hat{h}_r)(x_j-\hat{x}_j)h_q]\hat{} \tag{2.77}$$

$$+ \sum_{r,s=1}^{p} B_{ir}B_{js}[[(h_r-\hat{h}_r)(h_s-\hat{h}_s)h_q]\hat{} + R_{rs}\hat{h}_q] - C_{ij}\hat{h}_q\}$$

$$X\ \{[(h-\hat{h})(h-\hat{h})^T]\hat{} + R\}^{-1}_{q\ell}.$$

Equation (2.77) clearly reveals the magnitude of the problem in the determination of the coefficients for the update equations (2.72-3). A further simplification would involve disregarding the coefficients $D_{ij\ell}$. The resulting filter is referred to as a modified

filter. Equations (2.74-6) have to be evaluated for the truncated second-order or the Gaussian second-order filters. At an observation instance t_k, the modified truncated second-order filter equations are given by

$$\hat{x}(k|k) = \hat{x}(k|k) + P(k|k-1)H^T[X^t]^{-1} \tag{2.78}$$

$$x[y(k) - (h(\hat{x}, t) + \frac{1}{2}(P\partial^2 h))] ,$$

$$P(k|k-1) = P(k|k-1) - P(k|k-1)H^T[X^t]^{-1}HP(k|k-1), \tag{2.79}$$

where

$$X^t = HP(k|k-1)H^T + R(k) - \frac{1}{4}(P\partial^2 h)(P\partial^2 h)^T . \tag{2.80}$$

The modified Gaussian second-order filter is given by equations identical to (2.78-9) with the exception that X^t has to be replaced by X^G defined as

$$X^G = HP(k|k-1)H^T + R(k) + \frac{1}{2}(\partial^2 h P^2 \partial^2 h). \tag{2.81}$$

It is important to note that for a linear observation model, X^t and X^G are identical. As a result, the two filters have the same update equations and the only difference between the two would be due to $\widehat{GQG}^T$.

III. STATE ESTIMATION FOR BILINEAR SYSTEMS

3.1 Stochastic Models for Bilinear Systems

In this chapter we consider a general bilinear system described by

$$\dot{x}(t) = [A(t) + \sum_{i=1}^{m} B_i(t)\bar{u}_i(t)]x(t) + C(t)\bar{u}(t), \tag{3.1}$$

$$t \geq t_0, \quad x(t_0) = x_0 ,$$

where

$x(t)$ is an n vector of state variables,

$A(t)$ is an n x n matrix,

$B_i(t)$ is an n x n matrix,

$$\bar{u}(t) = [\bar{u}_1(t), \bar{u}_2(t), \ldots\ldots, \bar{u}_m(t)]^T$$

is an m vector of controls,

$C(t)$ is an n x m matrix.

If the controls $\bar{u}(t)$ in (3.1) are assumed to be corrupted by some additive white Gaussian noise

$$\bar{u}(t) = u(t) + n(t) \tag{3.2}$$

where

$u(t) = [u_1(t), \ldots\ldots, u_m(t)]^T$ is an m vector of deterministic controls,

$n(t)$ is an m vector of independent white Gaussian disturbances with the covariance

$$E[n(t_1)n^T(t_2)] = Q(t_1)\delta_D(t_1-t_2) ,$$

then, the bilinear system of (3.1) can be represented by:

$$\dot{x}(t) = (A(t) + \sum_{i=1}^{m} B_i(t)u_i(t))\, x(t) + \sum_{i=1}^{m} B_i(t)n_i(t)x(t) \tag{3.3}$$

$$+ C(t)u(t) + C(t)n(t) .$$

Define

$$f(x, t) = (A(t) + \sum_{i=1}^{m} B_i(t)u_i(t))\, x(t) + C(t)u(t), \tag{3.4}$$

$$G(x, t) = \begin{bmatrix} B_1^1(t)x + c_{11} & \ldots\ldots & B_m^1(t)x + c_{1m} \\ \vdots & & \vdots \\ B_1^n(t)x + c_{n1} & \ldots\ldots & B_m^n(t)x + c_{nm} \end{bmatrix}, \tag{3.5}$$

where

$C(t) = [c_{ij}]$,

B_j^i corresponds to the ith. row of the matrix B_j.

Incorporating the notation of (3.4-5), the bilinear system of (3.1) can be represented by

$$\dot{x}(t) = f(x, t) + G(x, t)\, n(t) \,, \qquad t \geq t_0 \,, \tag{3.6}$$

the modelling problem of which was discussed in Chapter II. There is a great deal of research and interest in determining a stochastic differential equation driven by a Wiener process which would have the same solution as (3.6) in the mean square sense (B3), (C1), (M2). Clark (C1) claims (3.6) corresponds to a Stratonovich equation given by

$$dx(t) = f(x, t)dt + G(x, t)dw(t), \tag{3.7}$$

where w(t) is a Wiener process with a covariance

$$E[\, dw(t) d^T w(t)] = Q(t)\, dt \,.$$

The Stratonovich equation, as discussed in Chapter II, is not very appealing for estimation purposes and it is, therefore, desirable to obtain the equivalent Ito equation for the bilinear system of (3.3). Application of Theorem 2.1 to (3.7) yields the desired Ito equation. Recall, the correction term for (3.7) is given as

$$\frac{1}{2} \sum_{k=1}^{n} \sum_{j=1}^{m} Q_{kj} G_{kj} \frac{\partial G_{ij}}{\partial x_k} \, dt \,, \tag{3.8}$$

and after some matrix manipulations, the correction term to (3.3) can be written as

$$[\frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} Q_{ij} B_i(t) B_j(t)] \; x(t) \; dt. \tag{3.9}$$

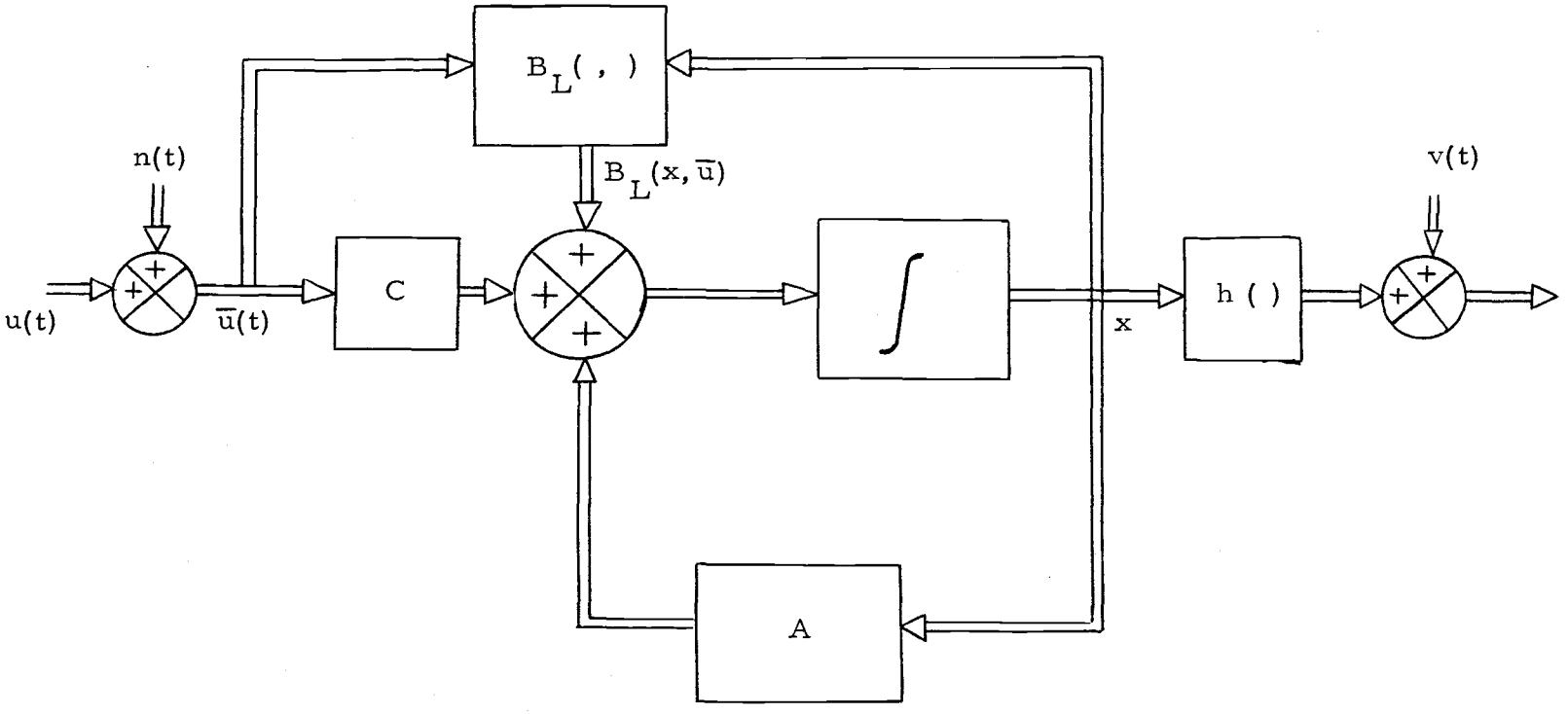
Introducing the correction term in (3.3), the following bilinear Ito differential equation is derived:

$$\begin{aligned} dx(t) = [(A(t) + \sum_{i=1}^{m} B_i(t) u_i(t) + \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} Q_{ij} B_i(t) B_j(t) \; x(t) \\ + C(t)u(t)] \; dt + \sum_{i=1}^{m} B_i(t) x(t) dw_i(t) + C(t) \; dw(t) \end{aligned} \tag{3.10}$$

where

$$dw(t) = [dw_1(t), \; . \; . \; . \; . \; . \; . \; , \; dw_m(t)]^T.$$

Figure 3.1 represents a block diagram for a stochastic bilinear system.



$$B_L(x, \bar{u}) = \sum_{i=1}^{m} B_i \bar{u}_i x$$

Figure 3.1. Stochastic bilinear state diagram.

3.2 Problem Statement

Consider the stochastic bilinear system described by

$$dx(t) = [(A(t) + \sum_{i=1}^{m} B_i(t)u_i(t))x(t) + C(t)u(t)]\, dt$$

$$+ \sum_{i=1}^{m} B_i(t)x(t)dw_i(t) + C(t)dw(t) \tag{3.11}$$

$$t \geq t_0, \; x(t_0) = x_0 \; .$$

To avoid excessive notation the correction term in (3.9) has been omitted. There is no loss of generality since one can always define

$$A(t) = A_0(t) + \sum_{i,j=1}^{m} Q_{ij}B_i(t)B_j(t), \tag{3.12}$$

where $A_0(t)$ would correspond to $A(t)$ in (3.9).

The observation model can be either continuous

$$dz(t) = h(x, t)\, dt + dv(t) \; , \tag{3.13}$$

or discrete

$$y(k) = h(x(k), t_k) + v(k), \tag{3.14}$$

where the dimensions in (3.13-14) are appropriately defined in Chapter II.

The objective of this chapter is to examine the problem of state estimation for a general class of bilinear systems of the type (3.11) based on a set of observations given by (3.13) or (3.14) and to provide a computationally feasible filter algorithm in the case where the observations are linear in the state. It is also intended to reduce the sub-optimal filters of Chapter II for the bilinear case and examine the implications of adopting such strategies.

3.3 The Optimal State Estimator for Bilinear Systems

As mentioned repeatedly throughout Chapter II, the state estimation problem is completely solved if the conditional density function $p(x, t|Y)$ can be found. Theorem 2.5 provides the partial differential equation for the evolution of this density function, but as for the general nonlinear case, for bilinear systems, the solution to this partial differential equation is available only in trivial cases. Under certain assumptions on the motion of the state vector (evolution on manifolds), the density function can be expanded in terms of generalized harmonic functions (M1), (W2). The method becomes quite complicated as the dimension of the state vector gets larger than two. In addition, certain structural properties must be met to ensure that the motion will stay on the manifold. If such assumptions are removed, the only direct approach to solving the

estimation problem is via the moment equations.

3.3.1 Continuous Case

Employing the notation of (3.4-5) the bilinear system of (3.11) can be represented by

$$dx(t) = f(x, t)\, dt + G(x, t)dw(t) \,. \tag{3.15}$$

The optimal minimum variance state estimation problem for (3.15) was addressed in Chapter II. Theorem 2.6 provides the equations for the mean and covariance for this system. These equations can be simplified for the bilinear case.

Define

$$F(t) = A(t) + \sum_{i=1}^{m} B_i(t)u_i(t) \,, \tag{3.16}$$

then, (3.11) reduces to

$$\begin{aligned} dx(t) = {} & [F(t)x(t) + C(t)u(t)]\, dt \\ & + \sum_{i=1}^{m} B_i(t)\, x(t)dw_i(t) + C(t)dw(t), \end{aligned} \tag{3.17}$$

for which (2.27-8) can be written as

$$d\hat{x}(t) = (F(t)\hat{x}(t) + C(t)u(t))\,dt$$

$$+ (\widehat{xh}^T - \hat{x}\hat{h}^T)R^{-1}(t)[dz(t) - \hat{h}dt] \,, \tag{3.18}$$

$$dP(t)_{ij} = [(F(t)P(t) + P(t)\,F^T(t) + \widehat{GQG}^T)_{ij}$$

$$- (\widehat{x_i h} - \hat{x}_i\hat{h})^T R^{-1}(t)(\widehat{hx_j} - \hat{h}\hat{x}_j)]\,dt \tag{3.19}$$

$$+ (\widehat{x_i x_j h} - \hat{x}_i\widehat{x_j h} - \hat{x}_j\widehat{x_i h} - \widehat{x_i x_j}\hat{h} + 2\hat{x}_i\hat{x}_j\hat{h})^T$$

$$X\, R^{-1}(t)\,[dz(t) - \hat{h}dt] \,.$$

Since $G(x, t)$ in (3.11) is a linear function of the state vector, GQG^T will only have up to second order terms in $x(t)$ and its expectation can be evaluated in terms of the mean and covariance functions. Let us first consider the case where $u(t)$ is a scalar. By definition

$$G(x, t) = B(t)x(t) + c(t) \,, \tag{3.20}$$

where G and c are n vectors. Consequently,

$$E[GQG^T] = E[(Bx + c)\,(Bx + c)^T]Q$$

$$= QE[Bxx^TB^T + cx^TB^T + Bxc^T + cc^T] \,.$$

However,

$$E[xx^T] = E[(\tilde{x} + \hat{x})(\tilde{x} + \hat{x})^T] ,$$

$$= P(t) + \hat{x}\hat{x}^T , \tag{3.22}$$

which yields

$$\widehat{GQG}^T = Q[BP(t)B^T + B\hat{x}\hat{x}^T B^T + c\hat{x}^T B^T + B\hat{x}c^T + cc^T] . \tag{3.23}$$

Equation (3.21) can be extended to the case where there is more than one input. Again, by definition

$$G(x, t) = [B_j^i x + c_{ij}] , \tag{3.24}$$

where

$i = 1, \ldots\ldots, n$,

$j = 1, \ldots\ldots, m$,

B_j^i corresponds to the ith. row of matrix B_j.

Consequently,

$$(G(x, t)QG^T(x, t))_{ij} = \sum_{r=1}^{m} Q_r(B_r^i + c_{ir})(B_r^j + c_{jr}) ,$$

after lengthy matrix manipulations the expected value of GQG^T can be written as

$$\widehat{GQG}^T = \sum_{i=1}^{m} [Q_i(B_i P B_i^T + B_i \hat{x}\hat{x}^T B_i^T)$$

$$+ CQ[B_i : \hat{x}]^T + [B_i : \hat{x}]\, QC^T \tag{3.25}$$

$$+ CQC^T$$

where

$$[B_i : \hat{x}] = [B_j^i \hat{x}] , \qquad n \times m \tag{3.26}$$

$i = 1, \ldots\ldots, n,$

$j = 1, \ldots\ldots, m.$

Equations (3.23-26) have been derived without any assumptions about the conditional density function $p(x, t|Y)$.

Any further simplification of the filter equations would require some knowledge about the nature of the observation function $h(.)$. As happens quite frequently the observation model is assumed to be a linear function of the states

$$dz(t) = H(t)x(t)dt + dv(t) . \tag{3.27}$$

Equation (3.27) can now be used to further simplify the mean equation (3.17). By definition

$$h(x, t) = H(t)x(t) ,$$

thus,

$$\hat{h}(x, t) = H(t)\hat{x}(t) ,$$

consequently,

$$\widehat{xh}^T - \hat{x}\hat{h}^T = E[xx^T H^T] - \hat{x}\,\hat{x}^T H^T(t) = P(t)H^T(t) ,$$

which reduces (3.18) to

$$d\hat{x}(t) = (F(t)\hat{x}(t) + C(t)u(t))\,dt + P(t)H^T(t)R^{-1}(t)[dz(t) - H(t)\hat{x}(t)dt] . \tag{3.28}$$

Before proceeding to simplify the covariance equations, it is illuminating to consider the following scalar example.

