OPTIMAL PARAMETER ADAPTIVE ESTIMATION OF STOCHASTIC PROCESSES

by

Alper K. Caglayan

Dissertation submitted to the Graduate Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Electrical Engineering

APPROVED:

Dr. Hugh F. VanLandingham, Chairman

Dr. Raymond C. Montgomery

Dr. A. Wayne Bennett

Dr. Leonard L. Grigsby

Dr. Frederick H. Lutze, Jr.

October 1974

Blacksburg, Virginia

B

MRJ 1-17-75

ACKNOWLEDGMENTS

I wish to express my sincere gratitude to Professor A. Wayne Bennett and Professor Lawrence Hasdorff for their help, advice, and encouragement while I was in residence at Virginia Polytechnic Institute and State University. I am deeply grateful to my thesis advisor, Dr. Raymond C. Montgomery, for his guidance and direction throughout this research. In particular, his help in the simulation results is appreciated. I also want to thank Professor Hugh F. VanLandingham, chairman of the advisory committee, for his suggestions and comments regarding the preparation of the dissertation.

Thanks go also to Dr. Douglas B. Price and Dr. Nesim Halyo who read the earlier drafts of this dissertation and made many helpful comments and suggestions. I also extend my thanks to Miss Margaret O. Scheier for her patience and understanding in typing the earlier manuscripts.

I am grateful to NASA Langley Research Center for the support of this research through the grant, NGR-47-004-116.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF FIGURES

LIST OF SYMBOLS

$\rightarrow$	logical implication when used between two statements
$\star$	end of a proof
$\equiv$	equality by definition
$\{x:P(x)\}$	the set of all objects with the property P
$\{x\}$	the set containing the element x
ϕ	the empty set
ε	is an element of
$\not\varepsilon$	is not an element of
$\forall$	for all
$\subset$	is a subset of
$\bigcup$	set union
$\bigcap$	set intersection
$I_A(.)$	characteristic function of the set A
$A_1 x A_2$	Cartesian product of the set A_1 and A_2
R	real line
$[a,b]$	$\{x \varepsilon R: a \leq x \leq b\}$
$[a,b)$	$\{x \varepsilon R: a \leq x < b\}$
(a,b)	$\{x \varepsilon R: a < x < b\}$
R^n	n-dimensional Euclidean space
R^{nT}	the set of all R^n-valued functions on T
$[\mu]$	except at a set of μ-measure zero

Chapter I

INTRODUCTION

The optimal parameter adaptive estimation problem may be stated as follows: Assume that we are given a collection (possibly countably infinite) of random processes with known distributions one of which is being observed with additive noise. The *a priori* probability, that a specific random process in this collection will be observed, is specified for each one in the collection. The problem is to find the least squares estimate of the random process that is being observed.

That is, the statistical model is specified up to a set of unknown parameters with known *a priori* probabilities. Knowledge of the parameter vector completely specifies the statistics of the problem. Therefore, the optimal estimator should learn the value of the parameter vector. (This characteristic justifies the name parameter adaptive.) The problem can be visualized as joint detection and estimation [41]. There is a large class of physical problems that can be considered in this context. For instance, joint estimation and system identification [28], joint failure detection and estimation [40], [34], estimation under uncertainty [42], parameter adaptive self-organizing control [43], [44], [45], etc..

Optimal parameter adaptive estimation of sampled Gaussian processes has been investigated by Magill [24]. Magill has shown that the optimal estimator is composed of a set of elemental estimators corresponding to each possible value of the parameter vector and a set of weighting coefficients which is a nonlinear functional of the elemental estimates. In this case, the elemental estimators turn out to be the linear Kalman filters conditioned on each value of the parameter vector. This decomposition of the optimal estimator is known as the "Partition Theorem" in the literature. Optimal parameter adaptive estimation of continuous Gaussian random processes with linear dynamic models has been treated by Lainiotis [27], [28]. Other more specific works can be found in Sims et al. [26] and Park et al. [36].

The previous results on this subject have been limited to parameter adaptive estimation of Gaussian vector random processes with linear dynamic models corrupted by Gaussian noise. In this work, we solve the parameter adaptive estimation problem for a larger class of vector random processes. For instance, the results that we obtain are applicable to vector random processes which are the solutions of nonlinear stochastic differential equations and to second order vector random processes which are not necessarily the solutions of stochastic differential equations. The outline of the thesis is as follows.

Chapter II is a review of stochastic processes and statistical decision theory. The probabilistic framework, in which the problem will be analyzed, is set up in this chapter. We put particular emphasis on the Wiener process, the stochastic integral, linear stochastic differential equations, and statistical decision theory. The only novelty of this chapter is the measure-theoretic approach to the M-ary hypothesis testing problem.

Chapter III is on the joint detection and least squares state estimation problem for linear continuous stochastic dynamic systems. In Section A, the detection problem in linear continuous stochastic dynamic systems will be formulated. Computable expressions for the likelihood ratios and the stochastic differential equations that these likelihood ratios satisfy will be found in the first section. In Section B, joint detection and state estimation problem for linear continuous stochastic dynamic systems is solved using nonlinear filtering theory. It will be shown that the least squares estimate of the state is a weighted average of the hypothesis conditioned estimates weighted by the a posteriori probabilities of the hypotheses conditioned on the observations. The stochastic differential equations that the a posteriori probabilities satisfy will also be found. The stochastic differential equations derived in this chapter have not been treated previously in the literature.

Chapter IV is the main contribution of this dissertation. In this chapter, we will formulate the parameter adaptive estimation problem for a large class of vector random processes in a measure theoretic framework. The scope of the problem considered in this chapter is as follows: We are given a countably infinite collection of vector random processes with known distributions; one of which is being observed with additive white Gaussian noise. The a priori probability, that a specific one in this collection will be observed, is also given. The least squares estimate of the random process that is being observed will be found in terms of the hypothesis conditioned estimates. In this chapter, we will show that the optimal estimate is an infinite linear combination of the hypothesis conditioned estimates with coefficients as the a posteriori probabilities of the hypotheses conditioned on the observations, (Lemma 4.8). We will show that there is only one natural probability measure that one can possibly talk about in this problem;

namely, this probability measure will be the one provided by an extended version of the classical Product Measure Theorem (Lemma 4.1 and Corollary 4.2). A Radon-Nikodym derivative representation will be derived for the a posteriori probabilities (Theorem 4.5). The important part of this Representation Theorem is that this Radon-Nikodym derivative can be written in terms of known Radon-Nikodym derivatives of measures of the stochastic processes in the collection with respect to the Wiener measure (Lemma 4.7). The Radon-Nikodym derivative representation can be considered as generalization of the classical Theorem of Bayes (see the remark following Lemma 4.7). By using recent results of Duncan [17], [18], and Kailath [33], [15], on likelihood functions, we will obtain a general expression for the weighting coefficients of the elemental estimators in terms of the conditioned estimates. We will prove an extended version of "the Partition Theorem" of parameter adaptive estimation (Theorem 4.10); namely, it will be shown the optimal estimator for this parameter adaptive estimation problem consists of two parts: I.) A non-adaptive part in which the hypothesis conditioned estimates are found. II.) An adaptive part in which the a posteriori probability of each hypothesis conditioned on the observations is calculated using the conditioned estimates in Part I. We will then find the stochastic differential equations that the a posteriori probabilities satisfy for the case of finitely many hypotheses (Theorem 4.11). It will also be shown that the a posteriori probability is the unique solution of the stochastic differential equation that it satisfies with the a priori probability as the initial condition. An expression for the conditional error covariance will be found in terms of the hypothesis conditioned error covariances.

In Chapter V, some analogous results on joint detection and estimation will be derived for discrete stochastic dynamic systems. An application of the results to optimal sensor redundancy management in digital slight control systems is in Chapter VI. Real-time hybrid computer simulation results for a self-organizing digital flight control system which is optimally tolerant of sensor failures will be presented in this chapter. Conclusions and recommendations for further study will be presented in Chapter VII.

Appendix I is a review of linear deterministic dynamic systems and the optimal continuous linear regulator. The regulator problem is solved in a Hilbert space setting and an interesting relationship between observability and inner product is proved in this context. Appendix II is a summary of basic definitions of probability theory.

The recent results on likelihood functions of stochastic signals in white noise can not be over emphasized. Prior to Duncan and Kailath's work, methods for computation of likelihood functions were primarily on Gaussian signals in Gaussian noise. The evaluation of these likelihood functions involved solution of integral equations in terms of the covariance of the stochastic process. The work of Grenader [37], Kadota [38], [39], Shepp [32], is along this direction. These integral equations are hard to solve and usually difficult to implement on line. The likelihood functions of stochastic signals in white noise in terms of casual least-squares estimates were first treated by Sosulin and Stratonovich [40] for a restricted class of signals. Major generalizations are due to Duncan and Kailath.

Chapter II

STOCHASTIC PROCESSES AND STATISTICAL DECISION THEORY

In this chapter, we shall set up the probabilistic framework in which we will study the optimal parameter adaptive estimation problem. In Section A, the basic definitions of random processes are introduced. Section B deals with one of the most important building blocks of modern theory of random processes; namely, the Brownian motion process. Elementary properties of the Wiener process, which is used in stochastic integration, is discussed in this section. The stochastic integral for sure functions is defined in Section C. Some characteristic properties of this integral, which are similar to those of the Ito integral [19], are proved in this section. In Section D, the concept of a linear stochastic dynamical system is introduced. Some properties of the solution of the linear vector random integral equations are proved. The classical framework of statistical decision theory is set up and a measure-theroetic solution to the M-ary hypothesis testing problem is given in Section E. The treatments of the subjects in Section A through D are quite standard. The approach to the M-ary hypothesis testing problem is novel.

A. Basic Notions of Probability Theory and Stochastic Processes

A probability space is a triplet (Ω,A,P), where Ω is a set of points w, A is a σ-algebra of subsets of Ω, and P is a measure defined on the measurable space (Ω,A) for which $P(\Omega) = 1$. (For a brief summary of these measure theoretic notions, see Appendix II; for a more detailed analysis see [1], [7].) We can visualize the probability space $(\Omega, \mathcal{A}, P)$ as a random experiment such that the outcomes of the experiment are points w in Ω, the elements of $\mathcal{A}$ are the events determined by the experiment, and P assigns to each event its probability. A random variable x is a real valued measurable function defined on a probability space, i.e., $x: (\Omega, \mathcal{A}) \to (R, \mathcal{B})$ where $\mathcal{B}$ is the Borel sets of the real line which is the σ-algebra generated by the class of all open sets on the real line. The random variable x induces a probability measure P' on its range by:

$$P'(B) = P\{w \in \Omega: x(w) \in B\} \quad \text{for every } B \in \mathcal{B} \tag{2.1}$$

P'(.) is called the distribution of x. If we are interested in only one random variable, it is simpler to work with the induced probability space (R, P'). Every random variable x defines a real valued function $P_x(.)$ on R by

$$P_x(y) = P'[(-\infty, y)] \tag{2.2}$$

$P_x(.)$ is called the distribution function of the random variable x. $P_x(.)$ is a non-decreasing function on R, continuous from the left with $P_x(-\infty) = 0$ and $P_x(+\infty) = 1$. It can be shown if g is any Borel measurable function on R, then

$$E(g(x(w))) \equiv \int_\Omega g(x(w))dP(w) = \int_{-\infty}^{+\infty} g(y)dP_x(y) = \int_R g(y)dP'(y) \tag{2.3}$$

where the second integral should be interpreted as the Lebesgue-Stieltjes integral with respect to $P_x(y)$.

If $x_1, x_2, \ldots, x_n$ are n random variables, then $x = (x_1, x_2, \ldots, x_n)$ is a measurable mapping from (Ω, A) into $(R^n, \mathcal{B}^n)$. The joint distribution of x is defined by

$$P'(B) = P\{w \in \Omega: x(w) \in B\} \quad \text{for every } B \in \mathcal{B}^n$$

Similarly, the joint distribution function of x is defined by

$$P_x(y_1, y_2, \ldots, y_n) = P\{w: x_i(w) < y_i,\ i = 1, \ldots, n\}$$

If $(x_1, x_2, \ldots, x_n)$ are such that $P_x(x_1, \ldots, x_n)$ is absolutely continuous with respect to Lebesgue measure, then from an application of Radon-Nikodym Theorem, there exists a non-negative Borel measurable function $p_x(x_1, x_2, \ldots, x_n)$ such that

$$P_x(B) = \int_B p_x(x_1, \ldots, x_n)\, d(x_1, \ldots, x_n) \quad \text{for every } B \in \mathcal{B}^n$$

$p_x(x_1, \ldots, x_n)$ is called the probability density function for the random variables $x_1, \ldots, x_n$.

The random variable, x, is said to be a Gaussian random variable, if the distribution function of x is given by:

$$P_x(y) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\Pi\ \sigma^2}} \exp[-\frac{1}{2} \frac{(z - \mu)^2}{\sigma^2}]\, dz$$

It then follows that $Ex = \mu$ and $E(x - \mu)^2 = \sigma^2$. The distribution of this Gaussian random variable will be denoted by $N(\mu, \sigma^2)$.

The n random variables $(x_1,x_2,...,x_n)$ is said to have a joint Gaussian distribution, if the joint distribution function of $x = (x_1,x_2,...,x_n)$ is given by:

$$P_x(y) = \int_{(-\infty,y)} \frac{1}{(2\Pi)^{n/2}(\det Q)^{1/2}}$$
$$\exp(-\frac{1}{2}(z-\mu)'\, Q^{-1}(z-\mu))\, d(z_1,z_2,...,z_n)$$

where $y = (y_1,y_2,...,y_n)'$, $z = (z_1,z_2,...,z_n)'$, and Q is a nxn positive definite real matrix.