<u>Example 3.1:</u>

Derive the optimal minimum variance filter equations for the scalar bilinear system given by

$$dx(t) = (a(t)x(t) + b(t))\,dt + (c(t)x(t) + e(t))dw(t), \tag{3.29}$$

$$t \geq t_0 , \quad x(t_0) = x_0 ,$$

with the linear observation model

$$dz(t) = h(t)x(t) + dv(t) . \tag{3.30}$$

Using (3.24) the mean satisfies

$$d\hat{x} = (a(t)\,\hat{x} + b(t))\,dt$$
$$+\ h(t)P(t)R^{-1}(t)[\,dz(t) - h(t)\,\hat{x}\,dt]\ . \tag{3.31}$$

It is informative to derive the equations for the evolution of the moments around the origin. Define

$$M_n(t) = E[\,x^n|Y]\ , \tag{3.32}$$

$$m(x) = x^n\ ,$$

thus,

$$m_x = nx^{n-1}\ ,\quad m_{xx} = n(n-1)x^{n-2}\ .$$

Since m(x) is a scalar twice differentiable function of x, application of Ito's differential rule (Lemma 2.1) and an expectation operation on the resulting stochastic differential equation would yield the desired conditional moment equations;

$$dM_n(t) = [\,(na(t) + \frac{Q}{2}(n-1)nc^2(t)M_n + \frac{Q}{2}n(n-1)e^2(t)M_{n-2}$$
$$+ (nb(t) + Q\,n(n-1)c(t)e(t))\,M_{n-1}]\,dt \tag{3.33}$$
$$+ h(t)R^{-1}(t)\,(M_{n+1}(t) - M_n(t)\,\hat{x}\,(t))\,[\,dz(t) - h(t)\,\hat{x}(t)\,dt]\ ,$$

where

$M_0 = 0$ and $M_1 = E[\,x]$.

It is interesting to note, the moments around the origin depend only on the next higher order moment as well as the lower order moments. Even though this property reduces the complexity of the optimal state estimator for (3.29) the dimension of the problem is still infinite.

The second central moment for the process x(t) can be obtained from (3.19). Since the observation is linear

$$(\widehat{xh} - \hat{x}\hat{h})\, R^{-1}(t)\, (\widehat{hx} - \hat{h}\hat{x}) = h^2(t)P^2(t)R^{-1}, \tag{3.34}$$

$$\widehat{x^2 h}\hat{x} = h(t)\,(P(t) + \hat{x}^2)\,\hat{x}\,. \tag{3.35}$$

By definition

$$P_3(t) = E[(x-\hat{x})^3] = E[x^3] - 3P(t)\hat{x} + \hat{x}^3\,, \tag{3.36}$$

and equation (3.23) gives

$$\widehat{GQG}^T = Q(c^2P(t) + c^2\hat{x}^2 + 2ec\hat{x} + e^2). \tag{3.37}$$

By employing the results of (3.34-7),

$$\begin{aligned} dP(t) = [\,&(2a(t) + Qc^2(t))\,P(t) + Qc^2(t)\hat{x}^2(t) + 2Qc(t)e(t)\hat{x}(t) \\ &+ Qe^2(t) - h^2(t)P^2(t)R^{-1}(t)]\,dt \\ &+ h(t)P_3(t)R^{-1}(t)[\,dz(t) - h(t)\hat{x}(t)\,dt]\,. \end{aligned} \tag{3.38}$$

For a scalar homogeneous bilinear system ($b(t) = e(t) = 0$), the mean and covariance equations can be readily derived from (3.31) and (3.38) by setting b and e equal to zero. In order to determine the third order central moment $P_3(t)$, define

$$P_4(t) = E[x - \hat{x})^4] ,$$

$$= E[x^4 + \hat{x}^4 - 4x^3\hat{x} - 4x\hat{x}^3 + 6x^2\hat{x}^2] ,$$

thus,

$$P_4(t) = E[x^4] - 4M_3\hat{x} + 6M_2\hat{x}^2 - 3\hat{x}^4 . \tag{3.39}$$

Applying Theorem 2.4 and Lemma 2.2 the equation for the evolution of $P_3(t)$ is given

$$\begin{aligned} dP_3(t) = [\, & 3a(t)P_3(t) + 3c^2(t)P_3(t) + 6c^2(t)P(t)\hat{x}(t) \\ & + P(t)P_3(t) + 2x(t)P^2(t)]\, dt \\ & + [P_4(t) - 3P^2(t)]\, [dz(t) - h(t)\hat{x}(t)dt] . \end{aligned} \tag{3.40}$$

Equations (3.38) and (3.40) clearly show the interdependence of the moment equations. Again, as in the moments around the origin, any order central moment depends on the lower order central moments as well as the next higher order one. If the conditional probability density function $p(x, t|Y)$ is assumed to be normal, $P_3(t)$ would vanish, as would $(P_4(t) - 3P^2(t))$.

Let us return to the general vector case and try to simplify the covariance equation (3.19) for linear observations. It is quite simple to show

$$(\widehat{x_i h} - \hat{x}_i \hat{h})^T R^{-1}(t)(\widehat{h x_j} - \hat{h}\hat{x}_j) \Rightarrow (P(t)H^T(t)R^{-1}(t)H(t)P(t))_{ij} \, . \tag{3.41}$$

Define

$$P_{3_{ij\ell}}(t) = E[(x_i - \hat{x}_i)(x_j - \hat{x}_j)(x_\ell - \hat{x}_\ell)] \, , \tag{3.42}$$

then,

$$E[x_i x_j x_\ell] = P_{3_{ij\ell}} + \hat{x}_i P_{j\ell} + \hat{x}_j P_{i\ell} + \hat{x}_\ell P_{ij} + \hat{x}_i \hat{x}_j \hat{x}_\ell \, , \tag{3.43}$$

where $P_3(\ )$ is the third and P is the second central moment.

In order to determine the coefficient of the last term in (3.19), each one of the expectations has to be evaluated;

$$\widehat{x_i x_j h} = \widehat{x_i x_j H(t) x} = H(t)\widehat{x_i x_j x}, \tag{3.44}$$

define

$$P_3(ij) = \begin{bmatrix} P_{3_{ij\ell}} \\ \vdots \\ P_{3_{ijn}} \end{bmatrix} \, , \tag{3.45}$$

thus, (3.44) reduces to

$$\widehat{x_i x_j h} = H(t)[P_3(ij) + \hat{x}_i P_j + \hat{x}_j P_i + P_{ij}\hat{x} + \hat{x}_i \hat{x}_j \hat{x}] \, , \tag{3.46}$$

and the other terms are given by

$$\widehat{x_i x_j \hat{h}} = H(t)\,(P_{ij} + \hat{x}_i \hat{x}_j)\hat{x}\ , \tag{3.47}$$

$$\widehat{\hat{x}_i \widehat{x_j h}} = H(t)\hat{x}_i\ P_j\ , \tag{3.48}$$

$$\hat{x}_j \widehat{x_i h} = H(t)\ \hat{x}_j\ P_i\ . \tag{3.49}$$

The above results can be presented in the following theorem.

<u>Theorem 3.1:</u>

The optimal minimum variance filter for the continuous stochastic bilinear system (3.11) with a linear observation model (3.27) satisfies the following equations

$$\begin{aligned} d\hat{x}(t) &= [\,F(t)\ \hat{x}(t) + C(t)u(t)]\ dt \\ &\quad + P(t)H^T(t)R^{-1}(t)[\,dz(t) - H(t)\ \hat{x}(t)\ dt]\ , \end{aligned} \tag{3.50}$$

$$\begin{aligned} (dP(t))_{ij} &= (F(t)P(t) + P(t)F^T(t) + \widehat{GQG}^T \\ &\quad - P(t)H^T(t)R^{-1}(t)H(t)P(t))\ dt \\ &\quad + P_3(ij)H^T(t)R^{-1}(t)[\,dz(t) - H(t)\ \hat{x}\,(t)dt]\ , \end{aligned} \tag{3.51}$$

where $\widehat{GQG}^T$ is defined in (3.25) and $P_3(ij)$ is given by (3.45).

Proof: Equation (3.50) is identical to (3.28) which was derived for the bilinear system of (3.11) under linear observation assumption.

Employing results of (3.41-9), derivation of (3.51) is an exercise in algebraic operations.

Equations (3.50) and (3.51) reiterate the fact that the filter is dependent on the third central moment. As a matter of fact, it can be proven that the nth. order central moment for (3.11) is only dependent on the (n+1) st. and the lower order moments (for a linear observation model). This property, as in the scalar case, is a characteristic of bilinear systems. It is, however, important to recognize that the solution to the optimal filter is, in general, not available.

3.3.2 Continuous-Discrete Case

If the observations are made at discrete time points

$$y(k) = h(x(t_k), t_k) + v(k) , \tag{3.52}$$

then, the results of Theorem 2.9 can be used to derive the optimal filter for the continuous-discrete case.

Theorem 3.2:

The optimal minimum variance filter for the bilinear system (3.11) and the observation model (3.52) is given by

$$\frac{d}{dt}\hat{x}(t) = F(t)\,\hat{x}(t) + C(t)u(t) , \tag{3.53}$$

$$\frac{d}{dt} P(t) = F(t)P(t) + P(t)F^T(t) + \widehat{GQG}^T, \tag{3.54}$$

$$t_k \leq t < t_{k+1},$$

where $\widehat{GQG}^T$ is defined by (3.25). At an observation point t_k the update equations are given by

$$\hat{x}(k|k) = \frac{E[x(k)p(y(k)|x(k))]}{E[p(y(k)|x(k))]}, \tag{3.55}$$

$$P(k|k) = \frac{E[x(k)x^T(k)p(y(k)|x(k))]}{E[p(y(k)|x(k)]} - \hat{x}(k|k-1)\hat{x}^T(k|k-1). \tag{3.56}$$

Proof: Equations (3.53-4) follow from Theorem 3.1 by setting $R^{-1}(t) = 0$. Theorem 2.9 provided the update equations (for the continuous-discrete case) which are identical to (3.55-6).

Equations (3.53-4) are no longer stochastic or dependent on the higher order moments, so they can be evaluated using regular calculus rules. It should be noted though, that the mean equation is independent of the covariance and can be solved separately, however, the covariance equation is dependent on itself as well as the mean due to the term $\widehat{GQG}^T$. The difference equations (3.55-6) are still impossible to solve explicitly and approximate solutions to them were discussed in Chapter II. Even for a linear observation model there

is no obvious simplification since the conditional density function $p(x, t|Y_k)$ is not known.

3.4 Sub-optimal Filter

As evident from the results of the preceding section, the optimal filter is not computationally feasible. Even in the scalar continuous case one is faced with having to solve an infinite number of coupled Ito differential equations. In the continuous-discrete case the solutions to the update equations involve determination of the coefficients of an infinite series. In order to obtain any estimate of the state, a sub-optimal filtering approach must be pursued.

The approximate techniques discussed in Chapter II can be employed in the derivation of a sub-optimal state estimator for the bilinear system (3.11). The case where the observations are linear is studied here.

3.4.1 Continuous Case

Consider the bilinear system (3.11) with the linear continuous observation model (3.27). It was proven in Theorem 3.1 that the optimal minimum variance filter for such a system is given by (3.50-1) and it was also mentioned that the central moment equations are coupled to the moments of the next higher order. A logical

approximation, therefore, is to assume that for some n

$$P_{n+1}(t) \equiv 0 . \tag{3.56}$$

Under such an assumption, the number of filter equations is reduced to n, the last of which is no longer stochastic and the entire set can, therefore, be systematically solved. There is, however, no clear choice as to what n should be. This type of approximation can be applied to any nonlinear system but is specifically suited to the optimal filter for bilinear systems. It should be noted that even though the moment equations are finite they are coupled Ito differential equations and have to be treated as such for any simulation purposes. It is no secret that the simulation of stochastic differential equations is quite complicated (R2) and renders a value of $n > 2$ in (3.56) impractical. For $n = 2$, $P_3(t) = 0$ is assumed and the following sub-optimal filter is derived (truncated second-order).

<u>Proposition 3.1:</u>

For the bilinear system described by (3.11) and the observation model (3.27) under the assumption that $P_3(t) = 0$, the sub-optimal filter equations satisfy

$$d\hat{x}(t) = (F(t)\hat{x}(t) + C(t)u(t))\, dt$$

$$+ P(t)H^T(t)R^{-1}(t)[\, dz(t) - H(t)\hat{x}(t)dt] , \tag{3.57}$$

$$\frac{d}{dt} P(t) = F(t)P(t) + P(t)F^T(t) + \widehat{GQG}^T - P(t)H^T(t)R^{-1}(t)H(t)P(t) , \tag{3.58}$$

where $\widehat{GQG}^T$ is given by (3.25).

Since the bilinear system (3.11) and the observation model (3.27) only involve terms of first order in x(t), the filter equations do not depend on the fourth order central moment. As a result, the approximate filters in Chapter II (truncated second-order and Gaussian second-order) will yield identical results for bilinear systems with linear observation models. Proposition 3.1 provides the mean and covariance equations for such an approximation.

Let us now consider the following scalar homogeneous stochastic bilinear system

$$\begin{aligned} dx_1(t) &= ax_1 dt + bx_1 dw(t) , \\ & t \geq t_0 , \\ & x_1(t_0) = x_0 > 0 , \\ E[dw^2(t)] &= dt. \end{aligned} \tag{3.59}$$

It will be shown in Chapter V that (3.59) has a solution given by

$$x_1(t) = x_0 \, EXP\left(\int_{t_0}^{t} (a - \frac{1}{2} b^2) ds + \int_{t_0}^{t} b dw(s)\right) . \tag{3.60}$$

The correction term $-\frac{1}{2}b^2$ is due to the fact that

$$E[\,EXP\left(\int_{t_0}^{t} bdw(s)\right)] = EXP(\tfrac{1}{2}b^2(t-t_0)). \qquad (3.61)$$

Suppose (3.59) is approximated by

$$dx_2 = a\,x_2 dt + b\hat{x}_2\,dw(t)\,, \qquad (3.62)$$

$$t \geq t_0\,,$$

$$\hat{x}_2(t_0) = x_2(t_0) = x_0\,,$$

where, by definition

$$\hat{x}_2 = E[x_2]\,. \qquad (3.63)$$

Employing the results of linear system theory, the solution for the process $x_2(t)$ can be written as

$$\hat{x}_2(t) = x_0\,EXP(\int_{t_0}^{t} ads)\,, \qquad (3.64)$$

and, consequently,

$$x_2(t) = x_0\,EXP(\int_{t_0}^{t} ads)\,(1 + \int_{t_0}^{t} bdw(s)\,). \qquad (3.65)$$

There is a close relationship between (3.60) and (3.65) if the exponential term in (3.46) is expanded up to second order terms.

The appealing feature of (3.62) is that it represents a linear system for which the optimal filter is known. With this property in mind, the following approximation is suggested.

Proposition 3.2:

A sub-optimal filter for the bilinear system of (3.11) and (3.27) is given by

$$d\hat{x}(t) = (F(t)\hat{x}(t) + C(t)u(t))\,dt + P(t)H^T(t)R^{-1}(t)\,[dz(t) - H(t)\hat{x}(t)dt]\,, \tag{3.66}$$

$$\frac{d}{dt}P(t) = F(t)P(t) + P(t)F^T(t) + G_1QG_1^T + P(t)H^T(t)R^{-1}(t)H(t)P(t), \tag{3.67}$$

where (3.11) is approximated by

$$dx(t) = (F(t)x(t) + C(t)u(t))\,dt + \sum_{i=1}^{m} B_i(t)dw_i(t)\hat{x}(t) + C(t)dw(t)\,, \tag{3.68}$$

consequently,

$$G_1(x,t) = G_1(\hat{x},t) = [\,B_j^i(t)\,\hat{x}(t) + C_{ij}\,],\ n \times m \tag{3.69}$$

$i = 1, \ldots\ldots, n$,

$j = 1, \ldots\ldots, m.$

The term $G_1 Q G_1^T$ can be evaluated using (3.25) and (3.69),

$$G_1 Q G_1^T = \sum_{i=1}^{m} Q_i B_i \hat{x} \hat{x}^T B_i^T + CQ[B_i : \hat{x}]^T$$

$$+ [B_i : \hat{x}] QC^T + CQC^T , \qquad (3.70)$$

where $[B_i : \hat{x}]$ is defined by (3.26).

The obvious difference between the filter of Proposition 3.2 and the truncated second-order filter is in the term $\widehat{GQG}^T$.

3.4.2 Continuous-Discrete Case

Even with the two sub-optimal filters presented in the preceding section the continuous filter is not well suited for simulation purposes on a digital computer. The continuous-discrete model offers a much more appealing framework for the simulation of the state estimation problem. Since the whole problem of parameter estimation addressed in Chapter IV depends on the availability of estimates of the state based on discrete observations, it is, therefore, essential to develop a sub-optimal estimation strategy for the continuous-discrete case. The techniques discussed at the end of Chapter II can obviously be applied to the bilinear system (3.11) and the observation model

$$y(k) = H(k)\, x(k) + v(k) . \qquad (3.71)$$

In this section a similar argument to Proposition 3.2 will be used to derive an approximation to the optimal filter.

Proposition 3.3:

Consider the bilinear system (3.11) with a linear observation model (3.71). Between the observations, the conditional mean and covariance satisfy

$$\frac{d}{dt}\hat{x}(t) = F(t)\hat{x}(t) + C(t)u(t) , \tag{3.72}$$

$$\frac{d}{dt} P(t) = P(t)F^T(t) + F(t)P(t) + \widehat{GQG}^T , \tag{3.73}$$

where $\widehat{GQG}^T$ is defined by (3.25). At an observation point t_k, the update equations are approximated by the equations for the linear system

$$dx(t) = (F(t)x(t) + C(t)u(t))\, dt + \sum_{i=1}^{m} B_i(t)\hat{x}(t)dw_i(t) + C(t)dw(t), \tag{3.74}$$

consequently,

$$\hat{x}(k|k) = \hat{x}(k|k-1) + K(k)(y(k) - H(t)\hat{x}(k|k-1), \tag{3.75}$$

$$P(k|k) = P(k|k-1) - K(k)H(k)P(k|k-1) , \tag{3.76}$$

$$K(k) = P(k|k-1)H^T(k)\,[H(k)P(k|k-1)H^T(k) + R(k)]^{-1}. \tag{3.77}$$

The filter equations involve no approximations between the observations. The only approximation in Proposition 3.3 is the update equations (3.75-77).

In Chapter II the modified (truncated and Gaussian) second-order filters were derived as an approximation to a general nonlinear continuous-discrete problem. It was mentioned that for linear observations the update equations for the two filters are identical. Since $\widehat{GQG}^T$ was derived (3.25) without any assumptions about the conditional density function, the two filters yield the same results for a bilinear system with a linear observation model. The mean and covariance equations are given by (3.72-77). Proposition 3.3 clearly reveals the type of approximation involved in adopting the modified truncated second-order sub-optimal filtering strategy. Any further attempt in expansion of the series

$$\sum_{i=1}^{N} a_i(y(k) - \hat{y}(k))^i ,$$

would result in cumbersome mathematics which cannot be justified computationally. It is, however, essential to test the proposed filter under simulation conditions and this will be done in Chapter V.

It should be mentioned that the results of this chapter can be easily extended to a much wider class of systems, as for example, given by

$$dx(t) = ((A(t) + \sum_{i=1}^{m} B_i(t)u_i(t))\, x(t) + C(t)u(t))\, dt$$

$$+ \sum_{i=1}^{m} D_i(t)dw_i(t)x(t) + E(t)dw(t)\ . \tag{3.78}$$

By changing $B_i(t)$ and $C(t)$ in $G(x, t)$ to $D_i(t)$ and $E(t)$, all the theorems and propositions presented in this chapter can be modified to accomodate (3.78).

IV. PARAMETER ESTIMATION FOR BILINEAR SYSTEMS

4.1 Introduction

In this chapter we address the problem of bilinear system parameter identification. As mentioned in Chapter I bilinear systems can be used to approximate certain nonlinear physical processes. The choice of a bilinear system as a model is assumed to stem from a priori knowledge about the physical structure and behavior of the process under consideration. There is no available theory at present which would provide a systematic and straight forward mathematical scheme for selecting a model in the identification problem. However, whenever the controls are known to have a multiplicative effect on the process, bilinear systems offer a logical choice for the model.

For the sake of mathematical rigor, the model used in this chapter is again assumed to be a stochastic bilinear differential equation of the Ito type driven by a Wiener process.

4.2 Problem Statement

Assume that the model to be identified can be represented by the stochastic bilinear differential equation

$$dx(t) = ((A(t) + \sum_{i=1}^{m} B_i(t)u_i(t)x(t) + C(t)u(t))dt$$
$$+ \sum_{i=1}^{m} D_i(t)x(t)dw_i(t) + E(t)dw(t) , \tag{4.1}$$

where the dimensions of A, B_i, and C are appropriately defined in Chapter III and

D_i is an n x n matrix,

E is an n x m matrix .

Equation (4.1) is slightly different from the bilinear system of (3.11) to accomodate a larger class of system models. If the Wiener process w(t) in (4.1) is assumed to have unit covariance

$$E[dw(t)d^T w(t)] = I(t)dt , \tag{4.2}$$

then, (4.1) is equivalent to (3.11) under the following assumptions:

$$D_i(t) = Q_i^{1/2} B_i(t) , \tag{4.3}$$

$$E(t) = Q^{1/2} C(t) , \tag{4.4}$$

where Q corresponds to the covariance of the Wiener process in (3.11).

The output of the process is assumed to be a linear function of

the state vector x(t), which is sampled at discrete time points $\{t_i \; i=1, \ldots\ldots\}$

$$y(k) = H(k)x(k) + v(k) \; . \tag{4.5}$$

One can view v(k), a white Gaussian sequence, as the measurement error induced by the recording device which cannot physically respond to the rapid fluctuations in the output of the process. This output can be stored and is available for batch processing.

The input to the system u(t) being a deterministic signal can be recorded accurately and is also available for use in the parameter estimation procedure.

It is further assumed that the order of the process x(t) is known. In cases where this information is not available a series of tests on models of different orders have to be performed (U1), (W5).

The Wiener process w(t) in (4.1) can be viewed as the modelling error with both additive and multiplicative effects. It can also be treated as the corruption of the controls before they reach the plant (a view taken in Chapter III in determination of a stochastic model for a bilinear system).

Given the input-output information, it is desired to find estimates of the parameters in (4.1), and Figure 4.1 illustrates a schematic diagram of the identification procedure.

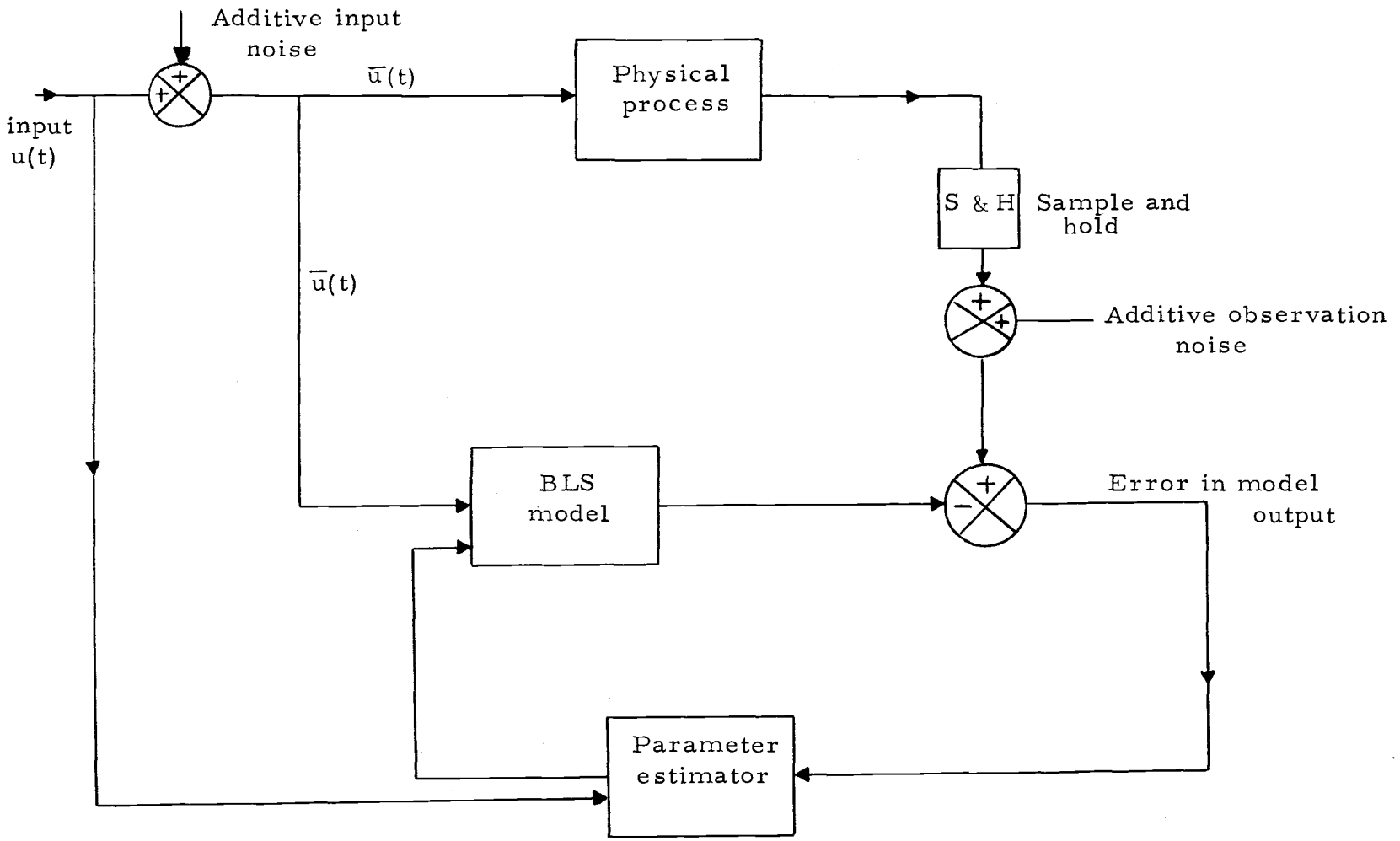


Figure 4.1. Schematic diagram of identification procedure.

4.3 On the Choice of the Parameter Estimation Scheme

As a starting point for our discussion of the parameter identification problem, note that in the absence of input noise, (4.1) would reduce to

$$\frac{d}{dt} x(t) = (A(t) + \sum_{i=1}^{m} B_i(t)u_i(t))\, x(t) + C(t)u(t) \,, \qquad (4.6)$$

thus reducing the bilinear system to a deterministic one. Equation (4.6) can be viewed as a linear time varying system for which the parameter estimation problem has been extensively studied (E1), (A4) and needs no further discussion. Even in cases where only the multiplicative noise can be neglected

$$dx(t) = ((A(t) + \sum_{i=1}^{m} B_i(t)u_i(t))x(t) + C(t)u(t))dt + E(t)dw(t), \qquad (4.7)$$

one is faced with a linear time varying system corrupted by a Wiener process. The parameter estimation problem for (4.7) can also be handled using identification techniques for linear systems (B2).

The difficulty arises, however, when the noise does enter in an additive, as well as, multiplicative fashion. From the discussion of Chapter III, (4.1) no longer depicts a linear system and the optimal state estimator is of infinite dimension.

Faced with a nonlinear system identification problem a short

survey of available methods is presented here (A5), (G3). Classical nonparametric approaches which involve correlation, or power spectra analysis are extremely difficult to apply to nonlinear cases, particularly in the presence of a multiplicative noise component. Alternatively, if the parameters are assumed to be time invariant, one can define

$\theta =$ vector of all the parameters to be estimated,

$$\theta = [\theta_1, \theta_2, \ldots\ldots, \theta_q]^T, \tag{4.8}$$

then,

$$\frac{d}{dt}\theta = 0. \tag{4.9}$$

Equation (4.1) can then be augmented to yield

$$\begin{bmatrix} dx \\ \\ d\theta \end{bmatrix} = \begin{bmatrix} (A(t) + \sum_{i=1}^{m} B_i(t)u_i(t))x(t) + C(t)u(t) \\ \\ 0 \end{bmatrix} dt \tag{4.10}$$

$$+ \begin{bmatrix} \sum_{i=1}^{m} D_i(t)x(t)dw_i(t) + E(t)dw(t) \\ \\ 0 \end{bmatrix} .$$

By defining

$$\bar{A}(t) = \begin{array}{c c} & \begin{array}{cc} n & q \end{array} \\ \begin{array}{c} n \\ \\ q \end{array} & \left[\begin{array}{c:c} A(t) & 0 \\ \hdashline 0 & 0 \end{array}\right] \end{array} , \quad \bar{B}_i(t) = \begin{array}{c c} & \begin{array}{cc} n & q \end{array} \\ \begin{array}{c} n \\ \\ q \end{array} & \left[\begin{array}{c:c} B_i(t) & 0 \\ \hdashline 0 & 0 \end{array}\right] \end{array} ,$$

$$\bar{D}_i(t) = \begin{array}{c c} & \begin{array}{cc} n & q \end{array} \\ \begin{array}{c} n \\ \\ q \end{array} & \left[\begin{array}{c:c} D_i(t) & 0 \\ \hdashline 0 & 0 \end{array}\right] \end{array} , \quad \bar{C}(t) = \begin{array}{c c} & m \\ \begin{array}{c} n \\ \\ q \end{array} & \left[\begin{array}{c} C(t) \\ \hdashline 0 \end{array}\right] \end{array} ,$$

$$\bar{E}(t) = \begin{array}{c c} & m \\ \begin{array}{c} n \\ \\ q \end{array} & \left[\begin{array}{c} E(t) \\ \hdashline 0 \end{array}\right] \end{array} , \quad Z = [x_1, \ldots, x_n, \theta_1, \ldots, \theta_q]^T,$$

$$\bar{H}(t) = p \begin{array}{c} \begin{array}{cc} n & q \end{array} \\ \left[\begin{array}{c:c} H(k) & 0 \end{array}\right] \end{array} ,$$

equations (4.5) and (4.9) can be written as

$$dZ = ((\bar{A}(t) + \sum_{i=1}^{m} \bar{B}_i(t)u_i(t))Z + \bar{C}(t)u(t)dt + \sum_{i=1}^{m} \bar{D}_i(t)\, Z dw_i(t) + \bar{E}(t)dw(t) , \tag{4.11}$$

$$y(k) = \bar{H}(k)\, Z(k) + v(k) . \tag{4.12}$$

Equations (4.11-2) represent a bilinear system of dimension $n + q$ with a linear observation model. However, very little is gained by adopting this standard procedure (B5-6), (D2) since the parameter estimation still requires a state estimator of even higher dimension than the original system. With no finite dimensional filter available in such cases the added dimensionality unnecessarily complicates any of the discussed sub-optimal state estimation strategies.

The two most widely used parametric or direct methods in parameter estimation are the least-squares and maximum likelihood procedures. In the least-squares approach the output error is minimized, but in the presence of state noise, the least-squares method will generate biased estimates for the parameters. The maximum likelihood procedure does not suffer from such a disadvantage and has been shown (A3) to be equivalent to a generalized least-squares in which the filter gain function is determined by the dynamics of the system.

It must, therefore, be concluded that perhaps the only available technique which can accommodate a stochastic bilinear system (4.1), without undue computational requirements, is the maximum likelihood approach.