A collection of random variables is called a stochastic or random process. The stochastic process $\{x(t,w): t \varepsilon T\}$ is measurable if $x(t,w)$ is measurable on the product space $T \times \Omega$. In this thesis T will always be a subset of the real line, R. Unless otherwise stated, T is implicitly assumed to be the interval $[0,\infty)$ in R. $\{x(t): t \varepsilon T\}$ is said to be a Gaussian process if the joint distribution of every finite set of samples $(x(t_1),...,x(t_n))$ of $\{x(t): t \varepsilon T\}$ is Gaussian. All finite joint distributions of a Gaussian process are completely determined by its mean and covariance. $\{x(t)\}$ is a second order process if $E|x(t)|^2 < \infty$ for all t. The set of all second order random variables is a Hilbert space with scalar product

$$\langle x,y\rangle = Exy \tag{2.4a}$$

The norm induced by this scalar product is

$$||x|| = (Ex^2)^{1/2} \tag{2.4b}$$

B. The Wiener Process

One of the most important stochastic processes is the Wiener process (the Brownian motion process). Wiener process is a Gaussian process $\{x(t),\ t \in T\}$ with mean and covariance defined by:

$$E(x(t)) \equiv 0 \quad \text{and} \quad E(x(t)x(s)) \equiv \min(t,s) \tag{2.5}$$

In order to show that min(t,s) defines a covariance we have to show that the matrix $[\min\{t_i,t_j\}]$ for $t_1<t_2<\ldots<t_n$ is positive-semidefinite [10]:

$$[\min\{t_i,t_j\}] = \begin{bmatrix} t_1 & t_1 & . & t_1 \\ t_1 & t_2 & . & t_2 \\ t_1 & t_2 & t_3 & t_3 \\ . & . & . & . \\ . & . & . & . \\ t_1 & t_2 & t_3 & t_n \end{bmatrix}$$

Multiplying the k^{th} row by -1 and adding it to the $k+1^{th}$ row we get the upper triangular matrix

$$[\min\{t_i,t_j\}] \sim \begin{bmatrix} t_1 & t_1 & t_1 & . & . & t_1 \\ 0 & t_2-t_1 & t_2-t_1 & . & . & t_2-t_1 \\ . & 0 & . & . & . & . \\ . & . & . & . & t_{n-1}-t_{n-2} & t_{n-1}-t_{n-2} \\ 0 & 0 & . & . & 0 & t_n-t_{n-1} \end{bmatrix}$$

$[\min\{t_i,t_j\}]$ is positive definite due to Sylvester's criterion.

Theorem 2.1. Let $\{x(t): t \in T\}$ be a Wiener process. Consider the increments of the process $\{x(t) - x(s)\}$ on the interval (s,t). $\{x(t) - x(s)\}$ has a Gaussian distribution $N(0,t-s)$. If (s_1,t_1) and (s_2,t_2) are disjoint, $(x(t_1) - x(s_1))$ and $(x(t_2) - (x(s_2))$ are independent.

Proof: Since $x(t)$ is a Gaussian process, the joint distribution of $\{x(t_1),x(t_2),\ldots,x(t_n), x(s_1),\ldots,x(s_n)\}$ is Gaussian. Since the joint distribution of any finite linear combination Gaussian random variables has a Gaussian joint distribution, $\{x(t_1) - x(s_1),\ldots,x(t_n) - x(s_n)\}$ have a joint Gaussian distribution. $\{x(t) - x(s)\}$ is a Gaussian process.

$$E(x(t) - x(s)) = E(x(t)) - E(x(s)) = 0$$

$$\begin{aligned} E(x(t) - x(s))^2 &= E(x(t))^2 + E(x(s))^2 - 2E(x(t)x(s)) \\ &= \min\{t,t\} + \min\{s,s\} - 2\min\{t,s\} \\ &= t + s - 2s = t - s \end{aligned}$$

Now let $s_1<t_1\leq s_2<t_2$. The covariance $\{x(t_1) - x(s_1), x(t_2) - x(s_2)\}$ is given by

$$\begin{aligned} &\operatorname{cov}\{x(t_1) - x(s_1),x(t_2) - x(s_2)\} \\ &= E(x(t_1) - x(s_1))(x(t_2) - x(s_2)) \\ &= Ex(t_1)x(t_2) + Ex(s_1)x(s_2) - Ex(t_1)x(s_2) - Ex(s_1)x(t_2) \\ &= \min\{t_1,t_2\} + \min\{s_1,s_2\} - \min\{t_1,s_2\} - \min\{s_1,t_2\} \\ &= t_1 + s_1 - t_1 - s_1 = 0 = (Ex(t_1) - x(s_1))(Ex(t_2) - x(s_2)) \end{aligned}$$

So $\{x(t_1) - x(s_1), x(t_2) - x(s_2)\}$ are uncorrelated. Since their joint distribution is Gaussian, they are independent. So $\{x(t) - x(s)\}$ has distribution $N(0,t-s)$.*

The Wiener process is first used as a model for Brownian motion. If $x(t)$ is thought of as the displacement of a particle under Brownian motion, from the above theorem it is clear that the displacement of the particle in successive time intervals are independent random variables and the distribution of the length of the displacement depends only on the length of the time interval. Our interest in Wiener process is in the modeling of noise. The process was first introduced by Brown and was later developed rigorously by Wiener. A widely accepted interpretation of white Gaussian noise is the formal derivative of a Wiener process. (This is just an interpretation and not a definition, since the sample function of a Wiener process is nowhere differentiable with probability 1 [8].)

A stochastic process $\{x(t)\}$ is continuous in the quadratic mean if

$$\lim_{h\to 0} E(x(t+h) - x(t))^2 = 0 \quad \text{for every } t \in T$$

Since, for the Wiener process, $W(t)$, by Theorem 2.1

$$E(W(t+h) - W(t))^2 = t + h - t = h$$

It follows that the Wiener process is continuous in the quadratic mean.

C. Stochastic Integration

In this section, we shall define the integral $\int_a^b f(t)dW(t)$ where $f(t)$ is a real valued function on $[a,b]$ and $W(t)$ is a Wiener process. As we have noted earlier, the sample function of a Wiener process is nowhere differentiable. Therefore, it would be very pleasing if the sample functions of a Wiener process were of bounded variation so that a standard Lebesgue-Stieltjes integral could be defined with respect to a Wiener process. However, the sample functions of a Wiener process are of unbounded variation with probability 1 [8] so that the integral $\int_a^b f(t)dW(t)$ cannot be defined by any of the customary recipes of integration theory. This difficulty can be overcome by defining the stochastic integral as a limit in the quadratic mean.

Let the time interval, T, be $[a,b]$. Consider the Wiener process $\{W(t), t \in [a,b]\}$. Let us define the function $F(t)$ on R by:

$$F(t) = E(W(t) - W(a))^2 = t - a \quad \text{for every } t \text{ in } [a,b]$$

Since $E(W(t) - W(s))^2 = t - s = t - a - (s-a) = F(t) - F(s)$, $F(t)$ is a bounded monotone increasing, continuous function of t. Now if $f(t)$ is any step function on $[a,b]$, then for $a = t_1<t_2<\ldots<t_n = b$

$$f(t) = f(t_k) \qquad t_k \le t < t_{k+1}$$

The integral of $f(t)$ with respect to the Wiener process is defined to be a random variable by:

$$\int_a^b f(t)dW(t) = \sum_{i=1}^{n-1} f(t_i)(W(t_{i+1}) - W(t_i))$$

Now if f and g are two step functions on [a,b] on the same partition

$$E\int_a^b f(t)dW(t)\int_a^b g(t)dW(t)$$

$$= \sum_{i=1}^{n-1}\sum_{j=1}^{n-1} f(t_i)g(t_j)E(W(t_{i+1}) - W(t_i))(W(t_{j+1}) - W(t_j))$$

From Theorem 2.1

$$= \sum_{i=1}^{n-1} f(t_i)g(t_i)E(W(t_{i+1}) - W(t_i))^2$$

By applying Theorem 2.1 once again

$$= \sum_{i=1}^{n-1} f(t_i)g(t_i)(F(t_{i+1}) - F(t_i))$$

So

$$E\int_a^b f(t)dW(t)\int_a^b g(t)dW(t) = \int_a^b f(t)g(t)dF(t) \tag{2.6}$$

where $dF(t)$ is the Lebesque-Stieltjes measure induced by $F(t)$. From Equation 2.6 it follows that

$$E[\int_a^b f(t)dW(t)]^2 = \int_a^b f^2(t)dF(t) \tag{2.6'}$$

So if $f_n(t)$ is a sequence of step function such that

$$\lim_n \int_a^b |f(t) - f_n(t)|^2\, dF(t) = 0$$

from Equation 2.6' it follows that

$$\lim_n E[\int_a^b f(t)dW(t) - \int_a^b f_n(t)dW(t)]^2 = 0 \tag{2.7}$$

Now it would be useful to introduce convergence in the quadratic mean. A sequence of random variables is said to converge to a random variable x if $\lim_n(E(x_n-x)^2)^{1/2} = 0$.

Notation $x = \lim_n x_n$ (q.m.)

So the random variable $\int_a^b f(t)dW(t)$ can be defined by:

$$\int_a^b f(t)dW(t) = \lim_n \int_a^b f_n(t)dW(t) \text{ (q.m.)} \tag{2.8}$$

It is easy to see that the above definition is independent of the particular sequence of step functions $\{f_n(t)\}$ chosen. This integral is defined for all functions which are measurable with respect to Lebesque-Stieltjes measure dF(t) and for which

$$\int_a^b |f(t)|^2 \, dF(t) < \infty$$

Vector random processes are vectors whose elements are stochastic processes. The stochastic integral with sure integrand functions was first defined by Paley and Wiener [19]. Therefore, we shall call the integral defined by Equation 2.8 as the Wiener integral. We summarize the properties of the Wiener integral in the next lemma.

Lemma 2.2. Let $\{W(t): t \varepsilon [a,b]\}$ be a Wiener process. Define $F(t) = E(W(t) - W(a))^2$. Let $g(t)$ be a measurable function with respect to the Lebesque-Stieltjes measure $dF(t)$ and $\int_a^b |g(t)|^2 dF(t) < \infty$. The Wiener integral $I(g)$ of $g(t)$ defined by Equation 2.8 has the following properties:

1. $E\, I(g) = 0$
2. $I(a_1 g_1 + a_2 g_2) = a_1 I(g_1) + a_2 I(g_2)$ where a_1, a_2 are real numbers
3. $E|I(g)|^2 = \int_a^b g^2(t) dF(t)$

Proof: Let g_n be a sequence of step functions such that $\lim_n \int_a^b |g(t) - g_n(t)|^2 dF(t) = 0$. It follows that

$$E\, I(g) = E \lim_n I(g_n) \text{ q.m.}$$

$$= \lim_n E(I(g_n)) \qquad \text{since } E\, I(g_n) = 0 \,\forall n \,\varepsilon\, N$$

$$E\, I(g) = 0$$

The limit and the expectation operator can be interchanged since convergence in the quadratic mean implies convergence in the mean in a probability space, i.e., if $x_n \to x$ q.m., then $x_n \to x$ in the mean. It follows that

$$\lim_n E|x_n - x| = 0$$

Since $|E(x_n - x)| \leq E|x_n - x|$, we have

$$\lim_n |E(x_n - x)| = 0$$

which implies that

$$\lim_n E\, x_n = E \lim_n x_n$$

The second property follows from the property of limits. Third property is true for step function g_n by Equation 2.6', i.e.,

$$E(I(g_n))^2 - \int_a^b g_n^2(t)dF(t) = 0$$

It follows that

$$E(I(g))^2 - \int_a^b g^2(t)dF(t) - (E(I(g_n))^2 - \int_a^b g_n^2(t)dF(t))$$

$$= E(I(g))^2 - E(I(g_n))^2 + \int_a^b g_n^2(t)dF(t) - \int_a^b g^2(t)dF(t)$$

$$= ||I(g)||^2 - ||I(g_n)||^2 + ||g_n||^2 - ||g||^2$$

where the norm for the random variable $I(g)$ is $||I(g)|| = (E(I(g))^2)^{1/2}$ and the norm for $g(t)$ is $||g(t)|| = (\int_a^b g^2(t)dF(t))^{1/2}$. The following inequalities follow from the definition of a norm [4]:

$$||I(g)|| - ||I(g_n)|| \leq ||I(g) - I(g_n)||$$

$$||g_n|| - ||g|| \leq ||g - g_n||$$

Since $I(g_n) \to I(g)$ q.m. and $g_n \to g$ q.m., it follows that [8] $E(I(g_n))^2 \to E(I(g))^2$ and $\int_a^b g_n^2(t)dF(t) \to \int_a^b g^2(t)dF(t)$. Therefore, the property 3 of the Wiener integral follows.*

A vector Wiener process $\{W(t)\}$ is a vector Gaussian process with $E\, W(t) = 0$ and

$$E(W(t)W'(s)) = \int_a^{\min(t,s)} Q(\tau)d\tau \quad \text{for } t,s \text{ in } [a,b]$$

where $Q(t)$ is a positive semidefinite matrix.