4.4 Properties of Maximum Likelihood Estimators

Given a set of N observations

$$Y = [y(1), y(2), \ldots\ldots, y(N)] , \tag{4.13}$$

and a set of parameters θ, a likelihood functional $L(Y, \theta)$ can be constructed. The maximum likelihood estimate of the parameters $\hat{\theta}$ is then defined as the maximizing θ of $L(Y, \theta)$ for a given Y (E1). Since the logarithm is a monotonic function of its argument, the maxima of the likelihood functional and its logarithm are identical;

$$\max_{\theta} L(Y, \theta) = \max_{\theta} Ln[L(Y, \theta)] . \tag{4.14}$$

A necessary condition for a (local) maximum is given by

$$\frac{\partial}{\partial \theta} (Ln(L(Y, \theta))) \Big|_{\theta = \hat{\theta}} = 0 . \tag{4.15}$$

The estimate obtained in this fashion is, in itself, a random variable and by defining

$$E[\hat{\theta}] = f(\theta) = \theta + b(\theta) , \tag{4.16}$$

$b(\theta)$ represents the bias in the estimate. If $b(\theta) = 0$, the estimate is said to be unbiased. Since the estimate is a random variable it is essential to have some idea about its mean and the dispersion around

it. In the case of a scalar parameter, a lower bound on the variance of the estimate can be derived (R1)

$$\mathrm{Var}\,[\hat{\theta}] \geq -\frac{\left(1+\frac{d}{d\theta}\,b(\theta)\right)^2}{E\left[\frac{\partial^2}{\partial\theta^2}\,\mathrm{Ln}(L(Y,\theta))\right]}. \tag{4.17}$$

Equation (4.17) is commonly referred to as the Cramer-Rao inequality. When the estimate is unbiased ($b(\theta) = 0$), the lower bound reduces to

$$\mathrm{Var}\,[\hat{\theta}] \geq -\frac{1}{E\left[\frac{\partial^2}{\partial\theta}\,(\mathrm{Ln}(L(Y,\theta))\right]}. \tag{4.18}$$

In the case where more than one parameter is involved a lower bound similar to (4.17) can be derived (E1) and when the estimate is unbiased

$$\mathrm{Cov}\,[\hat{\theta}] = E[(\hat{\theta}-\theta)\,(\hat{\theta}-\theta)^T] \geq J^{-1}, \tag{4.19}$$

for

$$J = -E\left[\left(\frac{\partial}{\partial\theta}\,\mathrm{Ln}(L(.,.))\right)\left(\frac{\partial}{\partial\theta}\,\mathrm{Ln}(L(.,.))\right)^T\right], \tag{4, 20}$$

where J is called the Fisher information matrix. We also define

$\hat{\theta}_k$ = estimate of θ after k observations,

then, if for an arbitrarily small ϵ

$$\lim_{k \to \infty} p\,[\,|\hat{\theta}_k - \theta| \leq \epsilon\,] = 1, \tag{4.21}$$

$\hat{\theta}_k$ is called a consistent estimate of θ. It has been shown (A4), in order for a process to be identifiable, it needs to have consistent estimates. A necessary condition for having a consistent estimate is that the Fisher information matrix be positive definite, thus, ensuring the existence of a unique extremum for $L(Y,\theta)$. It has been shown that as the number of observations increase, the maximum likelihood estimates become asymptotically unbiased, efficient (achieve the lower bound on the covariance), and have asymptotic normal behavior with a mean $\hat{\theta}$ and variance J^{-1}.

4.5 Maximum Likelihood Algorithm

In this section a maximum likelihood parameter estimation procedure is developed for the bilinear system described by (4.1) with the linear observation model (4.5).

Suppose the output of the system is observed up to time t_N and define the observation manifold

$$Y_N = [\,y(1),\ y(2),\ \ldots\ldots,\ y(N)]\ . \tag{4.22}$$

Also, define

θ = vector of unknown parameters ,

$$\theta = [\text{all the elements of } A, B_i, C, D_i, E, \text{ possibly } R \text{ and } Q]^T \tag{4.23}$$

where θ has q elements.

For the case under consideration a logical choice for the likelihood functional is the conditional probability density function $p(Y_N|\theta)$,

$$L(Y, \theta) = p(Y_N|\theta) . \tag{4.24}$$

As before $\hat{\theta}$ denotes the vector of the best parameter estimates, i.e. $\hat{\theta}$ is the maximizing value of θ in

$$\max_{\theta} p(Y_N|\theta) . \tag{4.25}$$

Since

$$p(Y_N|\theta) = p(y(1), y(2), \ldots\ldots, y(N)|\theta) , \tag{4.26}$$

using Baye's rule

$$p(Y_N|\theta) = p(y(N)|Y_{N-1}, \theta)\, p(Y_{N-1}|\theta) , \tag{4.27}$$

after successive application of Baye's rule

$$p(Y_N|\theta) = p(y(N)|Y_{N-1}, \theta)\, p(y(N-1)|Y_{N-2}, \theta) \ldots\ldots p(y(1)|\theta) , \tag{4.28}$$

or equivalently,

$$p(Y_N|\theta) = \prod_{i=1}^{N} p(y(i)|Y_{i-1}, \theta) . \tag{4.29}$$

As mentioned in the previous section, taking logarithm of (4.29) does not affect the maximization results and $\hat{\theta}$ can be obtained as the maximizing θ of

$$Ln(p(Y_N|\theta)) = Ln\left(\prod_{i=1}^{N} p(y(i)|Y_{i-1}, \theta)\right) , \tag{4.30}$$

or

$$Ln(p(Y_N|\theta)) = \sum_{i=1}^{N} Ln(p(y(i)|Y_{i-1}, \theta)). \tag{4.31}$$

Although greatly simplified, the parameter estimation problem still requires an expression for the conditional density function $p(y(i)|Y_{i-1}, \theta)$. The following theorem will shed some light on the subject.

<u>Theorem 4.1:</u> (Innovations Theorem)

Given the following observation model

$$y(k) = Z(k) + v(k) , \tag{4.32}$$

where v(k) is a white Gaussian sequence with a covariance

$$E[v(k)v^T(j)] = R(k)\delta(k-j) .$$

Then the innovations process is defined as

$$\nu(k) = y(k) - \hat{Z}(k|k-1) = \tilde{Z}(k|k-1) + v(k) , \tag{4.33}$$

where

$$\hat{Z}(k|k-1 = E[Z(k)|y(1), \ldots\ldots, y(k-1)] , \tag{4.34}$$

$$\tilde{Z}(k|k-1) = Z(k) - \hat{Z}(k|k-1) , \tag{4.35}$$

is a white sequence with

$$E[\nu(k)] = E[y(k) - \hat{Z}(k|k-1)] = 0 , \tag{4.36}$$

and

$$E[\nu(k)\nu^T(j)] = P_{\nu\nu}(k)\delta(k-j) , \tag{4.37}$$

where

$$P_{\nu\nu}(k) = E[(y(k) - \hat{Z}(k|k-1)(y(k) - \hat{Z}(k|k-1))^T] , \tag{4.38}$$

$$P_{\nu\nu}(k) = E[(\tilde{Z}(k|k-1) + v(k))(\tilde{Z}(k|k-1) + v(k))^T]$$

$$= P_{\tilde{Z}\tilde{Z}}(k) + R(k) . \tag{4.39}$$

For a proof of this theorem see Kailath (K1-2). Similar results are also available for the continuous case (K2) where the innovations process is also shown to be Gaussian. The lack of continuity for a discrete observation process renders the Gaussian assumption invalid, in general. In cases where the observations are "dense enough", however, the innovations process comes arbitrarily close to that of

a continuous process. This implies that the innovations process can be approximated by a normal density function if the sampling rate is high and in such cases the density function has a zero mean and $P_{\nu}(.)$ for its variance. In any case, the innovations process contains all the useful information contained in the observations,

$$y(k) \equiv \nu(k) \ . \tag{4.40}$$

Application of Theorem 4.1 to the problem at hand yields

$$y(k) = H(k)x(k) + v(k) \ , \tag{4.41}$$

where the innovations process is defined as

$$\nu(k) = y(k) - E[H(k)x(k) | Y_{k-1}] \ , \tag{4.42}$$

but

$$E[H(k)x(k) | Y_{k-1}] = H(k)\hat{x}(k|k-1) \ ,$$

as a result,

$$\nu(k) = y(k) - H(k)\hat{x}(k|k-1) \ . \tag{4.43}$$

The innovations process has a zero mean and a covariance given by

$$P_{\nu}(k) = E[(H(k)\tilde{x}(k|k-1) + v(k))(H(k)\tilde{x}(k|k-1) + v(k))^T]$$

$$= H(k)\ E[\tilde{x}\tilde{x}^T]\ H^T(k) + R(k) \ . \tag{4.44}$$

but from Chapter II

$$P(k) = E[\tilde{x}\tilde{x}^T] = P(k|k-1) ,$$

thus,

$$P_{\nu\nu}(k) = H(k)P(k)H^T(k) + R(k) . \tag{4.45}$$

Since the observation noise is Gaussian and the sampling rate is assumed to be high enough, the innovations process can then be approximated by the following normal density function

$$p(\nu(k)) = \frac{1}{(2\pi)^{\frac{p^1}{2}} \; |P_{\nu\nu}|^{1/2}} \text{EXP}\left(-\frac{1}{2}\nu^T(k)P_{\nu\nu}^{-1}(k)\nu(k)\right) , \tag{4.46}$$

where p^1 corresponds to the dimension of the observation vector $y(k)$.

Note that (4.46) is actually a conditional density function since $\nu(k)$ depends on the observations up to the last time interval

$$p(\nu(k)) = p(\nu(k)|Y_{k-1}) . \tag{4.47}$$

A consequence of Theorem 4.1 was that the innovations process and the observations contain the same information and it is, therefore, reasonable to assume that

$$p(\nu(k)|Y_{k-1}, \Theta) = p(y(k)|Y_{k-1}, \Theta) . \tag{4.48}$$

Equations (4.46) and (4.48), therefore, provide the conditional density function required in the evaluation of the likelihood functional (4.31) if $\hat{x}(k|k-1)$ and $P(k|k-1)$ required in (4.43-4) are available.

The estimation problem for obtaining $\hat{x}(k|k-1)$ and $P(k|k-1)$ was addressed in Chapter III and it was shown that these conditional moments are the solutions to a set of two ordinary differential equations. However, evaluation of these differential equations depends on the determination of the initial conditions $\hat{x}(k-1|k-1)$ and $P(k-1|k-1)$ which involves non-existing closed form solutions to a set of difference equations. The sub-optimal filters proposed in Chapter III can, however, be employed to obtain an approximate solution to the bilinear system parameter identification problem.

The approach suggested here is similar to that proposed by Mehra (M3-4) in principle. Mehra, however, does not allow for cases where the noise has a multiplicative effect in his derivation. The state estimation and parameter identification problems can, therefore, be simultaneously solved in a recursive manner; i.e. parameter estimates in the previous iteration are used in the filter construction and the process is repeated until the scheme converges. Some error in the estimation of the parameters is inevitable since a sub-optimal filter is employed.

The steps involved in the maximum likelihood parameter

estimation procedure are as follows.

Given a set of N observations define

$$M(j) = E[\nu(j)\nu^T(j)] \ . \tag{4.49}$$

Then from (4.31), $\hat{\theta}$ (maximum likelihood estimate of the parameters) is the maximizing θ of

$$\max_{\theta} \sum_{i=1}^{N} \mathrm{Ln} \frac{1}{(2\pi)^{\frac{p}{2}} |M(k)|^{1/2}} \mathrm{EXP}(-\frac{1}{2} \nu^T(i)M^{-1}(i)\nu(i)] \tag{4.50}$$

$$= \max_{\theta} \sum_{i=1}^{N} -\frac{1}{2} (p^1 \mathrm{Ln}(2\pi) + \mathrm{Ln}|M(i)| + \nu^T(i)M^{-1}(i)\nu(i)) \ .$$

Deleting constant terms, $\hat{\theta}$ is the maximizing θ of the newly defined likelihood functional

$$L(Y, \theta) = \sum_{i=1}^{N} -\frac{1}{2} (\mathrm{Ln}|M(i)| + \nu^T(i)M^{-1}(i)\nu(i)) \ , \tag{4.51}$$

where the maximization has to be performed subject to

$$\dot{\hat{x}}(t|i-1) = (A(t) + \sum_{j=1}^{m} B_j(t)u_j(t))x(t) + C(t)u(t), \tag{4.52}$$

$$\begin{aligned} \dot{P}(t|i-1) = {} & (A(t) + \sum_{j=1}^{m} B_j(t)u_j(t))P(t|i-1) \\ & + P(t|i-1)\,(A(t) + \sum_{j=1}^{m} B_j(t)u_j(t))^T + \widehat{GQG}^T, \end{aligned} \tag{4.53}$$

and $\widehat{GQG}^T$ is defined by (3.25),

$$\nu(i) = y(i) - H(i)\hat{x}(i|i-1) , \tag{4.54}$$

$$M(i) = H(i)P(i|i-1)H^T(i) + R(i) , \tag{4.55}$$

$$\hat{x}(i|i) = \hat{x}(i|i-1) + K(i)\nu(i) , \tag{4.56}$$

$$P(i|i) = P(i|i-1) - K(i)H(i)P(i|i-1) , \tag{4.57}$$

$$K(i) = P(i|i-1)H^T(i)[H(i)P(i|i-1)H^T(i) + R(i)]^{-1}. \tag{4.58}$$

This constrained maximization problem is, of course, non-trivial and the subject of discussion in the next section.

4.6 Maximization Procedures

The constrained maximization problem of (4.51) subject to (4.52-58) is actually an extremely difficult one. Often, the choice of the optimization procedure will have quite an effect on the final estimates of the parameters, particularly since the shape of the likelihood functional hill might become quite irregular as the number of the parameters increase. There is an enormous amount of literature available on optimization techniques; see for example, (A1), (C2), (F2). In general, the optimization algorithms suitable for our purposes can be divided into two categories (A5), (E1):

i) Direct search methods.

ii) Gradient methods.

The choice of a particular optimization method is largely dependent on the problem at hand and a short summary of three procedures is provided here.

4.6.1 Relaxation Method

In this method the likelihood functional is maximized with respect to each parameter individually. The procedure is then repeated until convergence occurs. Although simple to implement, computationally the relaxation method is known to experience convergence problems (A1). This maximization algorithm is quite appealing for our purposes when there are only a few parameters to be estimated.

4.6.2 Steepest Descent Method

In this method the gradient of the likelihood functional with respect to the vector of the parameters Θ is computed

$$\frac{\partial}{\partial \Theta} L(.,.) = [\frac{\partial}{\partial \Theta_1} L, \frac{\partial}{\partial \Theta_2} L, \ldots\ldots, \frac{\partial}{\partial \Theta_q} L]^T. \quad (4.59)$$

At every iteration the new parameter estimate is defined by

$$\theta(i+1) = \theta(i) + \Gamma(i)\frac{\partial L}{\partial \theta}\bigg|_{\theta=\theta(i)} , \tag{4.60}$$

where $\Gamma(i)$ is appropriately chosen to maximize the likelihood functional along the gradient vector. This approach is known to have very slow convergence properties near the optimum.

4.6.3 Newton-Raphson Method

The Newton-Raphson algorithm requires the calculation of both the gradient and second order partial derivatives with respect to the parameters. At each iteration the new estimates are obtained from

$$\theta(i+1) = \theta(i) - \left[\frac{\partial^2 L}{\partial \theta \partial \theta^T}\bigg|_{\theta=\theta(i)}\right]^{-1} \left(\frac{\partial L}{\partial \theta}\bigg|_{\theta=\theta(i)}\right) . \tag{4.61}$$

For a large number of parameters the calculation of the second partial derivative matrix and its inverse can be computationally quite time consuming. An added bonus, however, is that the second partials provide the Fisher information matrix, which is the lower bound on the covariance of the estimates of the parameters. To obtain correct results it is absolutely essential that the Hessian (matrix of second order partials) be negative definite to ensure the existence of a

maxima. At points where the computed Hessian becomes non-negative definite, the following adjustment can be made

$$\text{Hessian} = \left(\frac{\partial^2 L}{\partial\theta\partial\theta^T}\right|_{\theta=\theta(i)} - mI) , \tag{4.62}$$

where I is the identity matrix and m is a large enough positive number that would make the Hessian negative definite. The convergence of this approach is faster than the previous two methods around the true optimum. However, having to make sure that the Hessian is always negative definite is an added computational burden. In cases where the a priori estimates of the parameters are not very reliable, it might be a better idea to start off with another technique and then switch to the Newton-Raphson algorithm as the estimates start to converge.

Since both the steepest descent and Newton-Raphson methods require the calculation of the gradient vector, the derivation of general expressions for the gradient is provided below.

By definition

$$L(Y_N, \theta) = -\frac{1}{2}\sum_{i=1}^{N} \nu^T(i)M^{-1}(i)\nu(i) + \text{Ln}|M(i)| , \tag{4.63}$$

$$\frac{\partial L}{\partial\theta} = [\frac{\partial L}{\partial\theta_1}, \ldots\ldots, \frac{\partial L}{\partial\theta_q}]^T . \tag{4.64}$$

Before proceeding with the derivation of the gradient equations, a few remarks are in order.

Remark 1:

The covariance of the innovations process is symmetrical

$$E[\nu(i)\nu^T(i)] = E[\nu(i)\nu^T(i)]^T , \tag{4.65}$$

or

$$M(i) = M^T(i), \tag{4.66}$$

which is the property of the covariance function.