<u>Corollary 2.2</u>. If $W(t)$ is an n-vector Wiener process with $E(W(t)W'(s)) = \int_a^{\min(t,s)} Q(\tau)d\tau$, and $B(t)$ in an nxn matrix with integrable elements, then

$$E(\int_a^b B(t)dW(t))(\int_a^b B(t)dW(t))' = \int_a^b B(t)Q(t)B'(t)dt$$

<u>Proof</u>: We first note that if (s_1,t_1) and (s_2,t_2) are disjoint intervals in $[a,b]$ we have

$$E(W(t_1) - W(s_1))(W(t_2) - W(s_2))' = \int_a^{\min(t_1,t_2)} Q(\tau)d\tau$$

$$- \int_a^{\min(t_1,s_2)} Q(\tau)d\tau - \int_a^{\min(s_1,t_2)} Q(\tau)d\tau$$

$$+ \int_a^{\min(s_1,s_2)} Q(\tau)d\tau = 0$$

Assume that the elements of $B(t)$ were step functions such that $B(t) = B_i$ for $t \in [t_i,t_{i+1})$ $i = 1,\ldots,N-1$ with $t_1 = a$ and $t_{N-1} = b$.

$$E(\int_a^b B(t)dW(t))(\int_a^b B(t)dW(t))' = \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} EB_i(W(t_{i+1})$$

$$- W(t_i))W(t_{j+1} - W(t_j))'B_j'$$

In view of the remark we made at the start of the proof, we have

$$= \sum_{i=1}^{N-1} B_i \, E(W(t_{i+1}) - W(t_i))(W(t_{i+1}) - W(t_i))'B_i'$$

$$= \sum_{i=1}^{N-1} B_i \int_{t_1}^{t_{i+1}} Q(\tau)d\tau \; B_i'$$

$$= \int_a^b B \; Q(\tau) \; B'd\tau$$

The general case follows by using a similar argument as in Lemma 2.2.*

D. Linear Stochastic Dynamical Systems

In this section, we will consider linear dynamical systems driven by white noise input. The standard reference for this section is [20]. Consider the vector random integral equation:

$$x(t;w) = x(a;w) + \int_a^t A(s)x(s;w)ds + \int_a^t B(s)dW(s;w) \tag{2.9}$$

where $x(t;w)$ is a n-vector random process, $W(t;w)$ is a m-vector Wiener process with $E\,W(t) = 0$ and $E\,W(t)W'(s) = \int_a^{\min(t,s)} Q(\tau)d\tau$, $A(t)$, $B(t)$ are matrices of appropriate dimensions, elements of $A(t)$ are in $L_2(a,b)$ and elements of $B(t)$ are in $L_2(a,b)$ with respect to Stieltjes measure $dF(t)$ induced by $W(t)$; $x(a;w)$ is a Gaussian random variable independent of the $W(t)$ process. For any sample function the first integral is to be interpreted as a Legesgue integral. The second integral should be interpreted as a Wiener integral defined by Equation 2.8. The unique measurable solution of Equation 2.9 is given by [9]

$$x(t) = \Phi(t,a)x(a) + \int_a^t \Phi(t,s)B(s)dW(s) \tag{2.10}$$

where $\Phi(t,s)$ is the transition matrix of $A(t)$ (see Appendix I for the definition of the transition matrix). Random integral Equation 2.9 can be interpreted as the formal differential equation

$$\dot{x}(t) = A(t)x(t) + B(t)w(t) \tag{2.9'}$$

where $w(t)$ is a white Gaussian process with zero mean and

$$\mathrm{cov}\{w(t),w(s)\} = Q(t)\delta(t-s) \quad \text{where } \delta(t-s)$$

is the Dirac delta function. Using the properties of the Wiener integral, the mean and the covariance of the solution of the random integral equation 2.9 can easily be found.

<u>Theorem 2.3</u>. Consider the solution of the random integral Equation 2.9 given by

$$x(t) = \Phi(t,a)x(a) + \int_a^t \Phi(t,s)B(s)dW(s) \tag{2.10}$$

$x(t)$ is a Gaussian process with its mean and auto-correlation given by

$$E(x(t)) = \Phi(t,a)Ex(a) \tag{2.11}$$

$$Ex(t)x'(s) = \Phi(t,a)\ E(x(a))x'(a))\ \Phi'(s,a) + \int_a^{\min(t,s)} \Phi(t,\tau)B(\tau)Q(\tau)B'(\tau)\tau'(s,\tau)d\tau \tag{2.12}$$

and $E(x(t)x'(t)) = V(t)$ is the unique solution

$$\dot{V}(t) = A(t)V(t) + V(t)A'(t) + B(t)Q(t)B'(t) \tag{2.13}$$

with $V(a) = Ex(a)x'(a)$.

<u>Proof</u>: x(t) is a Gaussian process since it is a linear transformation of a Gaussian process. Equation 2.11 follows from property 1 of the Wiener integral.

$$\begin{aligned} Ex(t)x'(s) &= \Phi(t,a)E(x(a)x'(a))\Phi'(s,a) \\ &+ E\,\Phi(t,a)x(a)\Big(\int_a^s \Phi(s,y)B(y)dW(y)\Big)' \\ &+ E\int_a^t \Phi(t,y)B(y)dW(y)x'(a)\Phi'(s,a) \\ &+ E\int_a^t \Phi(t,y)B(y)dW(y)\Big(\int_a^s \Phi(s,y)B(y)dW(y)\Big)' \end{aligned}$$

The second and third terms are zero because x(a) and W(t) are independent and W(t) has zero mean. Using the fact that W(t) has independent increments and properly 1 of the Wiener integral we get

$$\begin{aligned} Ex(t)x'(s) &= \Phi(t,a)E(x(a)x'(a))\Phi'(s,a) \\ &+ E\int_a^{\min(t,s)} \Phi(t,y)B(y)dW(y)\Big(\int_a^{\min(t,s)} \Phi(s,y)B(y)dW(y)\Big)' \end{aligned}$$

Using Corollary 2.2 we obtain

$$\begin{aligned} Ex(t)x'(s) &= \Phi(t,a)E(x(a)x'(a))\Phi'(s,a) \\ &+ \int_a^{\min(t,s)} \Phi(t,y)B(y)Q(y)B'(y)\Phi'(s,y)dy \end{aligned}$$

It follows then

$$\begin{aligned} V(t) = Ex(t)x'(t) &= \Phi(t,a)E(x(a)x'(a))\Phi'(t,a) \\ &+ \int_a^t \Phi(t,y)B(y)Q(y)B'(y)\Phi'(t,y)dy \end{aligned}$$

Differentiating with respect to t we get Equation 2.13.*

E. Statistical Decision Theory and M-ary Hypothesis Testing

In the sequel we shall use hypothesis testing in the detection of random signals in additive noise. To this end we now develop some aspects of statistical decision theory that will be essential in our development.

The classical model that is used in statistical decision theory consists of three measurable spaces: The parameter space (Ω_1, A_1) consisting of elements θ, the observation space (Ω_2, A_2) consisting of elements y, and the decision space (Ω_3, A_3) consisting of elements a ([11],[12]). For a given element θ in the parameter space, the value that the observation can assume is described probabilistically: For a given $\theta \varepsilon \Omega_1$, a probability measure $P_2(\theta,.)$ will govern the observation when θ is the true value of the parameter. For instance, in the problem we are concerned with, the parameter space will be a countably infinite set of hypotheses each one corresponding to a certain second order random process. The observation space will correspond to measurements under one of these hypotheses. The decision space will consist of decisions that correspond to different hypotheses in the parameter space. The problem of statistical decision theory is to find a decision function, a mapping from the observation space to the decision space, that is best with respect to a cost function. We now give the following definitions to make the above discussions more precise.

Definition 2.4. A decision function is a measurable mapping, d(.), from the observation space, (Ω_2, A_2), into the decision space, (Ω_3, A_3).

That is, the decision function represents the action of deciding an element $a \in \Omega_3$ for a given element $y \in \Omega_2$.

Definition 2.5. The cost function, $C(a,\theta)$, is a nonnegative measurable function defined on $(\Omega_3, A_3) \times (\Omega_1, A_1)$. The numerical value of the cost function at the point (a,θ) represents the loss following a decision "a" when "θ" is the true value of the parameter.

Definition 2.6. Hypothesis testing is a statistical decision problem in which a probability measure, $P_1(.)$, is given on the parameter space, (Ω_1, A_1); and, for every $\theta \in \Omega_1$ a probability measure, $P_2(\theta,.)$ is given on the observation space (Ω_2, A_2) such that, for every fixed $A_2 \in A_2$, $P_2(.,A_2)$ is measurable on (Ω_1, A_1). The Bayes risk is the expected value of the cost function averaged over all $\theta \in \Omega_1$ and $a \in \Omega_3$.

At this stage, we would like to be able to make probability statements about events that are subsets of $\Omega_1 x \Omega_2$. That is, a probability measure, P(.), on $(\Omega_1 x \Omega, A_1 x A_2)$ should be constructed. Intuitively, this probability measure, P, should satisfy

$$P(A_1 x A_2) = \int_{A_1} P_2(\theta, A_2) dP_1(\theta) \quad \text{for } A_1 \in A_1 \text{ and } A_2 \in A_2 \qquad (2.14)$$

The fact that such a probability measure P on $(\Omega_1 x \Omega_2, A_1 x A_2)$ satisfying the above condition exists and is unique is given by an extended version of the classical product measure theorem (see Theorem

2.62 in [30] and Lemma 4.1 and Corollary 4.2). So the Bayes risk becomes

$$B = \int_{\Omega_1 \times \Omega_2} C(d(y),\theta)dP \qquad (2.15)$$

where P is given by 2.14.

Definition 2.7. M-ary hypothesis testing problem is a hypothesis testing problem where the parameter and the decision space contain M elements; i.e., $\Omega_1 = \{\theta_0,\theta_1,\ldots,\theta_{M-1}\}$ and $\Omega_3 = \{a_0,a_1,\ldots,a_{M-1}\}$.

Theorem 2.8. Consider the M-ary hypothesis testing problem described by definitions 2.6 and 2.7. Assume that there exists a σ-finite measure μ on (Ω_2,A_2) such that $P_2(\theta_i,.)$ is absolutely continuous with respect to μ for all i. The decision function, d, which minimizes the Bayes risk is given by

$$d^{-1}(a_j) = \{y \in \Omega_2\colon S_j(y) = \min_{0 \le k \le M-1} S_k(y)\} \quad j = 0,1,\ldots,M-1 \qquad (2.16)$$

where

$$S_k(y) = \sum_{i=0}^{M-1} C(a_k,\theta_i)P_1(\theta_i) \frac{dP_i}{d\mu}(y) \qquad (2.17)$$

where $P_i(.) = P_2(\theta_i,.)$ and $\frac{dP_i}{d\mu}$ is the Radon-Nikodym derivative of P_i with respect to μ. Furthermore, the minimum Bayes risk that can be achieved is given by

$$B_{min} = \int_{\Omega_2} \min_k\{S_k(y)\}\, d\mu \qquad (2.18)$$

Proof: From Equation 2.15, the Bayes risk is

$$B = \int_{\Omega_1 \times \Omega_2} C(d(y),\theta)\, dP$$

$$= \int_{\bigcup_{i=0}^{M-1} \{\theta_i\} \times \Omega_2} C(d(y),\theta)\, dP$$

$$= \sum_{i=0}^{M-1} \int_{\{\theta_i\} \times \Omega_2} C(d(y),\theta)\, dP$$

Since $\Omega_2 = \bigcup_{j=0}^{M-1} d^{-1}(a_j)$, we have

$$B = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} \int_{\{\theta_i\} \times d^{-1}(a_j)} C(d(y),\theta)\, dP$$

$$= \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} C(a_j,\theta_i) \int_{\{\theta_i\} \times d^{-1}(a_j)} dP$$

Using 2.14, we get

$$B = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} C(a_j,\theta_i)\, P_1(\theta_i)\, P_2(\theta_i, d^{-1}(a_j)) \tag{2.19}$$

Since $P_i = P_2(\theta_i,.)$ is absolutely continuous with respect to μ, $\frac{dP_i}{d\mu}$ exists. Also, the $\frac{dP_i}{d\mu} \geq 0$, and

$$P_2(\theta_i, d^{-1}(a_j)) = \int_{d^{-1}(a_j)} \frac{dP_i}{d\mu}(y)\, d\mu \tag{2.20}$$

Combining (2.19) and (2.20), the Bayes risk becomes

$$B = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} C(a_j,\theta_i)P_1(\theta_i) \int_{d^{-1}(a_j)} \frac{dP_i}{d\mu}(y)\, d\mu$$

$$B = \sum_{j=0}^{M-1} \int_{\Omega_2} I_{d^{-1}(a_j)}(y) \sum_{i=0}^{M-1} C(a_j,\theta_i)P_1(\theta_1) \frac{dP_i}{d\mu}(y)\, d\mu$$

where $I_A(.)$ denotes the characteristic function of the set A.

$$B = \sum_{j=0}^{M-1} \int_{\Omega_2} I_{d^{-1}(a_j)}\, S_j(y)\, d\mu \tag{2.21}$$

Note that $S_j(y)$ is non-negative. It is now clear that (2.21) is minimized if d is chosen such that

$$d^{-1}(a_j) = \{y \in \Omega_2 : S_j(y) = \min_k S_k(y)\} \quad j=0,1,\ldots,M-1$$

Also, the minimum value of B is clearly given by (2.18).*

When the observation space is R^n; that is, when the observation space represents a finite sequence of measurements, the σ-finite measure μ in Theorem 2.8 can be taken as Lebesque measure, so that $\frac{dP_i}{d\mu}$ becomes the probability density of the observations under hypothesis i. When the observation space represents a continuous measurement, μ can be taken as the Wiener measure. In the case where the probability density functions exist, an equivalent optimum decision function can be obtained in terms of the likelihood functions Λ_i where

$$\Lambda_i(y) = \frac{p(y|\theta_i)}{p(y|\theta_o)} \tag{2.22}$$

where $p(y|\theta_i)$ is the probability density of the observations when θ_i is the true value of the parameter. In this case, the decision function, d', which is equivalent to (2.16), is given by

$$d'^{-1}(a_j) = \{y \in \Omega_2 : L_k(y) = \min_k L_k(y)\} \quad j = 0,1,\ldots,M-1 \quad (2.23)$$

where $L_k(y)$ is given by

$$L_k(y) = \sum_{i=0}^{M-1} P_1(\theta_i)(C(a_k,\theta_i) - C(a_i,\theta_i))\ \Lambda_i(y)$$

$$+ P_1(\theta_0)(C(a_k,\theta_0) - C(a_0,\theta_0)) \quad (2.24)$$

Of course, the assumption that $C(a_k,\theta_i) \geq C(a_i,\theta_i)$ has to be made to get those equations which is equivalent to asserting that wrong decisions cost more than the right decisions.