Remark 2:

$$\frac{\partial}{\partial \theta_k} \operatorname{Ln}|M(i)| = \operatorname{tr}\left(M^{-1}(i)\frac{\partial}{\partial \theta_k} M(i)\right). \tag{4.67}$$

To determine the gradient

$$\begin{aligned}\frac{\partial}{\partial \theta_k} L = -\frac{1}{2}\sum_{i=1}^{N} \Big[& \nu^T(i)M^{-1}(i)\frac{\partial}{\partial \theta_k}\nu(i) + \frac{\partial}{\partial \theta_k}\nu^T(i)M^{-1}(i)\nu(i) \\ & - \nu^T(i)M^{-1}(i)\frac{\partial}{\partial \theta_k} M(i)M^{-1}(i)\nu(i) \\ & + \operatorname{tr}\left(M^{-1}(i)\frac{\partial}{\partial \theta_k} M(i)\right)\Big] .\end{aligned} \tag{4.68}$$

Since M(i) is symmetric

$$\frac{\partial}{\partial \theta_k} \nu^T(i)M^{-1}(i)\nu(i) = \nu^T(i)M^{-1}(i)\frac{\partial}{\partial \theta_k} \nu(i) , \tag{4.69}$$

$$\frac{\partial}{\partial \theta_k} L = -\frac{1}{2} \sum_{i=1}^{N} [2\, \nu^T(i)M^{-1}(i)\frac{\partial}{\partial \theta}\nu(i) - \nu^T(i)M^{-1}(i)\frac{\partial}{\partial \theta_k} M(i)M^{-1}(i)\nu(i)$$

$$+ \operatorname{tr} (M^{-1}(i)\frac{\partial}{\partial \theta_k} M(i))] . \tag{4.70}$$

But, by definition

$$\nu(i) = y(i) - H(i)\hat{x}(i|i-1) \tag{4.71}$$

hence,

$$\frac{\partial}{\partial \theta_k} \nu(i) = -[\frac{\partial}{\partial \theta_k}H(i)\hat{x}(i|i-1) + H(i)\frac{\partial}{\partial \theta_k}\hat{x}(i|i-1)], \tag{4.72}$$

and with

$$M(i) = H(i)P(i|i-1)H^T(i) + R(i) ,$$

$$\frac{\partial}{\partial \theta_k}M(i) = \frac{\partial}{\partial \theta_k} H(i)P(i|i-1)H^T(i) + H(i)\frac{\partial}{\partial \theta_k} P(i|i-1)H^T(i)$$

$$+ H(i)P(i|i-1)\frac{\partial}{\partial \theta_k} H^T(i) + \frac{\partial}{\partial \theta_k}R(i) . \tag{4.73}$$

Since the gradient of (4.72-3) also depends on $\frac{\partial}{\partial \theta_k}\hat{x}(i|i-1)$

and $\frac{\partial}{\partial \theta_k} P(i|i-1)$, these partials are derived from the filter equations (4.52-8). Define

$$F(t) = A(t) + \sum_{j=1}^{m} B_j(t)u_j(t) , \tag{4.74}$$

then

$$\frac{\partial}{\partial \theta_k} \dot{\hat{x}}(t|i-1) = \frac{\partial}{\partial \theta_k} F(t)\hat{x}(t|i-1) + F(t)\frac{\partial}{\partial \theta_k} \hat{x}(t|i-1) + \frac{\partial}{\partial \theta_k} C(t)u(t) , \tag{4.75}$$

and

$$\frac{\partial}{\partial \theta_k} P(t|i-1) = \frac{\partial}{\partial \theta_k} F(t)P(t|i-1) + F(t)\frac{\partial}{\partial \theta_k} P(t|i-1) + \frac{\partial}{\partial \theta_k} P(t|i-1)F^T(t) + P(t|i-1)\frac{\partial}{\partial \theta_k} F^T(t) + \frac{\partial}{\partial \theta_k} \widehat{GQG}^T, \tag{4.76}$$

$$\frac{\partial}{\partial \theta_k} \hat{x}(i|i) = \frac{\partial}{\partial \theta_k} \hat{x}(i|i-1) + \frac{\partial}{\partial \theta_k} K(i)\nu(i) + K(i)\frac{\partial}{\partial \theta_k} \nu(i) , \tag{4.77}$$

$$\frac{\partial}{\partial \theta_k} P(i|i) = -\frac{\partial}{\partial \theta_k} K(i)H(i)P(i|i-1) - K(i)\frac{\partial}{\partial \theta_k} H(i)P(i|i-1) - P(i|i-1)H^T(i)M^{-1}(i)\frac{\partial}{\partial \theta_k} M(i)M^{-1}(i). \tag{4.79}$$

Equations (4.68-79) now provide the necessary expressions for the recursive computation of the gradient vector. In the Newton-Raphson method it is also required to compute the second partials

of the likelihood functional. Since the computations can be very time consuming, the second order terms are neglected (M4).

$$\frac{\partial}{\partial \theta_k \partial \theta_\ell} L = \sum_{i=1}^{N} \frac{\partial}{\partial \theta_k} \nu^T(i) M^{-1}(i) \frac{\partial}{\partial \theta_\ell} \nu(i) - \nu^T(i) M^{-1}(i) \frac{\partial}{\partial \theta_k} M(i) M^{-1}(i) \frac{\partial}{\partial \theta_\ell} \nu(i)$$

$$- \nu^T(i) M^{-1}(i) \frac{\partial}{\partial \theta_\ell} M(i) M^{-1}(i) \frac{\partial}{\partial \theta_k} \nu(i) \tag{4.80}$$

$$- \frac{1}{2} \operatorname{tr} \left(M^{-1}(i) \frac{\partial}{\partial \theta_k} M(i) M^{-1}(i) \frac{\partial}{\partial \theta_\ell} M(i) \right) .$$

Equation (4.80) in conjunction with (4.65-79) provides an estimate of the Hessian matrix. If the Hessian is negative definite the estimates of the parameters can be improved by using

$$\theta(i+1) = \theta(i) - \left(\frac{\partial^2 L}{\partial \theta \partial \theta^T} \bigg|_{\theta=\theta(i)} \right)^{-1} \left(\frac{\partial}{\partial \theta} L \bigg|_{\theta=\theta(i)} \right) . \tag{4.81}$$

The parameter identification algorithm discussed in this chapter must be tested under simulation conditions to provide computational evidence of its feasibility.

V. SIMULATION OF STOCHASTIC BILINEAR SYSTEMS

5.1 Introduction

In this chapter the problem of simulating solutions to stochastic bilinear differential equations is discussed and the effectiveness of the proposed filter of Chapter III and the parameter estimation algorithm of Chapter IV are illustrated at the hand of a few computational examples.

It is intuitively obvious that the value of any approximate method should only be judged by its performance in computations and the degree of difficulty involved in its implementation. It is, therefore, essential to examine the performance of the sub-optimal state estimator of Chapter III as well as the maximum likelihood parameter estimator of Chapter IV (based on the filter) under simulation conditions.

Any of the proposed algorithms require input-output information for a stochastic bilinear system. Such data can be either collected from a physical process known to exhibit bilinear properties or simulated by a stochastic bilinear differential equation. Since no such physical system data was available to us the simulation route had to be followed. Solutions to Ito stochastic bilinear differential equations can be generated using standard numerical techniques available for

ordinary differential equations, but as mentioned in Chapter II, Ito equations do not follow the rules of ordinary calculus and certain adjustments or corrections are necessary to any numerical method before it can be applied to stochastic cases (R2). There is, however, a great deal of uncertainty about the convergence of such techniques and even in cases where covergence is assured there is no guarantee that the limiting solution is that of the original stochastic differential equation (B3). To avoid erroneous results we restricted ourselves to cases where a closed form solution to the Ito bilinear differential equation is available. This in no way implies that the proposed state estimation and parameter identification techniques applicability are limited to such cases. This precaution simply ensures the data generated is indeed that of a stochastic bilinear system.

5.2 Closed Form Solution

A closed form solution to a stochastic bilinear differential equation of the form

$$dx(t) = ((A(t) + \sum_{i=1}^{m} B_i(t)u_i(t))x(t) + C(t)u(t))\, dt + \sum_{i=1}^{m} D_i(t)x(t)dw_i(t) + E(t)dw(t) , \tag{5.1}$$

is available only in certain cases as outlined on the following pages.

5.2.1 Scalar Cases

Only results of direct relevance to previous discussions are summarized here and for a comprehensive study of scalar stochastic bilinear differential equations the reader is referred to (A2), (D3), (G2).

The following theorem provides the conditions for the existence and uniqueness of solutions to scalar differential equations.

Theorem 5.1:

Consider

$$dx(t) = f(x, t)dt + g(x, t)dw(t) , \tag{5.2}$$

where

$$x(t_0) = x(0) \neq 0 , \quad t \geq t_0 \geq 0 ,$$

$$\int_{t_0}^{t_N} |f(x, t)|dt < \infty \quad , \quad \int_{t_0}^{t_N} |g(n, t)|dt < \infty \quad ,$$

and $x(0)$ is independent of all increments $w(t) - w(t_0)$, there then exists a unique solution $x(t)$ continuous with probability 1, if the functions $f(x, t)$ and $g(x, t)$ satisfy jointly:

1) the uniform Lipschitz condition in x,

$$|f(x, t) - f(y, t)| + |g(x, t) - g(y, t)| < C_1 |x-y| , \tag{5.3}$$

2) Fillipov's growth condition (F1)

$$|f(x,t)|^2 + |g(x,t)|^2 \leq C_2(1+|x|^2) . \tag{5.4}$$

For a proof see (G2).

It is simple to prove that for bounded parameters and inputs a scalar bilinear stochastic differential equation does meet the conditions of Theorem 5.1 and a unique solution is assured.

Theorem 5.2:

Consider the homogeneous bilinear Ito differential equation

$$dx(t) = b(t)x(t)dt + e(t)x(t)dw(t) \tag{5.5}$$

$$x(t_0) = x(0) \neq 0 , \qquad t \geq t_0 \geq 0 ,$$

if the conditions of Theorem (5.1) hold, (5.5) has a closed form solution described by

$$x(t) = x(0)\ \mathrm{EXP}[\int_{t_0}^{t} (b(s) - \frac{1}{2} e^2(s))ds + \int_{t_0}^{t} e(s)dw(s)] . \tag{5.6}$$

A proof can be constructed using the transformation $y(t) = \mathrm{Ln}(x(t))$ and using Ito's differential rule (G2).

Equation (5.6) provides a closed form solution to a homogeneous stochastic bilinear differential equation and reveals that such a system is strictly positive (negative) for all time if $x(0)$ is positive (negative).

This property has been used in the modelling of economic systems (F3). It should be noted that the integrals involved in (5.6) are no longer of the Ito type and they can be evaluated using standard numerical integration techniques. Equation (5.6) also provides a means for the determination of a discrete equation equivalent to (5.5). By definition

$$x(t_{i+1}) = x(0)\ \mathrm{EXP}\left[\int_{t_0}^{t_i} (b(s) - \frac{1}{2} e^2(s))ds + \int_{t_0}^{t_i} e(s)dw(s)\right] \tag{5.7}$$

$$\times\ \mathrm{EXP}\left[\int_{t_i}^{t_{i+1}} (b(s) - \frac{1}{2} e^2(s))\, ds + \int_{t_i}^{t_{i+1}} e(s)dw(s)\right],$$

thus,

$$x(t_{i+1}) = x(t_i)\mathrm{EXP}\left[\int_{t_i}^{t_{i+1}} (b(s) - \frac{1}{2} e^2(s))ds + \int_{t_i}^{t_{i+1}} e(s)dw(s)\right]. \tag{5.8}$$

Equation (5.8) does not represent a discrete bilinear equation, but it provides an important relationship that can be used in the recursive generation of a closed form solution to (5.5).

<u>Theorem 5.3:</u>

Consider the inhomogeneous stochastic bilinear system described by

$$dx(t) = (a(t) + b(t)x(t))\, dt + (c(t) + e(t)x(t))\, dw\,(t), \quad (5.9)$$

$$x(t_0) = x(0) \neq 0\,, \qquad t \geq t_0 \geq 0\,,$$

assuming conditions of Theorem 5.1 hold, (5.9) has a closed form solution given by

$$x(t) = \mathrm{EXP}\{\int_{t_0}^{t} (b(s) - \frac{1}{2} e^2(s))\, ds + \int_{t_0}^{t} e(s)dw(s)\}$$

$$X\,[x(0) + \int_{t_0}^{t} \mathrm{EXP}\{ -\int_{t_0}^{s} (b(u) - \frac{1}{2} e^2(u))\, du - \int_{t_0}^{s} e(u)dw(u)\}$$

$$X\,[a(s) - c(s)\; e(s)]\, ds \qquad (5.10)$$

$$+ \int_{t_0}^{t} \mathrm{EXP}\{ -\int_{t_0}^{s} (b(u) - \frac{1}{2} e^2(u))du - \int_{t_0}^{s} e(u)dw(u)\}\, e(s)dw(s)]\,.$$

For a proof see (G2).

Derivation of an equivalent discrete model for (5.10) is still possible but due to unnecessary complex notation it will not be presented here.

5.2.2 Vector Cases

The closed form solutions for the vector cases are rarely

available. However, under certain parametric conditions such solutions can be obtained.

Theorem 5.4:

If any finite collection of real square normal matrices $A_0, A_1, \ldots\ldots, A_m$ commute pairwise, then there exists a bounded nonsingular matrix P such that $P^{-1}A_iP$'s are diagonal for all $i = 0, \ldots, m$.

For a proof see (G1).

Consider the homogeneous bilinear system

$$dx(t) = (\bar{A}(t) + \sum_{i=1}^{m} \bar{B}(t)u_i(t))\, x(t)dt + \sum_{i=1}^{m} \bar{C}_i(t)x(t)dw_i(t) , \qquad (5.11)$$

$$x(t_0) = x(0) \neq 0 , \qquad t \geq t_0 \geq 0 ,$$

if the matricies $\bar{A}$, $\bar{B}_i$, and $\bar{C}_i$ commute pairwise, then according to Theorem 5.4, they can be diagonalized simultaneously. Define

$$A = P^{-1}AP , \quad B_i = P^{-1}B_iP , \quad C_i = P^{-1}C_iP ,$$

$$y(t) = Px(t) , \qquad (5.12)$$

thus

$$dy(t) = (A(t) + \sum_{i=1}^{m} B(t)u_i(t))\, x(t)dt + \sum_{i=1}^{m} C(t)x(t)dw_i(t) . \qquad (5.13)$$

Since (5.13) involves n decoupled scalar stochastic differential equations it can be solved using (5.6) and the inverse transformation

$$x(t) = P^{-1}y(t) , \tag{5.14}$$

will provide the solution for the vector $x(t)$.

Lemma 5.1:

A Lie algebra of real matricies $\mathcal{L}$ is solvable iff there exists a nonsingular matrix P such that $P^{-1}A_iP$ is upper triangular for all $A_i \varepsilon \mathcal{L}$.

For proof see (S2).

Consequently, if A, B_i's, and C_i's in (5.11) do generate a solvable Lie algebra, they can be upper triangularized simultaneously. Using this property a closed form solution for the two dimensional solvable cases with a single input can be obtained (F3).

For example, consider the two dimensional system

$$dx(t) = \bar{A}(t)\, x(t)dt + \bar{B}(t)x(t)dw(t) , \tag{5.15}$$

where $\bar{A}(t)$ and $\bar{B}(t)$ generate a solvable Lie algebra and can be upper triangularized according to Lemma 5.1. Define

$$y(t) = P(t)\, x(t) ,$$

$$A(t) = P^{-1}(t)\bar{A}(t)P(t) \quad , \quad B(t) = P^{-1}(t)\bar{B}(t)P(t) ,$$

thus,

$$dy(t) = A(t)x(t)dt + B(t)x(t)dw(t) , \tag{5.16}$$

where

$$A(t) = \begin{bmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{bmatrix} , \quad B(t) = \begin{bmatrix} b_{11} & b_{12} \\ 0 & b_{22} \end{bmatrix} .$$

A solution for $y_2(t)$ can be easily obtained using the scalar homogeneous results of (5.6). Now define the quotient process

$$z(t) = \frac{y_1(t)}{y_2(t)} . \tag{5.17}$$

Since the initial conditions were nonzero, there is danger of degeneracy. Using Ito's differential rule a scalar inhomogeneous stochastic equation in terms of z(t) can be derived and then solved using (5.10). Consequently,

$$y_1(t) = z(t)y_2(t) ,$$

$$x(t) = P^{-1}(t)y(t) ,$$

and the closed form solution for x(t) is obtained.