The advantage of working with the likelihood ratios is that equivalent decision regions are easier to characterize in the new M-1 dimensional $(\Lambda_1,\Lambda_2,\ldots,\Lambda_{M-1})$ space.

In the next chapter, we shall use the results of this chapter to find a solution to the joint detection and estimation problem in linear stochastic dynamic systems.

Chapter III

SIMULTANEOUS DETECTION AND LEAST SQUARES ESTIMATION IN LINEAR CONTINUOUS STOCHASTIC DYNAMIC SYSTEMS

In this chapter we formulate the joint detection and optimum state estimation problem for linear continuous stochastic dynamical systems. The formulation incorporates actuator and sensor failure detection problems. The results are also applicable to system identification.

This general area in engineering literature is known as optimal parameter adaptive estimation, which is essentially the problem of estimating a random process described by an initially unknown parameter vector. Knowledge of the parameter vector (knowledge of the active hypothesis) completely specifies the statistics of the process so that the optimal adaptive estimator should learn the value of the parameter vector. Optimal adaptive estimation of scalar sampled Gaussian processes has been investigated by Magill [24]. Magill has shown that the optimal estimator consists of a set of elemental estimators and a corresponding set of weighting coefficients, one pair for each possible value of the parameter vector (or for each active hypothesis, whichever way the problem is visualized). The elemental estimator corresponding to a certain value of the parameter vector turns out to be the linear Kalman filter [25] conditioned on that value of the parameter vector. The evaluation of weighting coefficients involves nonlinear operations on the conditioned estimates. This decomposition of the optimal estimator into these two parts is known as "the partition theorem" in the literature. The extension of this problem to the sampled vector Gaussian processes has been reported by Sims et al. [26]. The optimal parameter adaptive estimation of continuous Gaussian vector random processes with linear dynamic models has been treated by Lainiotis [27] using a formal approach. A more analytical derivation of the same results has also been reported by Lainiotis in [28].

In this chapter, we will solve the optimal parameter adaptive estimation problem in linear stochastic dynamical systems using Kushner's results [29] on nonlinear filtering. Although this derivation provides a short derivation of the needed equations for implementation, it does suffer from loss of insight into the problem. In the next chapter, we will solve a more general problem through a more direct route. In this approach, we will prove a generalized version of "the partition theorem" which is applicable to a wide class of problems including Markov processes with nonlinear dynamics and random processes that do not satisfy stochastic differential equations. The results obtained in the next chapter will include the results of this chapter as special cases.

In Section A, we formulate the detection problem in linear continuous stochastic dynamic systems. Computable expressions for the likelihood ratios and the stochastic differential equations that these likelihood ratios satisfy are found in this section. Considerations for on-line implementation of the detection logic is discussed in Section B. In the next section, joint detection and state estimation problem for linear continuous stochastic dynamic systems is solved by using Kushner's results on nonlinear filtering. The least squares estimate of the state will be shown to be a weighted average of the hypothesis conditioned estimates weighted by the a posteriori probabilities. The stochastic differential equations that a posteriori probabilities satisfy will be found. The stochastic differential equations that the likelihood ratios satisfy in the detection problem of Section A, which are stated in Theorem 3.3, and the stochastic differential equations that the weighting coefficients of the elemental estimators satisfy, which are given by Theorem 3.6 in Section C, have not been treated previously in the literature.

A. Detection in Linear Stochastic Dynamic Systems

In this section, we will formulate the detection problem in linear continuous stochastic dynamic systems. The likelihood ratios for the detection problem and the stochastic differential equations that these likelihood ratios satisfy will be found.

Consider the following hypothesized system structure

$$x(t) = c_k + \int_a^t A_k(s)x(s)ds + \int_a^t B_k(s)dW(s) \quad k = 1,2,\ldots,M \tag{3.1}$$

$$Y(t) = \int_a^t C_k(s)x(s)ds + \int_a^t dV(s), \quad t \in [a,b] \tag{3.2}$$

where x is the n-dimensional state vector, Y is the m-dimensional observation vector, W(t) is a n-vector Wiener process with $EW(t) = 0$ and $E[W(t) - W(a)][W(s) - W(a)]' = \int_a^{\min(t,s)} Q(\tau)d\tau$ for some positive semi-definite matrix Q(t), V(s) is a m-vector Wiener process independent of the process W(t) with $EV(t) = 0$ and

$$E[V(t) - V(a)][V(s) - V(a)]' = \int_a^{\min(t,s)} R(\tau)d\tau$$

for some positive definite matrix $R(t)$. $A_k(t)$, $B_k(t)$ nxn matrices whose elements are continuous functions of time and they are known matrices for each hypothesis. $C_k(t)$ is a mxn matrix whose elements are continuous functions of time and is known for each hypothesis. c_k is a Gaussian random variable independent of the Wiener processes $W(t)$ and $V(t)$. The mean and variance of c_k is assumed to be known for each hypothesis.

The formulation given by (3.1) is broad enough to model sensor and actuator failures in linear dynamic stochastic systems. For instance, in the case of sensor failure modeling, we can set $A_k(t) = A(t)$; $B_k(t) = B(t)$, $k = 1,\ldots,M$, and C_k $(k = 1,\ldots,M-1)$ may represent the failure state corresponding to the failure of k^{th} subgroup of sensors and C_M may represent the normal operation with no sensor failures. That is, C_k $(k = 1,\ldots,M-1)$ is the C_M matrix with rows corresponding to the k^{th} subgroup of sensors deleted. Actuator failures can be modeled by changing the columns in the B_k matrix. We have, without loss of generality, assumed that the Wiener processes $W(t)$ and $V(t)$ have zero mean. In the practical applications, solving the problem as if the mean values were known and using the maximum likelihood estimates of the means yield very good results. We shall say more about this in the last chapter. Also, the inherent assumption that (C_k,A_k) is observable is made. The assumption is the amount of redundancy needed so that the resulting system will be practical. If state variable feedback is used in the control of the system, the hopeful state of affairs should be that the system remains observable with the remaining unfailed sensors.

It is well known [20] that, for each fixed hypothesis, the best estimate of the state in the least-squares sense out of all functions which are measurable with respect to the σ-algebra generated by the observations is given by the conditional expectation which can be physically realized by a Kalman-Bucy filter [13]. We shall denote the least squares estimate conditioned on the k^{th} hypothesis by $\hat{x}_k(t)$, i.e.,

$$\hat{x}_k(t) = E(x(t)|Y_t,h_k) \tag{3.3}$$

where (Y_t,h_k) is the minimal σ-algebra generated by the observations up to the time t under the hypothesis h_k.

We shall now proceed to find the likelihood ratio for the detection problem described by 3.1 and 3.2. The likelihood ratio definition (Equation 2.22) that we introduced in Chapter II poses no mathematical problems as long as one works with detection problems with finite discrete measurements. For instance, suppose that $(y(k), k=1,\ldots,n)$ is a m-vector random process under the two hypotheses h_0 and h_1. Then, from Chapter II, we know that the probability density functions $p_y(y_1,\ldots,y_n|h_0)$ and $p_y(y_1,\ldots,y_n|h_1)$ exist under the suitable absolute continuity assumptions. In this case using Equation 2.22 we can define the likelihood ratio

$$\Lambda_1(y_1,\ldots,y_n) = \frac{p_y(y_1,\ldots,y_n|h_1)}{p_y(y_1,\ldots,y_n|h_0)} \tag{3.4}$$

When dealing with continuous measurements complications arise in defining likelihood ratios in terms of the probability density functions. Therefore, it is desirable to define the likelihood ratio given by 3.4

in a more general way. To simplify the notation, let Y_n represent $(y_1,\ldots,y_n)$. In Chapter II, the probability densities were defined by

$$P_{Y_n}(B|h_i) = \int_B p_{y_n}(y_1,\ldots,y_n|h_i)\, d\mu(y_1,\ldots,y_n) \quad i = 0,1 \tag{3.5}$$

for every Borel set of R^{mn} where μ is the Lebesque measure on R^{mn}. That is, the probability densities are given by the Radon-Nikodym derivatives

$$p_{Y_n}(y_1,\ldots,y_n\ |h_i) = dP_{Y_n}(.|h_i)/d\mu \tag{3.6}$$

It follows that the likelihood ratio is given by the Radon-Nikodym derivative that

$$\Lambda_1(y_1,\ldots,y_n) = dP_{Y_n}(.|h_1)/dP_{Y_n}(.|h_0) \tag{3.7}$$

so that the likelihood-ratio can be defined independent of the probability density functions. This approach is more feasible in defining ratios in the case of continuous parameter vector random processes. Every stochastic process $(y(t,w),\ t \in T)$ with prescribed finite dimensional joint distributions induces a measure on the real valued functions in the interval T in the following way [8]. Now $y(t,w)$: $(\Omega \times T) \to R^m$ where (Ω,A,P) is the probability space on which the stochastic process is defined. Define a new mapping $Y: \Omega \to R^{mT}$ by $Y(w) = y(.,w) \in R^{mT}$, where R^{mT} is all R^m valued functions on T. Then $(R^{mT},\mathcal{B}^{mT},PY^{-1})$ is a new probability space where $\mathcal{B}^{mT}$ is the class of Borel sets of R^{mT} and $PY^{-1}(B) = P[Y^{-1}(B)]$ for every $B \in \mathcal{B}^{mT}$. We shall denote PY^{-1} by P_Y. In the simple detection problem we are considering under each hypothesis, h_1, h_0, two probability measures will be induced on R^{mT}; namely, P_{Y_1}, P_{Y_0}. The likelihood ratio is defined as the Radon-Nikodym derivative of P_{Y_1} with respect to P_{Y_0} if it exists. Of course, the likelihood ratio will exist if P_{Y_1} is absolutely continuous with respect to P_{Y_0}. The Radon-Nikodym derivatives of measures induced on the space of continuous functions by the solution of stochastic differential equations have been resolved by Girsanov [16]; a convenient formula is given in Duncan [17], [18]. We summarize these results in the following lemma.

Lemma 3.1. Consider the following stochastic processes in the same probability space.

$$h_1: Y(t,w) = \int_a^t Z(s,w)ds + \int_a^t dV(s,w) \qquad t \in [a,b] = T$$

$$h_0: Y(t,w) = \int_a^t dV(s,w)$$

where $V(t)$ is a m-dimensional Wiener process with $EV(t) = 0$ and $EV(t)V'(s) = \int_a^{\min(t,s)} Q(\tau)d\tau$, $Z(t)$ is a measurable m-dimensional vector random process, which is the solution of a stochastic differential equation, independent of $V(t)$. If $\int_a^b Z'(t,w)Z(t,w)dt$ is finite almost surely, the probability measures induced on the space of R^m valued continuous functions under the two hypotheses, P_{Y_1} and P_{Y_0}, are mutually absolutely continuous and the likelihood ratio, the Radon-Nikodym derivative of P_{Y_1} with respect to P_{Y_0} is given by

$$\Lambda = (dP_{Y_1}/dP_{Y_0}) = \exp\left(\int_a^t \hat{Z}'(s)R^{-1}(s)dY(s) - \frac{1}{2}\int_a^t \hat{Z}'(s)R^{-1}(s)\hat{Z}(s)ds\right) \tag{3.8}$$

where $\hat{Z}(t) = E(Z(t)|Y(s),\ a\leq s\leq t,\ h_1)$.

Proof: See Theorem 3 in [17] and [18].*

Remark: The first integral in 3.8 is an Ito integral. The Ito integral is an extension of the Wiener integral that we have defined in Chapter II to the case where the integrand is random. The definition of Ito integral is the same as in Chapter II and this integral enjoys the same properties as the Wiener integral except that property 3 now becomes [8]:

$$E\left|\int_a^b g(t,w)dW(t)\right|^2 = \int_a^b (E\ g(t,w)|^2)dF(t)$$

We also note that Y(t) is not a Wiener process but since in each hypothesis Y(t) can be written as a stochastic integral it is well defined [19]. The Ito integral is closed under the application of a wide class of smooth functions. This property is known as Ito's Differentiation Lemma [19]. We shall now derive a differential equation that the likelihood ratio of 3.8 satisfies; this equation could be useful in implementation.