5.3 Importance of Data Generation

As mentioned earlier all the data tested had to be generated on a

digital computer. It was, however, of absolute importance to make certain the noise sequences generated by the computer indeed met the assumptions made about them. Let us return to the homogeneous scalar case

$$dx(t) = b(t)x(t)dt + e(t)x(t)dw(t) \tag{5.18}$$

$$x(t_0) = x(0) \neq 0. \quad , \quad t \geq t_0 \geq 0. \; .$$

Taking expectation of (5.18),

$$\frac{d}{dt}\hat{x}(t) = b(t)\hat{x}(t) \; , \tag{5.19}$$

or equivalently,

$$\hat{x}(t) = x(0)\ \mathrm{EXP}\left[\int_{t_0}^{t} b(s)ds\right] . \tag{5.20}$$

Recall the closed form solution to (5.18),

$$x(t) = x(0)\ \mathrm{EXP}\left[\int_{t_0}^{t} (b(s) - \frac{1}{2}e^2(s))ds + \int_{t_0}^{t} e(s)dw(s)\right] , \tag{5.21}$$

taking expectation of (5.21),

$$\hat{x}(t) = x(0)\ \mathrm{EXP}\left[\int_{t_0}^{t} (b(s) - \frac{1}{2}e^2(s))ds\right] E\left[\mathrm{EXP}\left[\int_{t_0}^{t} e(s)dw(s)\right]\right], \tag{5.22}$$

where

$$E[\,EXP[\int_{t_0}^{t} e(s)dw(s)]] = E[\,1 + \int_{t_0}^{t} e(s)dw(s) + \frac{1}{2}(\int_{t_0}^{t} e(s)dw(s))^2 + \ldots\,]. \tag{5.23}$$

Since w(s) is a zero mean Wiener process all odd terms vanish and

$$E[EXP[\int_{t_0}^{t} e(s)dw(s)]] = EXP[\int_{t_0}^{t} \frac{1}{2} e^2(s)ds] \,. \tag{5.24}$$

It is now clear as to why the correction term must be included in the closed form solution. It should be noted that (5.24) is only true if the process w(s) has a zero mean. In generating the noise sequences on a digital computer, care must be taken to ensure the zero mean property assumed about it. If the mean is not corrected the expected value of the closed form solution will be different from that of the original equation and serious problems in the parameter estimation algorithms occur.

5.4 Simulation Results

In this section we present a few of the many tests that were performed to determine the effectiveness of the proposed state and parameter estimation algorithms. Figure 5.1 provides a flow

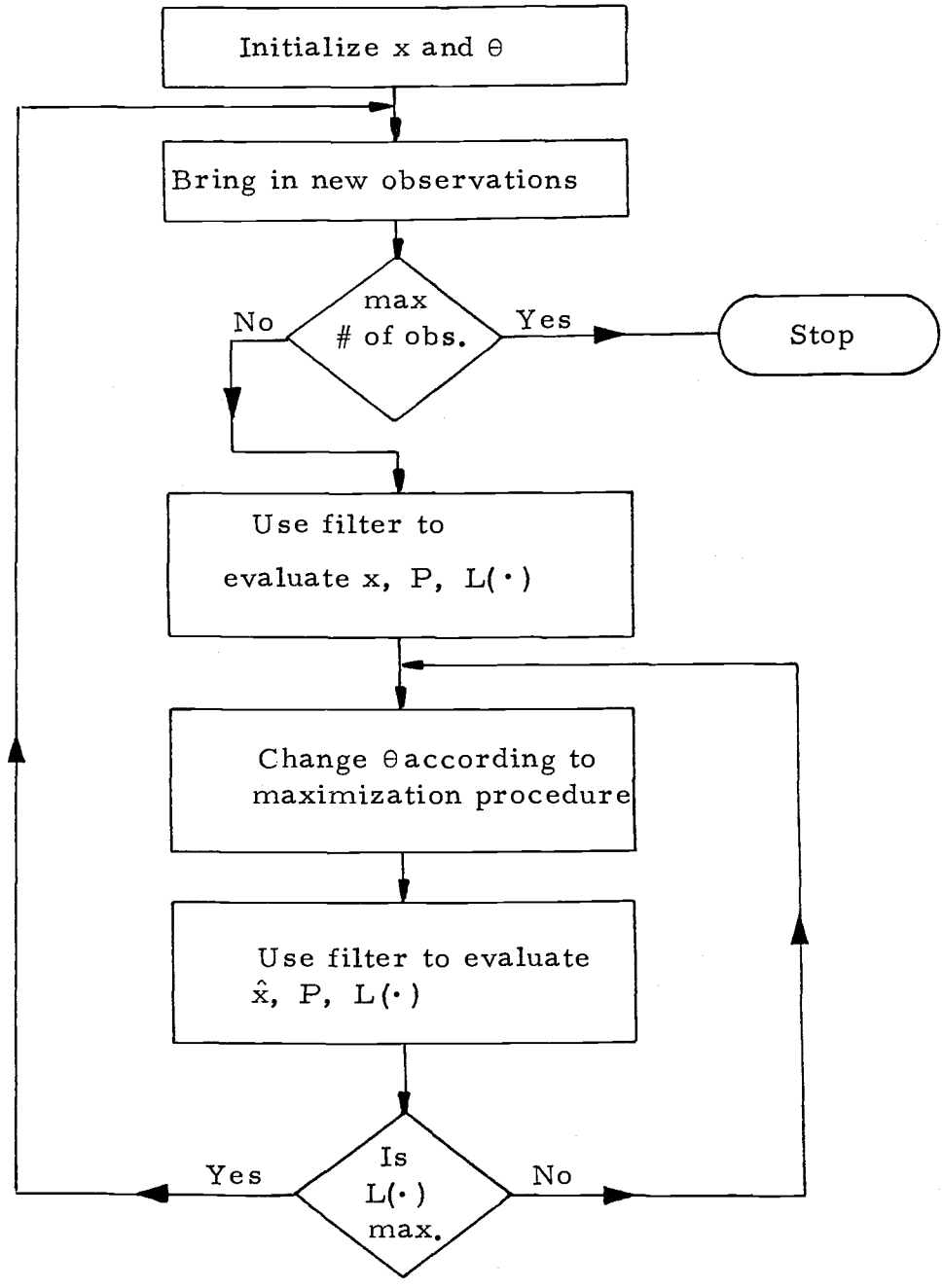


Figure 5.1. Flow chart of state and parameter estimation procedure.

chart of the steps involved in the identification procedure.

Example 1:

Consider the scalar homogeneous bilinear model

$$dx(t) = au(t)\, x(t)dt + bx(t)\, dw(t), \tag{5.25}$$

$$x(0) = 10.,$$

with the observation model

$$y(k) = x(k) + v(k)\ . \tag{5.26}$$

The data was generated for a = -2., b = 1., Q = 0.09., R = 1.0, Dt = 0.01, and u(t) = 3. Figure 5.2 provides the plots of the exact solution as well as the estimates of the state using the sub-optimal filter of Chapter III. It is easy to see that even in the presence of quite a large observation noise the filter tracks the actual state very closely.

Example 2:

Consider the system of example two

$$dx(t) = a\, u(t)x(t)dt + bx(t)dw(t)\ , \tag{5.27}$$

$$y(k) = x(k) + v(k)\ . \tag{5.28}$$

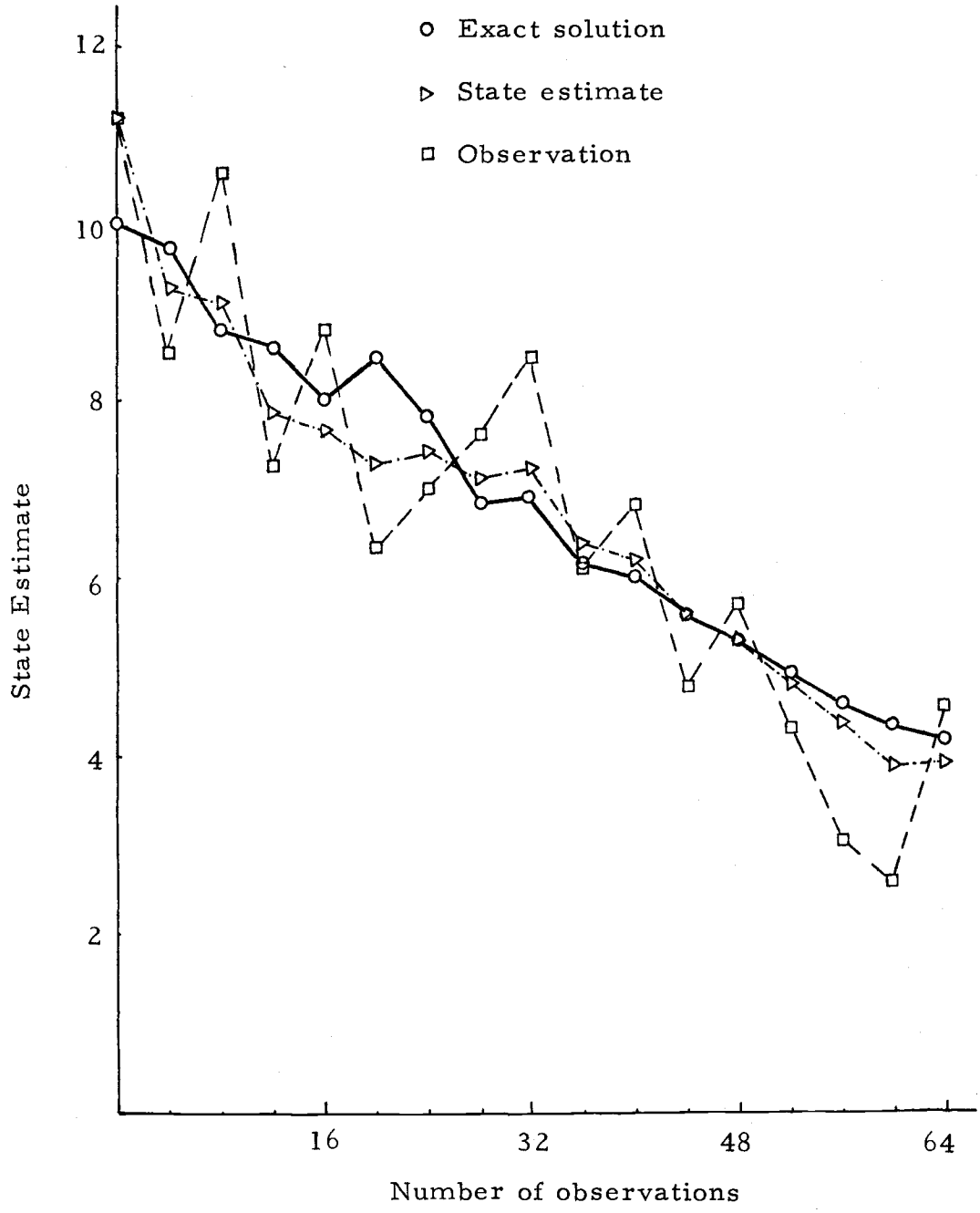


Figure 5.2. State estimates for a scalar homogeneous bilinear system.

The data was generated for the following parameters

$$a = -2.0, \quad b = 1.0, \quad R = 0.16, \quad Q = 1.0,$$

$$Dt = 0.01, \quad x(0) = 10., \; u(t) = 3.$$

First,the noise generated was not corrected for a zero mean and Figure 5.3 shows the estimates for a and b as being incorrect. The data was then generated for the same parameters but the noise was corrected and Figure 5.3 reveals that the estimates of a and b improved considerably. The same problem was simulated for a time varying input $u(t) = 3e^{t}$ and the parameter estimates did not differ much from the constant input case.

The identification algorithm recursively provides the estimates of the state. Figure 5.4 presents the result of such a filter after the parameters were estimated for 100 observations. The parameter estimates of Figure 5.3 are very close to the true value and the state estimates of Figure 5.4 follow the exact solution quite accurately.

Example 3:

Consider the inhomogeneous scalar bilinear system

$$dx(t) = (au(t)x(t) + cu(t))\, dt + (bx(t) + e)\, dw(t) \,, \qquad (5.29)$$

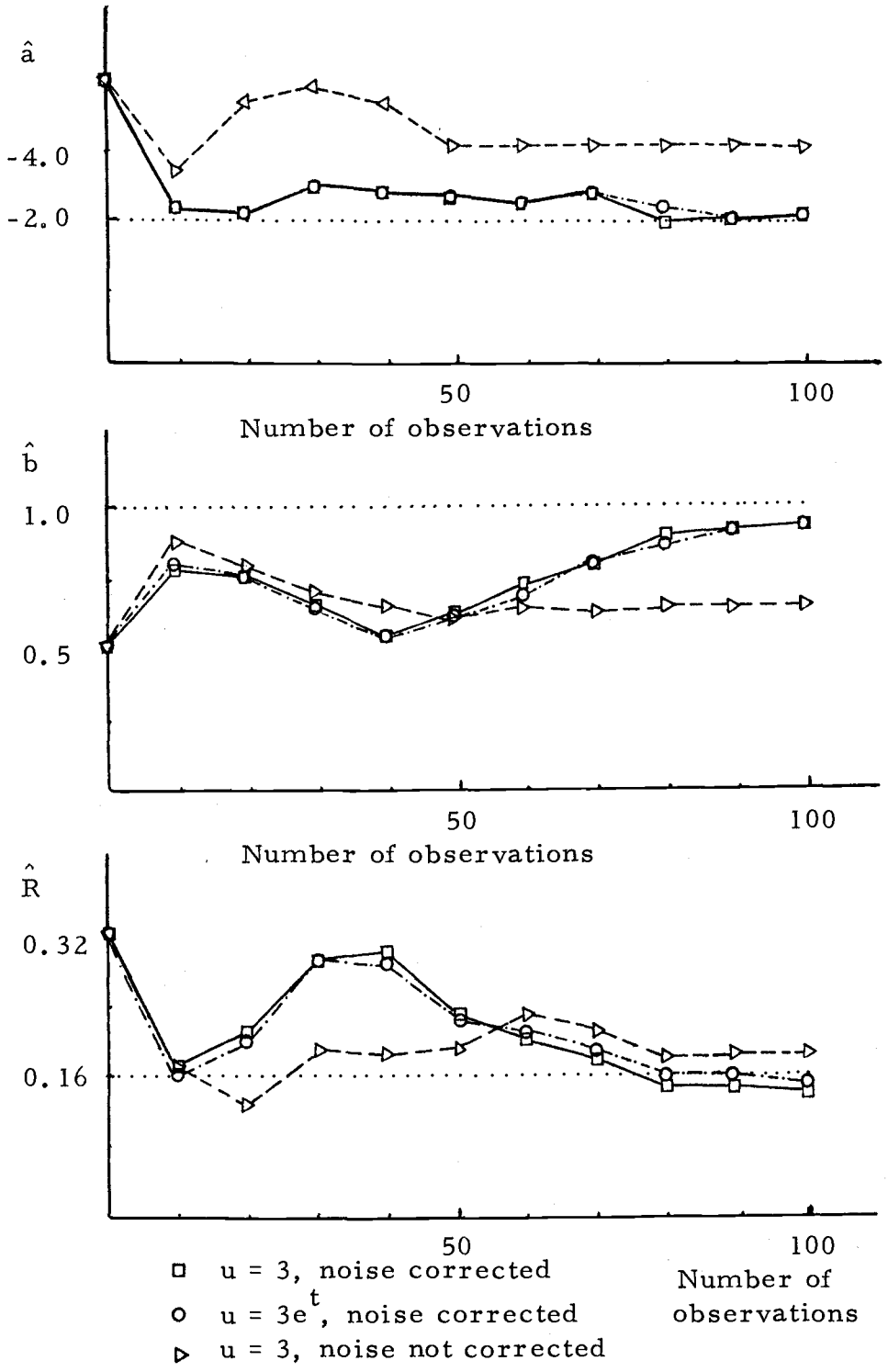


Figure 5.3. Parameter estimates for a scalar bilinear system.