Lemma 3.2. The likelihood ratio $\Lambda(t)$ of Equation 3.8 is the solution of the stochastic differential equation

$$d\Lambda(t) = \Lambda(t)\ \hat{Z}'(t)R^{-1}(t)dY(t) \tag{3.9}$$

Proof: The proof is a direct application of the differentiation Lemma on 3.8. Without loss of generality, assume in the proof that R(t) is the identity matrix. Under each hypothesis the likelihood ratio can be written as

$$\Lambda(t) = \exp\left(\int_a^t \hat{Z}'(s)dV(s) + \int_a^t f(s)ds\right) \tag{3.10}$$

where

$$\begin{aligned} f(s) &= -\frac{1}{2}\hat{Z}'(s)\hat{Z}(s) && \text{under } h_0 \\ f(s) &= -\frac{1}{2}\hat{Z}'(s)\hat{Z}(s) + \hat{Z}'(s)\hat{Z}(s) && \text{under } h_1 \end{aligned} \tag{3.11}$$

Now we can write the likelihood ratio Λ in terms of m stochastic integrals:

$$\Lambda = \exp \sum_{i=1}^{m} s_i$$

where $s_i = \int_a^t \hat{Z}_i(s)dV_i(s) + 1/m \int_a^t f(s)ds$. Since the required continuous partials exist, we can apply Ito's differentiation Lemma (see [19]) by noting

$$\frac{\partial\Lambda}{\partial t} = 0, \quad \frac{\partial\Lambda}{\partial s_i} = \frac{\partial^2\Lambda}{\partial s_i \partial s_j} = \Lambda$$

$$d\Lambda = \Lambda \sum_{i=1}^{m} ds_i + \frac{1}{2} \sum_{i=1}^{m}\sum_{j=1}^{m} ds_i ds_j \tag{3.12}$$

where $ds_i = \hat{Z}_i dV_i + f/m\ dt$.

$$d\Lambda = \Lambda \sum_{i=1}^{m} \hat{Z}_i\ dV_i + \quad fdt + \frac{1}{2}\Lambda \sum_{i=1}^{m} \hat{Z}_i^2\ dt \tag{3.13}$$

Reverting back to matrix notation

$$d\Lambda = \Lambda\hat{Z}'dV + \Lambda\ fdt + \frac{1}{2}\Lambda\ \hat{Z}'\hat{Z}dt$$

Under h_0: $d\Lambda = \Lambda\ \hat{Z}'dV = \Lambda\ \hat{Z}'dY$

Under h_1:
$$\begin{aligned} d\Lambda &= \Lambda\ \hat{Z}'dV + \Lambda\ \hat{Z}'\hat{Z}dt \\ &= \Lambda\ \hat{Z}'(Zdt + dV) \end{aligned}$$

$$d\Lambda = \Lambda\ \hat{Z}'dY$$

The proof is complete.*

<u>Theorem 3.3</u>. Consider the detection problem posed by Equation 3.1 and 3.2:

$$x(t) = c_k + \int_a^t A_k(s)x(s)ds + \int_a^t B_k(s)dW(s)$$

$$Y(t) = \int_a^t C_k(s)x(s)ds + \int_a^t dV(s)$$

The likelihood ratios for this detection problem (Λ_k, $k = 1,\ldots,M-1$) where Λ is the Radon-Nikodym derivative of the measure induced by observations under hypothesis h_k with respect to the measure induced under h_M satisfy the following stochastic differential equations:

$$\begin{aligned} d\Lambda_k(t) = \Lambda_k(t)(C_k(t)\hat{x}_k(t) - C_M(t)\hat{x}_M(t))'R^{-1}(t)(dY(t) \\ - C_M(t)\hat{x}_M(t)dt) \qquad k = 1,2,\ldots,M-1 \end{aligned} \tag{3.14}$$

where $\hat{x}_k(t) = E(x(t)|Y(s)\colon a\leq s\leq t,\ h_k)$, $k = 1,2,\ldots,M$. The likelihood ratios are given by

$$\Lambda_k = \frac{\exp(\int_a^t \hat{x}'_k(s)C'_k(s)R^{-1}(s)dY(s) - \frac{1}{2}\int_a^t \hat{x}'_k(s)C'_k(s)R^{-1}(s)C_k(s)\hat{x}_k(s)ds)}{\exp(\int_a^t \hat{x}'_M(s)C'_M(s)R^{-1}(s)dY(s) - \frac{1}{2}\int_a^t \hat{x}'_M(s)C'_M(s)R^{-1}(s)C_M(s)\hat{x}_M(s)ds)}$$

$$k = 1,2,\ldots,M-1 \qquad (3.15)$$

Proof: The likelihood ratios can be obtained by introducing a dummy hypothesis h_o: $Y(t) = \int_a^t dV(t)$. By Lemma 3.1, $P_{Y_k} << P_{Y_o}$ and $P_{Y_o} << P_{Y_M}$. (The symbol, "<<," stands for "absolutely continuous with respect to.) So that by the chain rule of Radon-Nikodym derivatives, we have

$$\frac{dP_{Y_k}}{dP_{Y_M}} = \frac{dP_{Y_k}}{dP_{Y_o}} \cdot \frac{dP_{Y_o}}{dP_{Y_M}} = \frac{dP_{Y_k}}{dP_{Y_o}} \Bigg/ \frac{dP_{Y_M}}{dP_{Y_o}} \qquad [P_{Y_M}]$$

The last equality follows since $P_{Y_M} << P_{Y_o}$ by Lemma 3.1. Equation 3.15 follows from above by using Lemma 3.1. From Lemma 3.1, we see that the likelihood ratios are given by:

$$\xi_k = \frac{\exp(\int_a^t \hat{Z}'_k(s)R^{-1}(s)dY(s) - \frac{1}{2}\int_a^t \hat{Z}'_k(s)R^{-1}(s)\hat{Z}_k(s)ds)}{\exp(\int_a^t \hat{Z}'_M(s)R^{-1}(s)dY(s) - \frac{1}{2}\int_a^t \hat{Z}'_M(s)R^{-1}(s)\hat{Z}_M(s)ds)}$$

$$= \exp(\int_a^t (\hat{Z}_k-\hat{Z}_M)'R^{-1}dY - \frac{1}{2}\int_a^t (\hat{Z}_k-\hat{Z}_M)'R^{-1}(\hat{Z}_k-\hat{Z}_M)ds$$

$$- \int_a^t (\hat{Z}_k-\hat{Z}_M)'R^{-1}\,\hat{Z}_M ds)$$

A similar argument to that used in the proof of the Lemma 3.2 shows that

$$d\xi_k = \xi_k(\hat{Z}_k-\hat{Z}_M)'R^{-1}dY - \xi_k(\hat{Z}_k-\hat{Z}_M)'R^{-1}\,\hat{Z}_M dt$$

Equation 3.14 follows from the preceding relation and the proof is complete.*

B. Implementation of Detection Filters

Of course, the hypothesis conditioned least squares estimates $\hat{x}_k$ in 3.14 and 3.15 are given by the Kalman-Bucy equations [13], [9], [20]:

$$d\hat{x}_k(t) = A_k(t)\hat{x}_k(t)dt + P_k(t)C'_k(t)R^{-1}(t)(dY(t) - C_k(t)\hat{x}_k(t)dt)$$

where $\hat{x}_k(a) = Ec_k$ and $P_k(t)$ is the solution of

$$\dot{P}_k = A_k(t)P_k + P_kA_k'(t) - P_kC_k'(t)R^{-1}(t)C_k(t)P_k + B_k(t)Q(t)B_k'(t) \tag{3.16}$$

with $P_k(a) = Ec_kc_k'$; $k = 1,\ldots,M$. The solution of the detection problem requires M Kalman-Bucy filters that can be realized by an analog computer components in the form

$$\frac{d}{dt}\hat{x}_k(t) = A_k(t)\hat{x}_k(t) + G_k(t)(y(t) - C_k(t)\hat{x}_k(t)) \tag{3.17}$$

where $Y(t) = \int_a^t y(s)ds$ and the filter gain $G_k(t)$ is given by

$$G_k(t) = P_k(t)C_k'(t)R^{-1}(t) \qquad (P_k \text{ is given by 3.16})$$

The reason that the stochastic differential equation can be manipulated to give the ordinary differential equation 3.17 is that the gain of the optimal filter is deterministic [21]. Unfortunately, in the stochastic differential equation we have derived for the likelihood ratios (3.14) the coefficient of dY is random and the formal manipulation would yield erroneous results. Methods to obtain the Ito integral as the limit of ordinary integrals using polygonal approximations [22], [23], can be used to get around this difficulty. So that M-1 likelihood ratios can be computed and based on the results of the last chapter, we can choose the most likely system structure in the Bayesian sense. The evaluation of the likelihood ratios requires M conditioned least mean-square estimates which can be realized by M Kalman-Bucy filters. The detection problem is completely solved for this problem.

C. Joint Detection and State Estimation

The state estimation problem for the system described by 3.1 and 3.2 will be solved in this section. The problem is clearly one of joint detection and estimation. The least squares estimate will be shown to be a weighted average of the hypothesis conditioned estimates weighted by the *a posteriori* probabilities. The stochastic differential equations that the *a posteriori* probabilities satisfy will be found.

We will solve the problem using the results on nonlinear filtering [29]. To this end. we will set up an equivalent nonlinear filtering formulation of the joint detection-estimation problem described by 3.1 and 3.2. One way of doing this is to introduce M new random processes (I_k, $k = 1,\dots,M$) as follows:

Let $k = 1,2,\dots,M$

$I_k(t,w) = 1$ for all $t \in [a,b]$ when hypothesis h_k is true

$I_k(t,w) = 0$ for all $t \in [a,b]$ when hypothesis h_k is not true.

Also, $P(I_k(a,w) = 1) = P_k$, $k = 1,2,\ldots,M$ is specified. (These are the a priori probabilities of the M hypotheses.) Also let

$$\sum_{k=1}^{M} P_k = 1$$

That is, one of the hypotheses is bound to be true. From this assumption it follows that for all t in [a,b] $I_k(t) = 1$ almost surely $\rightarrow$ $I_j(t) = 0$ almost surely for $j \neq k$, $j = 1,\ldots,M$. Now we are in a position to write an equivalent description of the system described by Equation 3.1 and 3.2:

$$x(t) = \sum_{k=1}^{M} I_k(a)c_k + \int_a^t [\sum_{k=1}^{M} I_k(s)A_k(s)]\, x(s)ds$$

$$+ \int_a^b [\sum_{k=1}^{M} I_k(s)B_k(s)]dW(s) \qquad t \in [a,b] \tag{3.18}$$

$$Y(t) = \int_a^t [\sum_{k=1}^{M} I_k(s)C_k(s)]\, x(s)ds + \int_a^t dV(s) \tag{3.19}$$

To simplify notation let us define the vector $\theta(t) = [I_1(t)\ldots I_M(t)]'$. Now by augmenting the state vector in 3.18 and 3.19 by the indicator variables (I_k, $k = 1,\ldots,M$) we get

$$\begin{bmatrix} x(t) \\ \theta(t) \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^{M} I_k(a)c_k \\ \theta(a) \end{bmatrix} + \int_a^t \left[\begin{array}{c|c} \sum_{k=1}^{M} I_k(s)A_k(s) & 0 \\ \hline 0 & 0 \end{array}\right] \begin{bmatrix} x(s) \\ \theta(s) \end{bmatrix} ds$$

$$+ \int_a^t \begin{bmatrix} \sum_{k=1}^{M} I_k(s)B_k(s) \\ 0 \end{bmatrix} dW(s) \qquad t \in [a,b] \tag{3.20}$$

$$Y(t) = \int_a^t \left[\sum_{k=1}^{M} I_k(s) C_k(s) \;\vdots\; 0 \right] \begin{bmatrix} x(s) \\ \theta(s) \end{bmatrix} ds + \int_a^t dV(s) \qquad (3.21)$$

Now the problem is one of non-linear filtering. Least-squares estimate of $x(t)$ is given by [20]

$$\hat{x}(t) = E(x(t)|Y_t) \qquad (3.22)$$

where Y_t is the σ-algebra generated by the observations $(Y(s),\ a\leq s\leq t))$ where $Y(t)$ is as in 3.21. Because of the special structure of the system in 3.20 and 3.21 we can find a representation of $\hat{x}(t)$ in terms of the conditioned estimates.

<u>Lemma 3.4</u>. Consider the system described by 3.20 and 3.21. The conditional expectation of $x(t)$ is given by

$$\hat{x}(t) = \sum_{k=1}^{M} \hat{I}_k(t)\hat{x}_k(t) \qquad (3.23)$$

where $\hat{I}_k(t) = E(I_k(t)|Y_t)$ and $\hat{x}_k(t) = E(x(t)|Y_t,\ I_k = 1)$.

<u>Proof</u>: By the smoothing property of the conditional expectations [7]

$$\begin{aligned} E(x(t)|Y_t) &= E(E(x(t)|Y_t,\ \theta)\Big|Y_t) \\ &= \sum_{k=1}^{M} P(I_k(t) = 1|Y_t)\ E(x(t)|Y_t,\ I_k = 1) \\ &= \sum_{k=1}^{M} P(I_k(t) = 1|Y_t)\ \hat{x}_k(t) \qquad (3.24) \end{aligned}$$

Since $E(I_k(t)|Y_t) = 1 \cdot P(I_k(t) = 1|Y_t) + 0 \cdot P(I_k(t) = 0|Y_t)$

$$= P(I_k(t) = 1|Y_t) \tag{3.25}$$

it follows that $E(x(t)|Y_t) = \sum_{k=1}^{M} \hat{I}_k(t)\hat{x}_k(t)$.*

Note that conditional expectation is a weighted average of the hypothesis conditioned estimates by the a posteriori probabilities. Also, $\hat{x}_k(t)$ is given by the linear Kalman-Bucy filter (3.16 and 3.17). We shall now find a stochastic differential equation that the a posteriori probabilities satisfy using Kushner's results on nonlinear filtering [24]. We state the part of the nonlinear filtering that we will use in the next lemma.