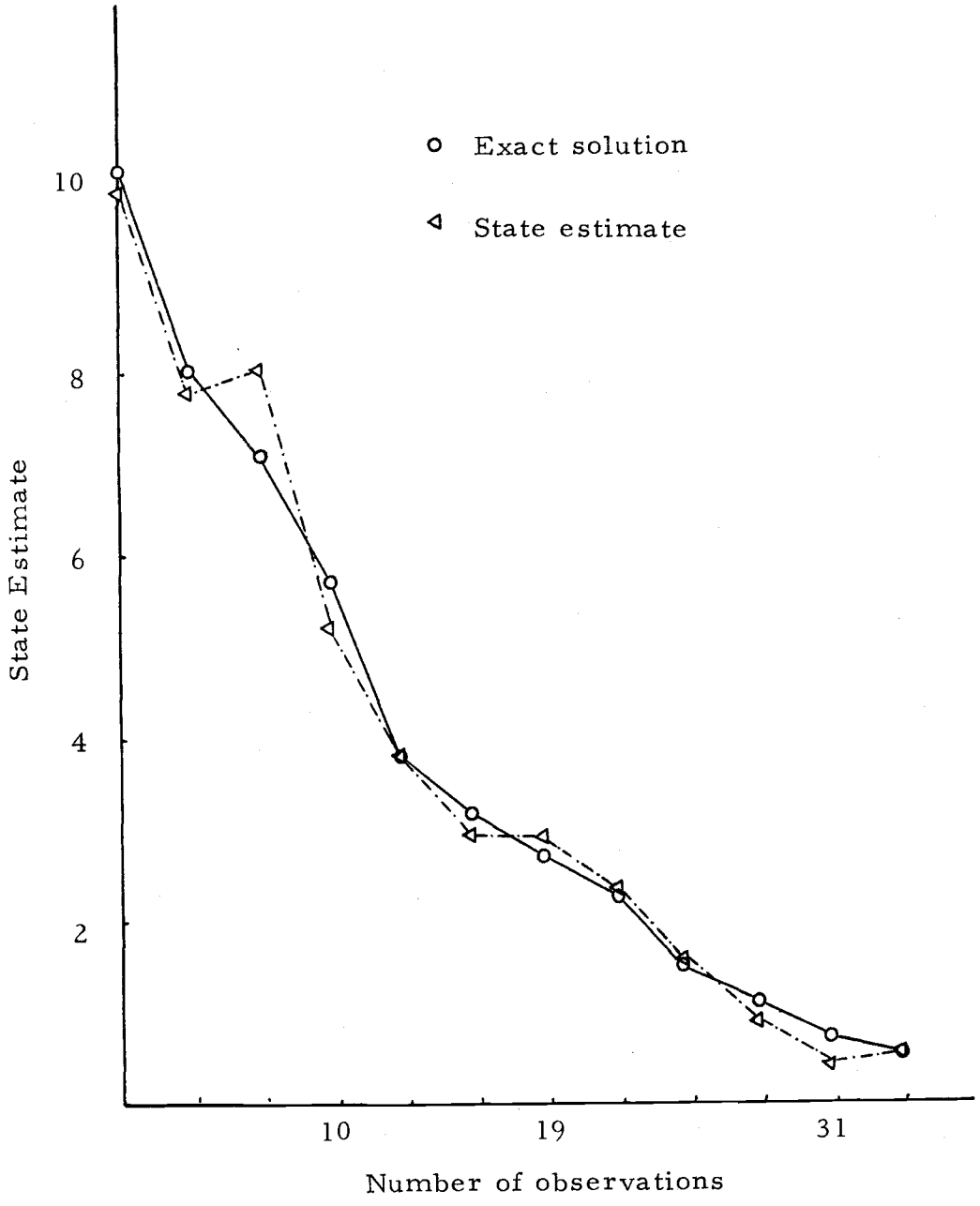


Figure 5.4. State estimate for a scalar homogeneous bilinear system.

with the observation model

$$y(k) = hx(k) + v(k) \ . \tag{5.30}$$

The data was generated for the following parameters

$$a = -3, \quad b = 1 \ , \quad c = 2, \quad e = 0.6$$

$$R = 0.36, \quad Q = 1, \quad DT = 0.01, \quad x(0) = 10,$$

$$u(t) = 3e^{-1.5t} \ .$$

The results of the parameter identification procedure are presented in Table 5.1 and the parameter estimates after 100 observations were used to generate the state estimates of Figure 5.5.

The state estimates are quite satisfactory and the error of around 10% in the parameter estimates can be attributed to the use of a sub-optimal filter as well as a small number of observations. It is, however, important to realize that the effectiveness of any parameter estimation algorithm should be judged on how well the identified model can approximate the states of the original process. As shown in Figure 5.5 the state estimates closely follow the actual values thus making the 10% error in the parameters tolerable.

Example 4:

Consider the two dimensional Abelian bilinear system described by

$$dx(t) = Ax(t)u(t)\,dt + Bx(t)dw(t)\ , \tag{5.31}$$

with the observation model

$$y(k) = hx(k) + v(k)\ , \tag{5.32}$$

where

$$A = \begin{bmatrix} -1. & 6. \\ -0.5 & -5. \end{bmatrix}\ , \qquad B = \begin{bmatrix} 0.8 & -1.2 \\ 0.1 & 1.1 \end{bmatrix}\ ,$$

$$u(t) = 3e^{t},$$

$$h = [0.25 \quad 0.5], \quad Dt = 0.01, \quad R = 0.16, \quad Q = 1.$$

The matricies in (5.31) can be diagonalized using $P^{-1}AP$ and $P^{-1}BP$ where

$$P = \begin{bmatrix} 1. & 2. \\ 0.5 & 3. \end{bmatrix}\ .$$

The diagonalized version is used to generate the data and the identified parameters are provided in Table 5.2. Again, using the parameter estimates after 100 observations, the state estimates of Figure 5.6 for x_1 and Figure 5.7 for x_2 are provided. Most parameters are within a 10% error range which, from the large

number of sample runs made, seems to be the best one could do especially when there are a large number of parameters to be identified. The state estimates are very close to the exact solution which confirms that the identification method used is providing a rather accurate model for the original process.

Table 5.1. Parameter estimates for a scalar inhomogeneous bilinear system.

Observations	$\hat{a}$	$\hat{b}$	$\hat{c}$	$\hat{e}$	$\hat{h}$	$\hat{R}$
10	-6.1	0.1	9.9	0.7	1.0	0.66
20	-3.5	0.4	10.0	1.3	1.0	0.5
30	-5.2	1.1	10.0	0.6	1.0	0.3
40	-4.5	1.0	3.6	0.5	1.0	0.48
50	-4.5	1.0	4.9	1.2	1.0	0.5
60	-4.4	1.1	4.7	0.6	1.0	0.44
70	-3.7	1.0	2.2	0.8	1.0	0.44
80	-3.8	1.1	2.4	1.2	1.0	0.38
90	-3.8	1.1	2.4	0.9	1.0	0.38
100	-3.6	1.0	2.2	0.8	1.0	0.38
Initial Guess	-0.5	2.0	1.4	1.0	0.6	0.56
True Value	-3.0	1.0	2.0	0.6	1.0	0.36

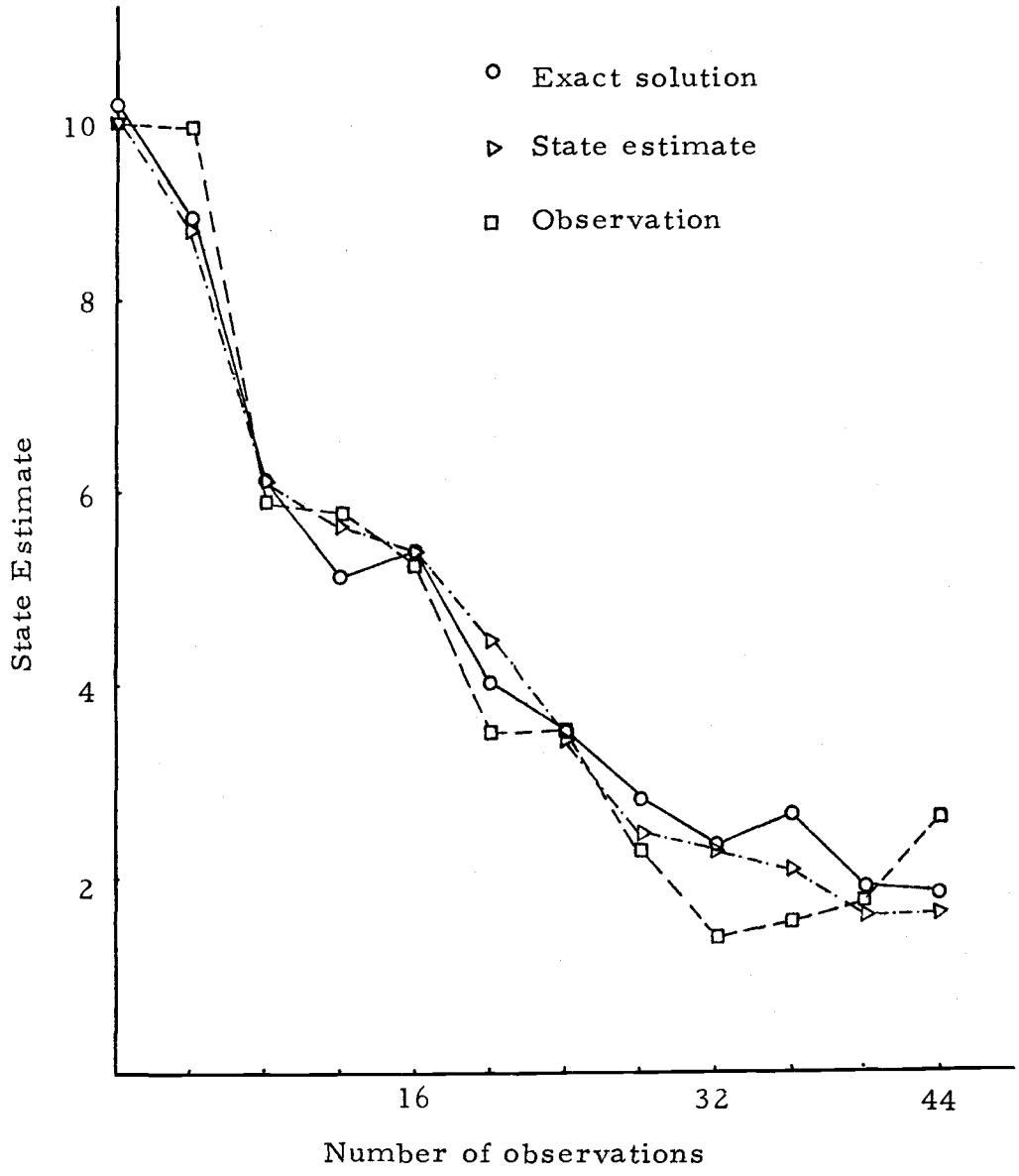


Figure 5.5 State estimates for an inhomogeneous scalar bilinear system.

Table 5.2. Parameter estimates for a second order bilinear system.

Observations	$\hat{a}_{11}$	$\hat{a}_{12}$	$\hat{a}_{21}$	$\hat{a}_{22}$	$\hat{b}_{11}$	$\hat{b}_{12}$	$\hat{b}_{21}$	$\hat{b}_{22}$
40	-2.0	3.0	-0.25	-4.2	0.5	-1.5	0.125	1.65
50	-1.85	3.9	-0.325	-4.45	0.5	-1.8	0.15	1.7
60	-1.125	5.7	-0.475	-5.05	0.5	-1.8	0.15	1.7
70	-0.85	6.9	-0.575	-5.05	0.7	-1.2	0.1	1.5
80	-0.85	6.9	-0.575	-5.04	0.7	-1.2	0.1	1.5
90	-0.85	6.6	-0.575	-5.04	0.7	-1.2	0.1	1.5
100	-1.0	6.6	-0.55	-5.04	0.7	-1.2	0.1	1.5
Initial Guess	-5.0	0.0	0.0	-6.0	3.0	0.0	0.0	3.0
True Value	-1.0	6.0	-0.5	-5.0	0.8	-1.2	0.1	1.1

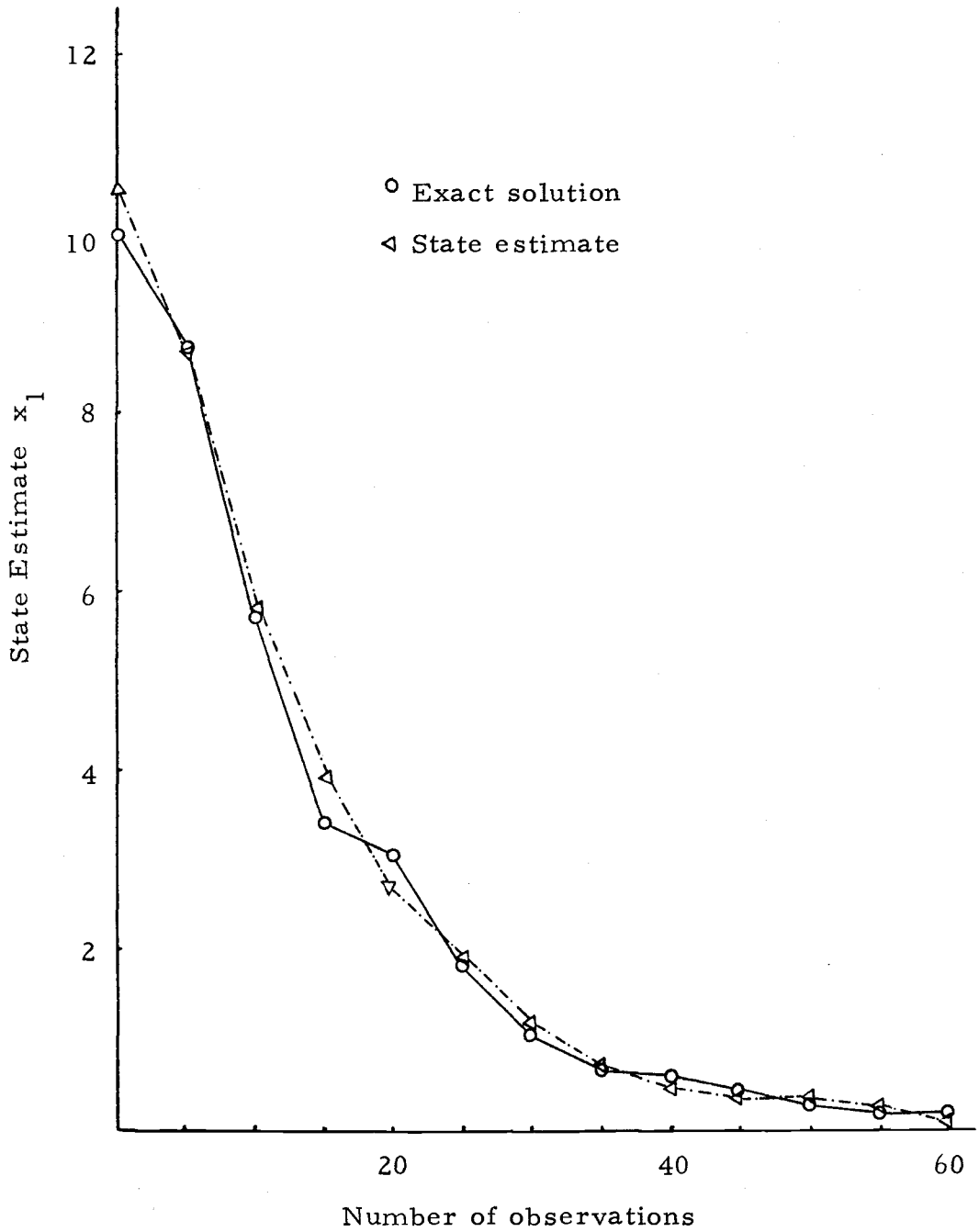


Figure 5.6. State estimates for a second order bilinear system.

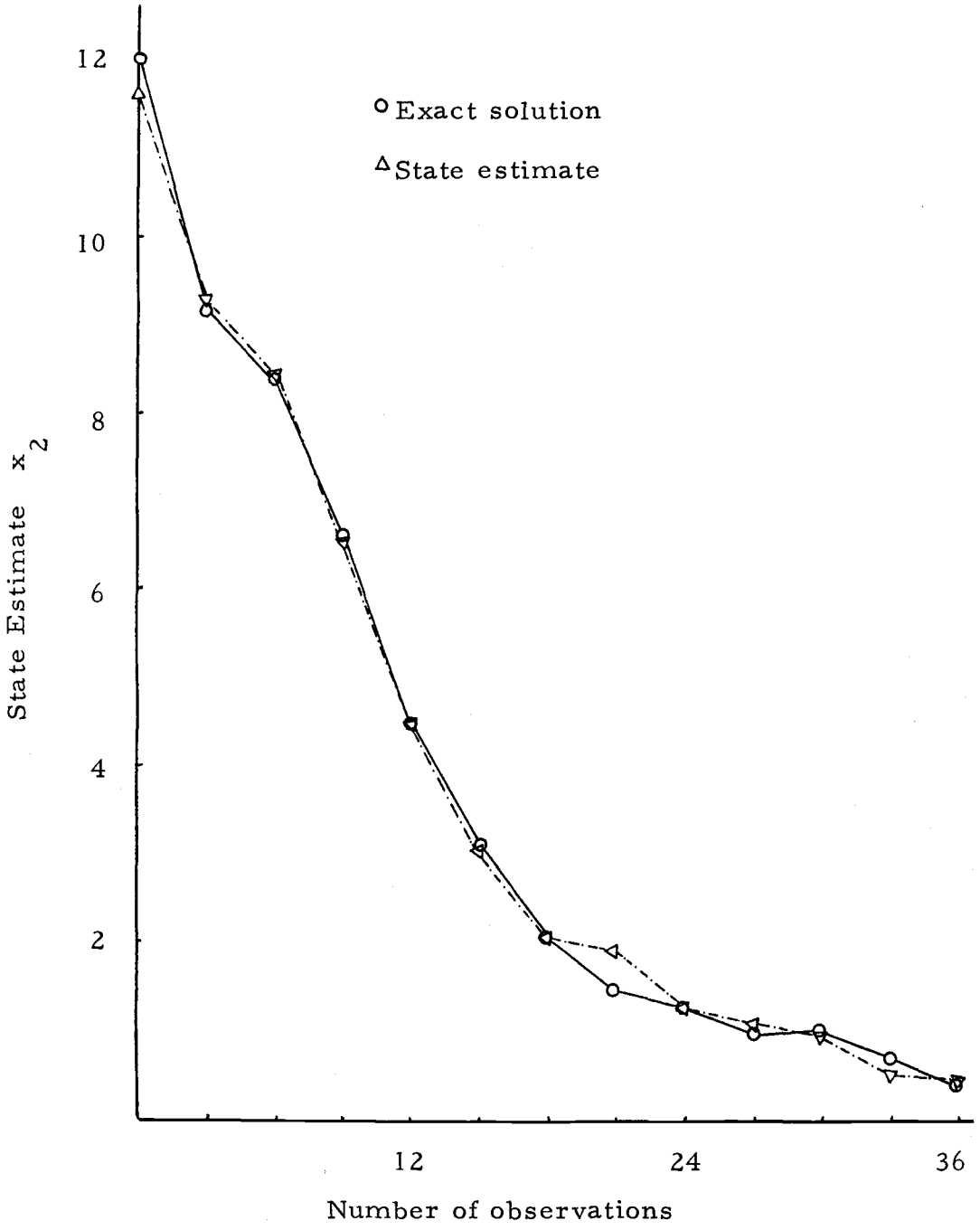


Figure 5.7. State estimates for a second order bilinear system.