Lemma 3.5. Consider the nonlinear stochastic dynamic system

$$x(t) = c + \int_a^t f(x(s),s)ds + \int_a^t g(x(s),s)dW(s) \tag{3.26}$$

$$Y(t) = \int_a^t h(x(s),s)ds + \int_a^t dV(s) \tag{3.27}$$

where W, V, c are as in 3.1. The vectors $f(x,t)$, $h(x,t)$, $h(x,t)$ and the matrix $g(x,t)$ are continuous in t. The components of $f(x,t)$ and $g(x,t)$ are globally Lipschitz continuous in x. Let $\hat{x}_i(t) = E(x_i(t)|Y_t)$ where $x_i(t)$ is the i^{th} component of x. The conditional mean of x_i satisfies the stochastic differential equation

$$d\hat{x}_i(t) = (dY(t) - E(h(x_t,t)|Y_t)dt)' R^{-1}(t) \tag{3.28}$$

$$(E(x_i(t)h(x_t,t)|Y_t) - \hat{x}_i(t)E(h(x_t,t)|Y_t)) + E(f_i(x_t,t)|Y_t)dt$$

Proof: See [24].*

We shall now derive a stochastic differential equation that the a posteriori probabilities satisfy in 3.24 using Lemma 3.5.

Theorem 3.6. Consider the system described by 3.20 and 3.21. The conditional means of the indicator variables (I_k, k=1,...,M) satisfy the following stochastic differential equations

$$d\hat{I}_k(t) = \hat{I}_k(t)(C_k(t)\hat{x}_k(t) - \sum_{j=1}^{M} \hat{I}_j(t)C_j(t)\hat{x}_j(t))'R^{-1}(t)$$
$$(dY(t) - \sum_{j=1}^{M} \hat{I}_j(t)C_j(t)\hat{x}_j(t)dt) \qquad (3.29)$$

where $\hat{I}_k(t) = E(I_k(t)|Y_t)$ and $\hat{x}_k(t) = E(x(t)|Y_t,\ I_k = 1)$

$$\hat{I}_k(a) = P_k$$

Proof: Applying Lemma 3.4 on 3.20 and 3.21 we get

$$d\hat{I}_k(t) = (E(I_k(t) \sum_{j=1}^{M} I_j(t)C_j(t)x(t)|Y_t) - \hat{I}_k(t)$$
$$E(\sum_{j=1}^{M} I_j(t)C_j(t)x(t)|Y_t))'R^{-1}(t)$$
$$(dY(t) - E(\sum_{j=1}^{M} I_j(t)C_j(t)x(t)|Y_t(dt) \qquad (3.30)$$

Consider the following term in 3.30

$$E(\sum_{j=1}^{M} I_jC_jx|Y_t) = E(E \sum_{j=1}^{M} I_jC_jx|Y_t,\theta)\ Y_t)$$
$$= \sum_{i=1}^{M} P(I_i(t) = 1|Y_t)E(\sum_{j=1}^{M} I_jC_jx|Y_t,I_i = 1)$$

From 3.25 and from the fact that $I_i = 1$ a.s. $\rightarrow I_j = 0$ a.s. for all $j \neq i$ it follows that

$$E\left(\sum_{j=1}^{M} I_j C_j x | Y_t\right) = \sum_{j=1}^{M} \hat{I}_j(t) C_j(t) \hat{x}_j(t) \tag{3.31}$$

Similarly

$$E\left(I_k \sum_{j=1}^{M} I_j C_j x | Y_t\right) = E\left(E\left(I_k \sum_{j=1}^{M} I_j C_j x | Y_t, \theta\right) | Y_t\right)$$

$$= \sum_{\ell=1}^{M} P(I_\ell = 1 | Y_t) E\left(I_k \sum_{j=1}^{M} I_j C_j x | Y_t, I \;\; = 1\right)$$

Since $I_i = 1$ a.s. $\rightarrow I_j = 0$ for all $j \neq i$, it follows that

$$E\left(I_k \sum_{j=1}^{M} I_j C_j x | Y_t\right) = \hat{I}_k(t) C_k(t) \hat{x}_k(t) \tag{3.32}$$

Combining 3.31 and 3.32 with 3.30 we get 3.29 and the proof is complete.*

So the solution of the joint detection and estimation problem for linear continuous stochastic dynamic systems can be realized as follows: Construct M linear Kalman-Bucy filters (3.17) conditioned on each hypothesis. Implement M stochastic differential equations for the posteriori probabilities (3.29) using the outputs of the filters. Form the weighted average of the conditioned estimates (3.23) to obtain the optimum estimate.

In the next chapter we will solve the joint detection and estimation problem for general second order random processes. The results of this chapter will be special cases of the more general results obtained in the next chapter.

Chapter IV

SIMULTANEOUS DETECTION AND LEAST SQUARES ESTIMATION OF RANDOM PROCESSES

This chapter is the main contribution of this dissertation. In this chapter, we shall formulate the parameter adaptive estimation of general second order random processes in a measure-theoretic framework. Previous work has been limited to parameter adaptive estimation of Gaussian vector random processes with linear dynamic models. The results of this chapter are applicable to vector random processes which may be the solutions of nonlinear stochastic differential equations and to vector random processes which are not necessarily the solutions of stochastic differential equations.

The scope of the problem considered in this chapter is as follows: We are given a countably infinite collection of vector random processes with known distributions; one of which is being observed with additive white Gaussian noise. The *a priori* probability, that a specific process in this collection is being observed, is also given. We will find the least-squares estimate of the random process that is being observed in terms of the hypothesis conditioned estimates.

We take a basic measure-theoretic approach in this chapter. We start by stating an extended version of the classical product measure theorem and then show the further characterization in the product probability for the countably infinite hypothesis case. Next, a Radon-Nikodym derivative representation is obtained for the *a posteriori* probabilities of the hypotheses conditioned on the observations. The

Representation Theorem for the _a posteriori_ probability can be considered as an extension of the classical Bayes Theorem. We then prove an extended version of "the partition theorem" of joint detection and estimation. Namely, it is shown that the best estimate in the least squares sense for this parameter adaptive estimation problem is a linear combination of the hypothesis conditioned estimates weighted by the _a posteriori_ probability of each hypothesis; and, the _a posteriori_ probability of each hypothesis is a function of the hypothesis conditioned estimates. So if the hypothesis conditioned estimates can be obtained, by using these estimates, we can get the _a posteriori_ probabilities. Then, we can obtain the least squares estimate by taking the linear combinations of the hypothesis conditioned estimates. That is, the least squares estimator for this problem is partitioned into two parts: I.) A non-adaptive part in which hypothesis conditioned estimates are found. II.) An adaptive part in which _a posteriori_ probability of each hypothesis conditioned on the observations is found by using conditioned estimates of Part I.

We then find the stochastic differential equations that the _a posteriori_ probabilities satisfy for the case when there is a finite number of hypotheses. It is also shown that the _a posteriori_ probability is the unique solution of the stochastic differential equation that it satisfies with the _a priori_ probability as the initial condition. We also derive an expression for conditional error covariance in terms of the conditioned error covariance for each hypothesis. We then specialize the results for linear continuous stochastic dynamic systems.

A. Product Probability

We start with the following version of product measure theorem.

Lemma 4.1: Extended Product Measure Theorem. Let (Ω_1, A_1, P_1) be a probability space and let Ω_2 be a set with a σ- algebra A_2. Assume that for each $w \varepsilon \Omega_1$ we are given a probability measure $P_2(w,.)$ on A_2. If $P_2(w,B)$ is Borel measurable in w for each fixed $B \varepsilon A_2$, then there exists a unique probability measure P on $(\Omega_1 x \Omega_2, A_1 x A_2)$ such that

$$P(A \times B) = \int_A P_2(w,B)dP_1(w) \quad \text{for all } A \varepsilon A_1 \text{ and } B \varepsilon A_2 \qquad (4.1)$$

Namely,

$$P(F) = \int_{\Omega_1} P_2(w,F_w(w'))dP_1(w) \text{ for all } F \varepsilon A_1 \times A_2 \qquad (4.2)$$

where $F_w(w')$ is the Ω_2-section [1] of F; i.e.,

$$F_w(w') = \{w' \varepsilon \Omega_2 : (w,w') \varepsilon F\}$$

Proof: See Theorem 2.62 in Ash [30].*

For instance, in the hypothesis testing problem, Ω_1 in 4.1 will correspond to the set of all hypotheses with an initial a priori probability measure (P_1) on it; Ω_2 will correspond to the observation space and, for each hypothesis, we will be given a probability measure (P_2) on this observation space.

We now show the further characterization in the product probability for the case where Ω_1 is countably infinite in the next corollary.

Corollary 4.2. Let (Ω_1, A_1, P_1), (Ω_2, A_2), and $P_2(w,.)$ be as in Lemma 4.1 except without the assumption that $P_2(w,B)$ be measurable in w for each fixed $b \varepsilon A_2$. If $\Omega_1 = \{w_1, w_2, \ldots, w_n, \ldots\}$ and A_1 is all subsets of Ω_1, then the unique probability measure induced on $A_1 \times A_2$ is given by

$$P(F) = \sum_{i=1}^{\infty} P_1(w_i) P_2(w_i, F_{w_i}(w')) \quad \text{for all } F \varepsilon A_1 \times A_2$$

(The reason that the above assumption about $P_2(w,B)$ can be relaxed in this case is that A_1 is all subsets of Ω_1 so that this assumption is automatically satisfied.)

Proof: Let $F \varepsilon A_1 \times A_2$, then by Lemma 4.1

$$P(F) = \int_{\bigcup_{i=1}^{\infty} \{w_i\}} P_2(w, F_w(w')) dP_1(w)$$

By the countable additivity of indefinite integrals [1]

$$= \sum_{i=1}^{\infty} \int_{\{w_i\}} P_2(w, F_w(w')) dP_1(w)$$

$$= \sum_{i=1}^{\infty} P_1(w_i) P_2(w_i, F_{w_i}(w')).*$$

We shall need the following lemma in the solution of parameter adaptive estimation.

Lemma 4.3. Let Ω be a set and A be a σ-algebra of subsets of Ω. Let $\{P_k\}_{k=1}^{\infty}$ be a sequence of probability measures on A and $\{a_k\}_{k=1}^{\infty}$ be a sequence on non-negative real numbers such that $\sum_{k=1}^{\infty} a_k = 1$. Then the set function $P(.)$ defined on A by

$$P(A) = \sum_{k=1}^{\infty} a_k P_k(A) \qquad \text{for all } A \in A \tag{4.3}$$

is a probability measure on (Ω, A).

Proof: Since $a_k P_k(A) \leq a_k$ for any k and $\sum_{k=1}^{\infty} a_k$ is convergent, P is certainly well-defined, $P(.)$ is non-negative, $P(\emptyset) = 0$, and $P(\Omega) = 1$. To show countable additivity, let $\{A_j\}_{j=1}^{\infty}$ be a sequence of disjoint sets in A. Then

$$\begin{aligned} P\left(\bigcup_{j=1}^{\infty} A_j\right) &= \sum_{k=1}^{\infty} a_k P_k\left(\bigcup_{j=1}^{\infty} A_j\right) \\ &= \sum_{k=1}^{\infty} a_k \sum_{j=1}^{\infty} P_k(A_j) \\ &= \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} a_k P_k(A_j) \end{aligned} \tag{4.4}$$

Now $a_k P_k(A_j) \leq \sum_{j=1}^{\infty} a_k P_k(A_j)$ implies $\sum_{k=1}^{n} a_k P_k(A_j) \leq \sum_{k=1}^{n} \sum_{j=1}^{\infty} a_k P_k(A_j)$. Taking the limits of both sides as $n \to \infty$, we get

$$\sum_{k=1}^{\infty} a_k P_k(A_j) \leq P\left(\bigcup_{j=1}^{\infty} A_j\right)$$

That is, $\sum_{k=1}^{\infty} a_k P_k(A_j)$ is convergent for each $j = 1,2,\ldots.$ Therefore, we can now form the sum

$$\sum_{j=1}^{n} \sum_{k=1}^{\infty} a_k P_k(A_j) = \sum_{k=1}^{\infty} \sum_{j=1}^{n} a_k P_k(A_j) \leq P(\bigcup_{j=1}^{\infty} A_j)$$

which implies that

$$\sum_{j=1}^{\infty} \sum_{k=1}^{\infty} a_k P_k(A_j) \leq P(\bigcup_{j=1}^{\infty} A_j) \tag{4.5}$$

Similarly,

$$\sum_{k=1}^{n} \sum_{j=1}^{\infty} a_k P_k(A_j) = \sum_{j=1}^{\infty} \sum_{k=1}^{n} a_k P_k(A_j) \leq \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} a_k P_k(A_j)$$

So

$$\sum_{k=1}^{\infty} \sum_{j=1}^{\infty} a_k P_k(A_j) \leq \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} a_k P_k(A_j) \tag{4.6}$$

Combining 4.5 and 4.6, it follows then

$$P(\bigcup_{j=1}^{\infty} A_j) = \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} a_k P_k(A_j)$$

$$= \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} a_k P_k(A_j)$$

$$= \sum_{j=1}^{\infty} P(A_j)$$

$P(.)$ is, therefore, a probability measure on (Ω,A).*

B. Representation Theorems for the A Posteriori Probability

Let us now try to cast the parameter adaptive estimation problem into the available measure-theoretic structure. Let Ω_1 be $(w_1, w_2, \ldots, w_n, \ldots)$ where w_i represents the event that i^{th} hypothesis is true, let

A_1 be all subsets of Ω_1. We are also given $P_1(w_i)$ which represents the <u>a priori</u> probability that the i^{th} hypothesis is true. For each $w_i \in \Omega_1$, we have a probability measure on Ω_2 where Ω_2 is the space of all R^m valued continuous functions in the problem described by 3.1. Let A_2 be the Borel sets of Ω_2. Now by Corollary 4.2, we have a unique probability measure P on the events which are in $A_1 \times A_2$. That is, $(\Omega_1 x\Omega_2, A_1 xA_2, P)$ is a probability space. From Lemma 3.4 we see that we need the <u>a posteriori</u> probability that the i^{th} hypothesis is true. Let us now define the transformation Y by Y: $(\Omega_1 x\Omega_2, A_1 xA_2, P) \to (\Omega_2, A_2)$ with

$$Y(w,w') = w' \tag{4.5}$$

In this setting we need to find the conditional probability of the event $w_i \times \Omega_2$ given that $Y = y$; that is, $P(w_i \times \Omega_2 | Y = y)$. Before we go on let us now define conditional probability.