VI. CONCLUSIONS AND SUGGESTIONS FOR FUTURE RESEARCH

In this thesis a systematic study of the state and parameter estimation problems for a general class of bilinear systems in the presence of additive as well as multiplicative state noise was presented. A computationally feasible solution to both problems was proposed and tested at the hand of a few examples.

6.1 Summary of Results

The state estimation problem for stochastic bilinear systems was first addressed and an optimal filter for a continuous state and observation model was derived. Even though the filter is infinite dimensional, an implementable filter could be obtained exploiting the appealing structural property that a given moment equation is coupled only to the next higher order moment as well as the lower order moments (for a linear observation model). Since no solution to the optimal filter is available, a sub-optimal procedure was proposed using a moment truncation argument. A second sub-optimal filter based on the use of the conditional best estimate of the state rather than the state itself in the multiplicative noise term was developed for the continuous case. The results were extended to the

continuous-discrete case. The underlying assumptions in using such an approximation were carefully analyzed.

The proposed continuous-discrete filter was employed in the derivation of a maximum likelihood parameter estimation algorithm. A complete set of equations for the evaluation of the likelihood functional as well as its gradient were provided.

A discussion of the importance of closed form solutions in the simulation of the filter and identification algorithms was presented. Closed form solutions for both scalar and certain vector cases were supplied. The importance of ensuring a zero mean in the generation of the noise sequence was emphasized.

Finally, a few examples were provided in which the performances of the proposed algorithms were examined for homogeneous scalar, inhomogeneous scalar, and homogeneous vector cases. The state estimator performed admirably and the parameter estimates were satisfactory.

6.2 Suggestions for Future Research

The following topics present themselves as logical extensions of our research:

1) testing the performance of the continuous sub-optimal filters of Chapter III and including the third order moment.

2) since the density function $p(x, t)$ for the homogeneous case is known to have log normal properties (skewed), it is worthwhile to examine the state estimation using the mode (maximum likelihood estimate) as the best estimate of the state rather than the mean.

3) developing an on-line identification algorithm where the observations do not have to be stored for future iterations.

4) derivation of closed form solutions to a larger class of bilinear stochastic differential equations and testing the performance of the algorithms for such cases.

BIBLIOGRAPHY

A1. M. Aoki. Introduction to Optimization Techniques. Macmillan, 1971.

A2. L. Arnold. Stochastic Differential Equation: Theory and Applications. New York, J. Wiley and Son, 1974.

A3. K. J. Astrom. Introduction to Stochastic Control Theory. New York, Academic Press, 1970.

A4. K. J. Astrom and T. Bohlin. "Numerical Identification of Linear Dynamic Systems from Normal Operating Records," Paper IFAC Symposium, Theory of Self-Adaptive Control Systems, Teddington, England. in Theory of Self-Adaptive Control Systems. Editor P. H. Hammond, Plenum Press, New York, 1966.

A5. K. J. Astrom and P. Eykhoff. "System Identification--A Survey," Automatica, Vol. 7, Pergamon Press, 123-162, 1971.

B1. R. S. Baheti, Parameter and Volterra-Kernel Estimation of Bilinear Systems, Ph.D. Thesis, Dept. of Electrical and Computer Engineering, O.S.U., Corvallis, OR., June 1977.

B2. A. V. Balakrishnan. "Modelling and Identification Theory: A Flight Control Application," in Theory and Applications of Variable Structure Systems, Edited by R. R. Mohler, A. Ruberti. Academic Press, 1972.

B3. A. V. Balakrishnan, "On the Application of Ito Integrals Using Band-Limited Processes," SIAM J. Control, Vol. 12., No. 2, May 1974.

B4. R. W. Bass, V. D. Norum, and L. S. Schwartz. "Optimal-Multichannel Nonlinear Filtering," J. Math. Anal. Appl. 16, 152-164, 1966.

B5. R. E. Bellman, H. Kagiwada, and R. E. Kalaba, "Quasi-linearization, System Identification, and Prediction," Report Rand Corp., RM-3812-PR., 1963.

B6. R. E. Bellman and R. E. Kalaba. "Quasilinearization and Nonlinear Boundary Value Problems," American Elsevier Pub., New York, 1965.

B7. R. W. Brockett. "System Theory on Group Manifolds and Coset Spaces," SIAM J. Control, Vol. 10, No. 2, May 1972, pp. 265-284.

B8. R. W. Brockett. "Lie Theory and Control Systems on Spheres," SIAM J. Appl. Math., Vol. 25, No. 2, Sept. 1973, pp. 213-25.

B9. R. W. Brockett, "On the Algebraic Structure of Bilinear Systems," in Theory and Application of Variable Structure Systems, R. R. Mohler and A. Ruberti, eds., Academic Press, New York, 1972.

B10. R. W. Brockett. "Lie Algebras and Lie Groups in Control Theory," in Geometric Methods in System Theory, D. Q. Mayne and R. W. Brockett, eds., Reidel Pub. Co., The Netherlands, 1973.

B11. R. S. Bucy, "Nonlinear Filtering Theory," IEEE Trans. Automatic Control, Vol. AC-10, 1965, p. 198.

B12. R. S. Bucy, C. Hecht, and K. D. Senne. "New Methods for Nonlinear Filtering," Revue Francaise d'Automatique, Informatique et de Recherche Operationnelle, Feb. 1973, pp. 3-54.

C1. J. M. C. Clark. The Representation of Nonlinear Stochastic Systems with Applications to Filtering, Ph. D. Thesis, Dept. Electrical Engineering, Imperial College, London, England, 1966.

C2. L. Cooper and D. Steinberg. Introduction to Methods of Optimization, W. B. Saunders, 1970.

D1. P. D'Alessandro. "Bilinearity and Sensitivity in Macroeconomy," in Variable Structure Systems with Applications to Economics and Biology, A. Ruberti and R. R. Mohler, eds., Springer Verlong, New York, 1975.

D2. D. M. Detchmendy and R. Sridhar. "Sequential Estimation of States and Parameters in Noisy Nonlinear Dynamical Systems," J. Basic Engr. (Trans. ASME), Vol. 88D, 1965, pp. 362-368.

D3. J. L. Doob. Stochastic Processes, J. Wiley and Son, 1953.

E1. P. Eykhoff. System Identification Parameter and State Estimation, J. Wiley, 1974.

F1. A. F. Fillipov. "On Certain Questions in the Theory of Optimal Control," SIAM J. Control, Series A., Vol. 1, No. 1, 1962.

F2. R. L. Fox. Optimization Methods for Engineering Design, Addison Wesley, 1971.

F3. P. A. Frick. "A Study of Stochastic Bilinear Systems as They Relate to an Economic Modelling Situation," R-76-21, Universita' di Roma, Oct. 1976.

F4. M. Fujisaki, G. Kallianpur, and H. Kunita. "Stochastic Differential Equations for the Nonlinear Filtering Problem," Osaka J. Math, Vol. 9, 1972, pp. 19-40.

F5. N. Funahashi and K. Nakamura, "Discrete Time System Identification by Correlation Method," IEEE Trans. on Aut. Control, Vol. AC-18, 1973, pp. 551-552.

G1. F. R. Gantmacher. The Theory of Matricies, Vol. I. New York: Chelsea, 1960.

G2. I. V. Gikhman and A. V. Skorokhod. Stochastic Differential Equations, Springer Verlog, 1972.

G3. I. Gustavson. "Comparison of Different Methods for Identification of Linear Models for Industrial Processes," 2nd. IFAC Symp. Identification and Process Parameter Estimation, Prague, Paper 11.4, 1970.

I1. K. Ito. Lectures on Stochastic Processes, Tata Institute of Fundamental Research, Bombay, India, 1961.

I2. K. Ito. "On Stochastic Differential Equations," Mem. Amer. Math. Soc., Vol. 4, 1951.

J1. A. H. Jazwinski. Stochastic Processes and Filtering Theory, Academic Press, 1970.

J2. A. H. Jazwinski. "Nonlinear Filtering with Discrete Observation," Paper No. 66-38, AIAA 3rd. Aerospace Sciences Meeting, New York, 1966.

J3. A. H. Jazwinski. "Filtering for Nonlinear Dynamical Systems," IEEE Trans. Aut. Control, Vol. AC-11, 1966, pp. 765-766.

J4. A. H. Jazwinski. "Stochastic Processes with Application to Filtering Theory," Analytical Mechanics Associates, Inc. Seabrook, Maryland, Rept. No. 66-6, 1966.

J5. C. Johnson and E. B. Stear. "Optimal Filtering in the Presence of Multiplicative Noise," Fifth Symp. on Nonlinear Estimation and Its Application, San Diego, Calif., Sept. 23-25, 1974.

K1. T. Kailath. "An Innovations Approach to Least Squares Estimation--Part I," IEEE Trans. on Aut. Control, Vol. AC-13, No. 6, 1968, pp. 646-55.

K2. T. Kailath, "A General Likelihood-Ratio Formula for Random Signals in Gaussian Noise," IEEE Trans. Inf. Theory. Vol. IT 15, May 1969, pp. 350-61.

K3. R. E. Kalman. "A New Approach to Linear Filtering and Prediction Problems," Trans. ASME, J. Basic Engr., Vol. 82D, March 1960, pp. 34-45.

K4. R. E. Kalman and R. S. Bucy. "New Results in Linear Filtering and Prediction Theory," Trans. ASME, J. Basic Engr., Vol. 83D, March 1961, pp. 95-108.

K5. H. J. Kushner. "On the Differential Equations Satisfied by Conditional Probability Densities of Markov Processes," SIAM J. Control, Vol. 2, 1964, pp. 106-119.

K6. H. J. Kushner. "Approximations to Optimal Nonlinear Filters," IEEE Trans. Aut. Control, Vol. AC-12, 1967, pp. 546-56.

K7. H. J. Kushner. "Dynamical Equations for Optimal Nonlinear Filtering," Int. J. Control, Vol. 11, No. 4, 1970, pp. 683-95.

L1. J. T. Lo and A. S. Wilsky. "Estimation for Rotational Processes with One Degree of Freedom I: Introduction and Continuous Time Processes," IEEE Trans. on Aut. Control, Vol. AC-20, No. 1, Feb. 1975, pp. 10-21.

M1. S. I. Marcus. Estimation and Analysis of Nonlinear Stochastic Systems, Ph. D. Thesis, Dept. Electrical Engr., M. I. T., Cambridge, Mass., 1975.

M2. H. P. J. McKean. Stochastic Integrals, New York, Academic Press, 1969.

M3. R. K. Mehra. "Maximum Likelihood Identification of Aircraft Parameters," Joint Aut. Control Conference, 1970, pp. 442-444.

M4. R. K. Mehra and J. S. Tyler. "Case Studies in Aircraft Parameter Identification," Proc. 3rd IFAC Symp-Identification and System Parameter Estimation, North Holland Publ. Comp., Amsterdam, 1973, pp. 117-144.

M5. R. R. Mohler. Bilinear Control Processes, Academic Press, 1973.

M6. R. R. Mohler. "Bilinear Systems; Genesis and Development," 12th. Annual Allerton Conference on Circuits and System Theory, 1974. pp. 814-822.

M7. R. R. Mohler and C. N. Shen. Optimal Control of Nuclear Reactors, New York, Academic, 1973.

N1. T. Nakamizo. "On the State Estimation for Nonlinear Dynamic Systems," Int. J. Control, Vol. 11, No. 4, 1970, pp. 683-695.

R1. C. R. Rao. "Information and Accuracy Attainable in the Estimation of Statistical Parameters," Bull. Calcutta Math. Soc., Vol. 37, 1945, pp. 81.

R2. N. J. Rao, J. D. Borwankar, and D. Ramkrishna. "Numerical Solution of Ito Integral Equations," SIAM J. Control, Vol. 12, No. 1, Feb. 1974, pp. 124-139.

S1. A. P. Sage and J. L. Melsa. Estimation Theory with Application to Communications and Control, McGraw Hill, 1971.

S2. A. A. Sagle and R. E. Walde. Introduction to Lie Groups and Lie Algebra, Academic Press, New York, 1973.

S3. G. N. Saridis and G. Stein. "Stochastic Approximation Algorithm for Linear System Identification," IEEE Trans. on Aut. Control, Vol. AC-13, 1968, pp. 515-523.

S4. L. Schwartz and E. B. Stear. "A Computational Comparison of Several Nonlinear Filters," IEEE Trans. on Aut. Control, Vol. AC-13, 1968, pp. 83-86.

S5. L. Schwartz. "Approximate Continuous Nonlinear Minimal Variance Filtering," Hughes Aircraft Company, Space Systems Division, El Segando, Calif. SSD60472R, Report No. 18, Dec. 1966.

S6. S. Sherman. "A Theorem on Convex Sets with Applications," Ann. Math. Statist. Vol. 26, 1955, pp. 763-767.

S7. S. Sherman. "Non Mean Square Error Criterion." IRE Trans. Inform. Theory, Vol. 4, 1958, pp. 125-126.

S8. H. W. Sorenson and A. R. Stubberud. "Nonlinear Filtering by Approximation of the A Posteriori Density," Int. J. Control, Vol. 8, 1968, pp. 33-51.

S9. R. L. Stratonovich. "A New Form of Representing Stochastic Integrals and Equations," SIAM J. Control, Vol. 4, 1966, pp. 362-371.

S10. R. L. Stratonovich. Topics in the Theory of Random Noise, Gordon and Breach, New York, 1963.

S11. Y. Sunahara. "An Approximation Method of State Estimation for Nonlinear Dynamical Systems," Brown University, Center for Dynamical Systems, Providence, Rhode Island, Tech. Rept. 67-8, Dec. 1967.

S12. P. Swerling. "Note on a New Computational Data Smoothing Procedure Suggested by Minimum Mean Square Estimation," IEEE Trans. Information Theory, Vol. 12, 1966, pp. 9-12.

S13. P. Swerling. "Classes of Signal Processing Suggested by Exact Minimum Mean Square Error Procedure," SIAM J. Applied Math., Vol. 14, 1966, pp. 1199-1224.

U1. H. Unbehauen and B. Goring. "Application of Different Statistical Tests for the Determination of the Most Accurate Order of the Model in Parameter Estimation," Proc. 3rd. IFAC Symp. - Identification and System Parameter Estimation, North Holland Publ. Co., Amsterdam, 1973, pp. 917-928.

W1. N. Wiener. The Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Application, John Wiley and Sons, New York, 1949.

W2. A. S. Wilsky. "Fourier Series and Estimation on the Circle with Applications to Synchronous Communication Parts I, II," IEEE Trans. on Inf. Theory, Vol. IT-20, No. 5, Sept. 1974, pp. 577-590.

W3. R. P. Wishner, J. A. Tabuczynski, and M. A. Athens. "A Comparison of Three Nonlinear Filters," Automatica, Vol. 5, 1968, pp. 487-496.

W4. E. Wong and M. Zakai. "On the Convergence of the Solution of Differential Equations Involving Brownian Motion," Report No. 65-5, AF-AFOSR-139-64, Electronics Research Lab., University of California, Berkley, 1965.

W5. C. M. Woodside. "Estimation of the Order of Linear Systems," 2nd. IFAC Symp.--Identification and Process Parameter Estimation, Prague, Paper 1.7, 1970.