<u>Definition 4.4</u>. Let X be a random variable on (Ω, A, P) and Y be a transformation such that Y: $(\Omega, A) \to (\Omega', A')$. If EX exists, then $E(X|Y = y)$: $(\Omega', A') \to (R, \mathcal{B})$ is defined by

$$\int_{Y^{-1}(A)} XdP = \int_A E(X|Y = y)dP_Y \quad \text{for all } A \in A'$$

If $B \in A$, $P(B|Y = y)$ is defined as $E(I_B|Y = y)$, where I_B is the characteristic function of the set B. If $h(w) = E(X|A_Y)$ and $g(y) = E(X|Y = y)$, then $h(w) = g(Y(w))$ [30]. We are now in the position to state the following important theorem.

Theorem 4.5: A Representation Theorem for the A Posteriori Probability. Let (Ω_1, A_1, P_1) be a probability space with $\Omega_1 = (w_1, w_2, \ldots, w_n, \ldots)$ and let Ω_2 be a set with a σ-algebra A_2. Suppose that we are given a probability measure $P_2(w_i, .)$ on A_2 for each $w_i \in \Omega_1$. Let P be the unique probability measure on $A_1 \times A_2$ given by Corollary 4.2. Let Y be the measurable transformation

$$Y: (\Omega_1 x \Omega_2, A_1 x A_2, P) \to (\Omega_2, A_2) \quad \text{by } Y(w,w') = w' \tag{4.6}$$

Then the conditional probability of the event $w_j \times \Omega_2$ conditioned on $Y = y$ is given by the Radon-Nikodym derivative

$$P(w_j \times \Omega_2 | Y = y) = \frac{d\mu_j}{d\mu}(y)[\mu] \tag{4.7}$$

where $\mu_j(.) = P_1(w_j)P_2(w_j,.)$ and $\mu(.) = \sum_{i=1}^{\infty} P_1(w_i)P_2(w_i,.)$.

Proof: By Lemma 4.3, $\mu(.)$ is a probability measure on A_2. It is clear that $\mu_j(.)$ is a finite measure on A_2. Furthermore, $\mu_j << \mu$ (μ_j is absolutely continuous with respect to μ) since for any $A \in A_2$

$$\sum_{i=1}^{\infty} P_1(w_i)P_2(w_i,A) = 0 \to P_1(w_j)P_2(w_j,A) = 0 \quad \text{for all } j$$

so that the Radon-Nikodym derivative $d\mu_j/d\mu$ exists.

Let us denote the event $w_j \times \Omega_2$ by B_j. From definition 4.4, it follows that $P(B_j|Y = y)$ is defined by the relation: (EI_{B_j} exists and $EI_{B_j} = P(w_j \times \Omega_2) = P_1(w_j)$)

$$\int_{Y^{-1}(A)} I_{B_j} dP = \int_A P(B_j|Y = y)dP_Y \quad \text{for } A \in A_2 \tag{4.8}$$

Remark 1. $P_Y(.) = \mu(.)$. This follows since for any $A \in A_2$

$$P_Y(A) = P(\{(w,w') \in \Omega_1 x \Omega_2 : Y(w,w') \in A\})$$
$$= P(\{(w,w') : w' \in A\})$$
$$= P(\Omega_1 x A)$$
$$= P(\bigcup_{i=1}^{\infty} (w_i \text{ x } A))$$

From now on, we shall write w_i for $\{w_i\}$. Since $\{w_i \text{ x } A\}_{i=1}^{\infty}$ is a disjoint class of sets in A_1 x A_2

$$P_Y(A) = \sum_{i=1}^{\infty} P(w_i \text{ x } A) \qquad (4.9)$$

From Corollary 4.2 it follows that

$$P(w_i \text{ x } A) = \sum_{k=1}^{\infty} P_1(w_k)P_2(w_k,(w_i \text{ x } A)_{w_k})$$

where $(w_i \text{ x } A)_{w_k}$ is the Ω_2-section of the set w_i x A. Since $(w_i \text{ x } A)_{w_k}$ $= \emptyset$ for all $k \neq i$ and $(w_i \text{ x } A)_{w_k} = A$ for $k = i$, it follows that

$$P(w_i \text{ x } A) = P_1(w_i)P_2(w_i,A) \qquad (4.10)$$

Combining 4.9 and 4.10, we get

$$P_Y(A) = \sum_{i=1}^{\infty} P_1(w_i)P_2(w_i,A) = \mu(A)$$

Remark 2. $\int_{Y^{-1}(A)} I_{B_j} dP = P_1(w_j)P_2(w_j,A)$. This follows since

$$\int_{Y^{-1}(A)} I_{B_j} dP = \int_{I_{Y^{-1}(A)\cap B_j}} dP \tag{4.11}$$

$$Y^{-1}(A)\cap B_j = (\bigcup_{i=1}^{\infty} (w_i x A))\cap(w_j x\Omega_2)$$

$$= \bigcup_{i=1}^{\infty} ((w_i x A)\cap(w_j x\Omega_2)$$

$$= \bigcup_{i=1}^{\infty} (w_i\cap w_j)x(A\cap\Omega_2)$$

$$Y^{-1}(A)\cap B_j = w_j \text{ x } A \tag{4.12}$$

Combining 4.11 and 4.12

$$\int_{Y^{-1}(A)} I_{B_j} dP = P(w_j x A) = P_1(w_j)P_2(w_j,A) = \mu_j(A)$$

So 4.8 becomes (using the results in Remarks 1 and 2)

$$\mu_j(A) = \int_A P(w_j x\Omega_2 | Y = y)d\mu \qquad \text{for all } A \in A_2$$

Since $\mu_j << \mu$, the Radon-Nikodym derivative $d\mu_j/d\mu$ exists and

$$\mu_j(A) = \int_A (d\mu_j/d\mu)d\mu \qquad \text{for all } A \in A_2$$

Since $d\mu_j/d\mu$ is unique up to a set of μ-measure zero, it must be that

$$P(w_j x \Omega_2 | Y = y) = d\mu_j / d\mu \; [\mu]$$

The proof is complete.*

The importance of the above theorem is that the Radon-Nikodym derivative in 4.7 can be computed in terms of the conditioned least-squares estimates. The expressions for the Radon-Nikodym derivative of the measures $\{P_2(w_i,.)\}_{i=1}^{\infty}$ with respect to the Wiener measure is fairly well-studied in the literature [16], [17], [18], [31], [32], [15], [33]. With these in mind, we shall derive an expression for equation 4.7 in the case where measures $P_2(w_i,.)$ are mutually absolutely continuous with respect to some other measure μ_o. Before we go on, however, we need the following elementary fact.

Lemma 4.6. Let (Ω, A, μ) be a measure space and let $\{f_n\}_{n=1}^{\infty}$ be a sequence of real-valued integrable functions on Ω such that

$$\sum_{n=1}^{\infty} \int |f_n| d\mu < \infty$$

then the series $\sum_{n=1}^{\infty} f_n(w)$ converges a.e. to an integrable function f and

$$\int fd = \sum_{n=1}^{\infty} \int f_n d\mu$$

Proof: Consider the partial sum $g_n(w) = \sum_{k=1}^{n} \int |f_k(w)|$. From the elementary property of integrals

$$\int g_n d\mu = \sum_{k=1}^{n} \int |f_k| d\mu$$

Since $\{g_n\}_{n=1}^{\infty}$ is an increasing sequence of non-negative real valued measurable functions such that $g_n \to \sum_{k=1}^{\infty} |f_n|$ for every w in Ω, by Lebesgue's Monotone Convergence Theorem [1]:

$$\lim_n \int g_n d\mu = \int \lim_n g_n d\mu$$

which implies

$$\int \sum_{n=1}^{\infty} |f_n| d\mu = \sum_{n=1}^{\infty} \int |f_n| d\mu$$

So $\{w \in \Omega: \sum_{n=1}^{\infty} |f_n(w)| = \infty\}$ must have measure zero. Therefore, $\sum_{n=1}^{\infty} |f_n|$ is convergent a.e. on Ω. Since absolute convergence implies convergence, $\sum_{n=1}^{\infty} f_n$ must be convergent a.e. on Ω. Since $\sum_{k=1}^{n} f_k(w) \leq \sum_{k=1}^{\infty} |f_k(w)|$ and $\sum_{k=1}^{n} f_k \to \sum_{k=1}^{\infty} f_k$ a.e., by Lebesgue's Bounded Convergence Theorem [1], we have

$$\lim_n \int \sum_{k=1}^{n} f_k d\mu = \int \lim_n \sum_{k=1}^{n} f_k d\mu$$

$$\sum_{k=1}^{\infty} \int f_k d\mu = \int \sum_{k=1}^{\infty} f_k d\mu \text{ .*}$$

<u>Lemma 4.7</u>. Let (Ω_1, A_1, P_1), (Ω_2, A_2), $\{P_2(w_i,.)\}_{i=1}^{\infty}$, P, and Y be as in Theorem 4.5. Let μ_o be some σ-finite measure on (Ω_2, A_2) such that

$$P_2(w_i,.) << \mu_o \quad \text{for } i = 1,2,...$$

If the probabilities $\{P_1(w_k)\}$ are all positive, then the conditional probability of the event $w_j \times \Omega_2$ is given by

$$P(w_j \times \Omega_2 | Y = y) = \frac{P_1(w_j)\,\frac{d\overline{\mu}_j}{d\mu_0}}{\sum_{i=1}^{\infty} P_1(w_i)\,\frac{d\overline{\mu}_i}{d\mu_0}} \tag{4.13}$$

where $\overline{\mu}_i(.) = P_2(w_i,.)$.

Proof: From Theorem 4.5 we have

$$P(w_j \times \Omega_2 | Y = y) = d\mu_j/d\mu \; [\mu]$$

where $\mu_j(.) = P_1(w_j)P_2(w_j,.)$ and $\mu(.) = \sum_{i=1}^{\infty} P_1(w_i)P_2(w_i,.)$. By hypothesis, $\overline{\mu}_i << \mu_0$. Since $\mu(.) = \sum_{i=j}^{\infty} P_1(w_i)\overline{\mu}_i(.)$, it follows then $\mu << \mu_0$. Let $A \in A_2$ and $\mu(A) = 0$. This implies

$$\sum_{i=1}^{\infty} \mu_i(A) = 0 \rightarrow \mu_i(A) = 0 \quad \text{for all } i.$$

Therefore, $\mu_i << \mu$. Since $\mu << \mu_0$ and $\mu_i << \mu$, by applying the chain rule for Radon-Nikodym derivatives [1], we get

$$\frac{d\mu}{d\mu_0}\,\frac{d\mu_j}{d\mu} = \frac{d\mu_j}{d\mu_0} \; [\mu_0] \tag{4.14}$$

Since $\{y: \frac{d\mu}{d\mu_0}(y) = 0\}$ has μ-measure zero [1], we can divide both sides of 4.14 by the $\frac{d\mu}{d\mu_0}$ to obtain:

$$P(w_j \times \Omega_2 | Y = y) = \frac{\frac{d\mu_j}{d\mu_0}}{\frac{d\mu}{d\mu_0}} \quad [\mu] \tag{4.15}$$

Now

$$\overline{\mu}_j(A) = \int_A \frac{d\overline{\mu}_j}{d\mu_o} d\mu_o \quad \text{for all } A \in A_2 \tag{4.16}$$

By multiplying both sides of 4.16 by $P_1(w_j)$, we get

$$\mu_j(A) = P_1(w_j)\overline{\mu}_j(A) = \int_A P_1(w_j)(d\overline{\mu}_j/d\mu_o)d\mu_o \quad \text{for all } A \in A_2$$

So

$$d\mu_j/d\mu_o = P_1(w_j)(d\overline{\mu}_j/d\mu_o) \tag{4.17}$$

From 4.17 it follows that

$$\sum_{j=1}^{\infty} P_1(w_j)\overline{\mu}_j(A) = \sum_{j=1}^{\infty} \int_A P_1(w_j)(d\overline{\mu}_j/d\mu_o)d\mu_o \quad \text{for all } A \in A_2 \tag{4.18}$$

Since $d\overline{\mu}_j/d\mu_o$ is a finite valued, non-negative measurable function so is $P_1(w_j)(d\overline{\mu}_j/d\mu_o)$. We, therefore, can apply Lemma 4.6 to get

$$\mu(A) = \sum_{j=1}^{\infty} P_1(w_j)\overline{\mu}_j(A) = \int_A \left(\sum_{j=1}^{\infty} P_1(w_j)(d\overline{\mu}_j/d\mu_o)\right)d\mu_o \quad \text{for all } A \in A_2 \tag{4.19}$$

From the uniqueness of the Radon-Nikodym derivative it follows that

$$d\mu/d\mu_o = \sum_{j=1}^{\infty} P_1(w_j)(d\overline{\mu}_j/d\mu_o) \quad [\mu_o] \tag{4.20}$$

Combining 4.15, 4.17, and 4.20, we get 4.13 and the proof is complete.*

<u>Remark</u>. The Radon-Nikodym derivative representation of the <u>a posteriori</u> probability, Theorem 4.5 and Lemma 4.7, can be considered

as an extension of the classical Theorem of Bayes. This can be seen as follows: Let $\Omega_2 = R^n$ and let μ_0 be the Lebesque measure on R^n. Then we would have

$$\frac{d\bar{\mu}_j}{d\mu_0} = p_Y(y|w_j)$$

where $p_Y(y|w_j)$ is the probability density of Y under the j^{th} hypothesis. The <u>a posteriori</u> probability would then become

$$p(w_j \times \Omega_2|Y = y) = \frac{p_1(w_j)p_Y(y|w_j)}{\sum_{i=1}^{\infty} p_1(w_i)p_Y(y|w_i)}$$

The above equation is one version of the classical Theorem of Bayes.

<u>Remark</u>. In the hypothesis of Lemma 4.7 we have assumed that $P_1(w_i)$ is positive for each i. That is, the <u>a priori</u> probability that i^{th} hypothesis is true should be nonzero. This is not a restriction since if $P_1(w_i)$ were zero then $\mu_j(.) = 0$ and by 4.7 we would have $P(w_j \times \Omega_2|Y = y) = 0$. We shall now find an analogue of Lemma 3.4.

<u>Lemma 4.8</u>. Let (Ω_1,A_1,P_1), (Ω_2,A_2) and $\{P_2(w_i,.)\}_{i=1}^{\infty}$ and Y be as in Theorem 4.5. Let x(t) be a second order vector random process on $(\Omega_1 x\Omega_2, A_1 xA_2, P)$. The conditional expectation of x(t) conditioned on the observation y is given by

$$E(x(t)|Y = y) = \sum_{i=1}^{\infty} P(w_i \times \Omega_2|Y = y)\, E[x(t)|Y = y,\, w = w_i] \tag{4.21}$$

<u>Proof</u>: Let $x_k(t)$ be the k^{th} component of $x(t)$. We will first show that

$$\sum_{i=1}^{\infty} \int_A \left| E(x_k | Y = y,\ w = w_i)(dP_{Y_i}\ \ dP_Y) \right| dP_Y < \infty \quad \text{for any } A \in A_2$$

where $P_{Y_i}(A) = P\{(w,w')\in\Omega_1 x\Omega_2\text{: } Y(w,w') \in A \text{ when } w = w_i\}$

$$= P(w_i \times A)$$

Consider $E(|x_k||Y = y,\ w = w_i)$, which is defined by

$$\int_{Y_i^{-1}(A)} |x_k| dP = \int_A E(|x_k||Y{=}y,\ w{=}w_i) dP_{Y_i} \quad \text{for all } A \in A_2 \tag{4.22}$$

Since $\sum_{i=1}^{\infty} \int_{Y_i^{-1}(A)} |x_k| dP = \sum_{i=1}^{\infty} \int_{w_i x A} |x_k| dP = \int_{\bigcup_{i=1}^{\infty} w_i x A} |x_k| dP$

$$\leq \int_{\Omega_1 x \Omega_2} |x_k| dP \leq \left(\int_{\Omega_1 x \Omega_2} |x_k|^2\ dP \right)^{1/2} < \infty \tag{4.23}$$

and since $P_{Y_i} << P_Y$ it follows from 4.22 and 4.23 that

$$\sum_{i=1}^{\infty} \int_A E(|x_k||Y{=}y,\ w{=}w_i)(dP_{Y_i}/dP_Y) dP_Y < \infty$$

Since in general $|E(x|Y{=}y)| \leq E(|x||Y{=}y)$

$$\sum_{i=1}^{\infty} \int_A |E(x_k|Y{=}y,\ w{=}w_i)(dP_{Y_i}/dP_Y)\ dP_Y < \infty \tag{4.24}$$

Now consider $E(x_k|Y=y,\ w=w_i)$

$$\int_{Y_i^{-1}(A)} x_k dP = \int_A E(x_k|Y=y,\ w=w_i)(dP_{Y_i}/dP_Y)dP_Y \text{ for all } A \in A_2$$

Since $\int_{Y_i^{-1}(A)} x_k dP = \int_{w_i xA} x_k dP$, we have

$$\int_{\bigcup_{i=1}^{\infty} w_i xA} x_k dP = \sum_{i=1}^{\infty} \int_A E(x_k|Y=y,\ w=w_i)(dP_{Y_i}/dP_Y)dP_Y$$

Because of 4.24 we can now apply Lemma 4.6 to get

$$\int_{Y^{-1}(A)} x_k dP = \int_A \sum_{i=1}^{\infty} E(x_k|Y=y,\ w=w_i)(dP_{Y_i}/dP_Y)dP_Y \text{ for all } A \in A_2$$

So we have from definition 4.4,

$$E(x_k|Y=y) = \sum_{i=1}^{\infty} (dP_{Y_i}/dP_Y)E(x_k|Y=y,\ w=w_i) \tag{4.25}$$

Since $P_{Y_i}(A) = P(w_i xA) = P_1(w_i)P_2(w_i,A)$ and $P_Y(A) = P(Y^{-1}(A)) =$

$$P(\bigcup_{i=1}^{\infty} w_i xA) = \sum_{i=1}^{\infty} P(w_i xA) = \sum_{i=1}^{\infty} P_1(w_i)P_2(w_i,A)$$

we can see from Theorem 4.5 that

$$dP_{Y_i}/dP_Y = P(w_i x\Omega_2|Y=y)$$

and the proof is complete.*

We shall also need an extension of Lemma 3.1. Namely, likelihood ratios of general second order vector random processes which are not necessarily the solutions of stochastic differential equations will be given in the next lemma.

Lemma 4.9. Consider the following detection problem:

1. $Y(t) = \int_a^t Z(s)ds + \int_a^t dV(s) \qquad t \in [a,b]$

2. $Y(t) = \int_a^t dV(s)$

where $V(t)$ is an n-vector Wiener process, Z is n-vector random process independent of V such that

$$E\, V(t)V'(s) = \min(t-a, s-a)I$$

$$\int_a^b E(Z'(t)Z(t))dt < \infty$$

Then P_{Y_1}, measure induced on the space of R^m valued continuous functions under hypothesis 1, is absolutely continuous with respect to P_{Y_2}, the measure induced under hypothesis 2. The Radon-Nikodym derivative is

$$dP_{Y_1}/dP_{Y_2} = \exp \int_a^t \hat{Z}_1'(s)dY(s) - \frac{1}{2}\int_a^t \hat{Z}_1'(s)\hat{Z}_1(s)ds \tag{4.26}$$

where $\hat{Z}_1(s) = E\{Z(s)\{Y(u)|\ a \leq u \leq s,\ w=w_1\}$ and Y in 4.26 has P_{Y_2} distribution.

Proof: See Theorem 1 in [18].*

C. Extended Partition Theorem

We are now in a position to state an extended version of "the partition theorem" of joint detection and estimation.

Theorem 4.10: Generalized Partition Theorem. Consider the joint detection and estimation problem

$$Y(t) = \int_a^t Z(s)ds + V(t) \qquad t \in [a,b]$$

where $Z(t) \in \{X_1(t), X_2(t) \ldots, X_n(t), \ldots\}$, $V(t)$ is a zero mean Wiener process with

$$EV(t)V'(s) = \int_a^{\min(t,s)} R(z)dz$$

(R is positive definite) such that $V(a) = 0$, $X_i(t)$ is a measurable random process independent of $V(t)$ with

$$\int_b^a E(X_i'(t)X_i(t))dt < \infty$$

The probability $P_1(w_i)$ of the event w_i which represents $Z(t) = X_i(t)$ is given such that

$$\sum_{i=1}^{\infty} P_1(w_i) = 1.$$

The least squares estimate of $Z(t)$ is then given by

$$\hat{Z}(t) = \sum_{n=1}^{\infty} a_n(t)\hat{Z}_n(t) \qquad (4.27)$$

$$a_n(t) = \frac{P_1(w_n)\exp(\int_a^t \hat{Z}_n'(s)R^{-1}(s)dY(s) - \frac{1}{2}\int_a^t \hat{Z}_n'(s)R^{-1}(s)\hat{Z}_n(s)ds}{\sum_{i=1}^{\infty} P_1(w_i)\exp(\int_a^t \hat{Z}_i'(s)R^{-1}(s)dY(s) - \frac{1}{2}\int_a^t \hat{Z}_i'(s)R^{-1}(s)\hat{Z}_i(s)ds)} \tag{4.28}$$

where $\hat{Z}_i(t) = E(Z(t)|Y_t, w = w_i)$.

Proof: Let us start with a basic probability space (Ω_o, A_o, P_i) (i is a fixed natural number) and the stochastic processes $\{Y_i(t,w_o), X_i(t,w_o), V(t,w_o);\ w_o \in \Omega_o,\ t \in [a,b]\}$ on this probability space such that $\{V(t,w_o)\}$ is a zero mean Wiener process with the prescribed covariance in the hypothesis, $\{X_i(t,w_o)\}$ is a vector random process, satisfying the hypotheses in the statement of the problem. $\{Y_i(t,w_o)\}$ is defined by:

$$P_i\{w_o \in \Omega_o\colon (Y_i(t_1,w_o), Y_i(t_2,w_o) \ldots, Y_i(t_n,w_o)) \in B\}$$
$$= P_i\{w_o \in \Omega_o\colon (\int_a^{t_1} X_i(s,w_o)ds$$
$$+ V(t_1,w_o), \ldots, \int_a^{t_n} X_i(x,w_o)ds + V(t_n,w_o)) \in B\}$$

for all finite collections $\{t_1, t_2, \ldots, t_n\}$ in $[a,b]$ and Borel sets of R^{mn}. Now let $T = [a,b]$. From the discussion in the beginning of this chapter, we know $Y_i(t,w_o)$ induces a measure $P_i Y_i^{-1}$ on the R^m valued real continuous functions on $[a,b]$. Let us denote the set of all continuous real valued functions taking their values in R^m by $\Omega_2 (=R^{mT})$. Let A_2 be the Borel sets of R^{mT}, so that we have now a probability space $(\Omega_2, A_2, P_i Y_i^{-1})$.

Now let $\Omega_1 = \{w_1, w_2, \ldots, w_n, \ldots\}$ and let A_1 be all subsets of Ω_1. Each w_i represents the event that $Z(t) = X_i(t)$ the probability of which is prescribed by $P_1(w_i)$. So we have the probability space (Ω_1, A_1, P_1).

If we now let $P_2(w_i,.) = P_i Y_i^{-1}$, we see that we have cast the problem into the format of extended product measure theorem (Lemma 4.1 and Corollary 4.2). Now by Corollary 4.2 there exists a unique probability measure P on $(\Omega_1 x \Omega_2,\ A_1 x A_2)$ such that

$$P(F) = \sum_{i=1}^{\infty} P_1(w_i) P_2(w_i, F_{w_i}(w')) \quad \text{for all } F \in A_1 x A_2$$

It is this probability space $(\Omega_1 x \Omega_2, A_1 x A_2, P)$ on which the conditional expectation of Z(t) will be found. Let us now see how we can visualize Z(t) as a stochastic process on $(\Omega_1 x \Omega_2, A_1 x A_2, P)$ so that we can apply Lemma 4.8. Now we can write Z(t) as

$$Z(t,w,w') = \sum_{w_i \in \Omega_1} I_{\{w_i\}}(w) X_i(t,w') \quad \text{for each } (w,w') \in \Omega_1 x \Omega_2$$

Note that the above sum is well-defined since, at each point (w,w'), the sum will contain countably many zeroes and a nonzero term. Since $I_{\{w_i\}}(w)$ is A_1-measurable and $X_i(t,w')$ is A_2-measurable for each $t \in T$, $I_{\{w_i\}}(w)X_i(t,w')$ will be $A_1 x A_2$ measurable. Since the limits of measurable functions are measurable, Z(t,w,w') will be $A_1 x A_2$ measurable for each $t \in T$. Therefore, Z(t,w,w') is a vector random process on $(\Omega_1, \Omega_2, A_1 x A_2, P)$. Note that we have reduced a joint estimation and detection problem into a pure estimation problem. So by Lemma 4.8, we have

$$\hat{Z}(t) = E(Z(t)|Y(t)=y) = \sum_{i=1}^{\infty} P(w_i x\Omega_2|Y(t)=y)\ \hat{Z}_i(t)$$

where $y: [a,t] \to R^m$ and $t \in [a,b]$

and $\hat{Z}_i(t) = E(Z(t)|Y(t) = y, w_i)$.

Now we can apply Lemma 4.7 with μ_o as Wiener measure to obtain

$$P(w_i x\Omega_2|Y(t)=y) = \frac{P_1(w_i)\ \dfrac{dP_2(w_i,.)}{d\mu_o}}{\displaystyle\sum_{j=1}^{\infty} P_1(w_j)\ \dfrac{dP_2(w_j,.)}{d\mu_o}} \qquad (4.29)$$

By Lemma 4.9

$$(dP_2/d\mu_o)(w_j,.) = \exp\left(\int_a^t \hat{Z}_j'(s)R^{-1}(s)dY(s) - \frac{1}{2}\int_a^t \hat{Z}_j'(s)R^{-1}(s)\hat{Z}_j(s)ds\right) \qquad (4.30)$$

Substituting 4.30 into 4.29 verifies 4.28 and the proof is complete.*

Theorem 4.10 shows that least squares estimator for this problem is partitioned into two parts: I.) A non-adaptive part in which the least squares estimate of Z(t) conditioned on each hypothesis is found. II.) An adaptive part in which the a posteriori probabilities are found using the conditioned estimates. The second part provides the learning or system identifying nature of the estimator. The structure of the estimator is shown in Figure 4.1. The problem is closely related to detection: $dP_2(w_i,.)/d\mu_o$ in 4.29 is the likelihood function for the detection problem

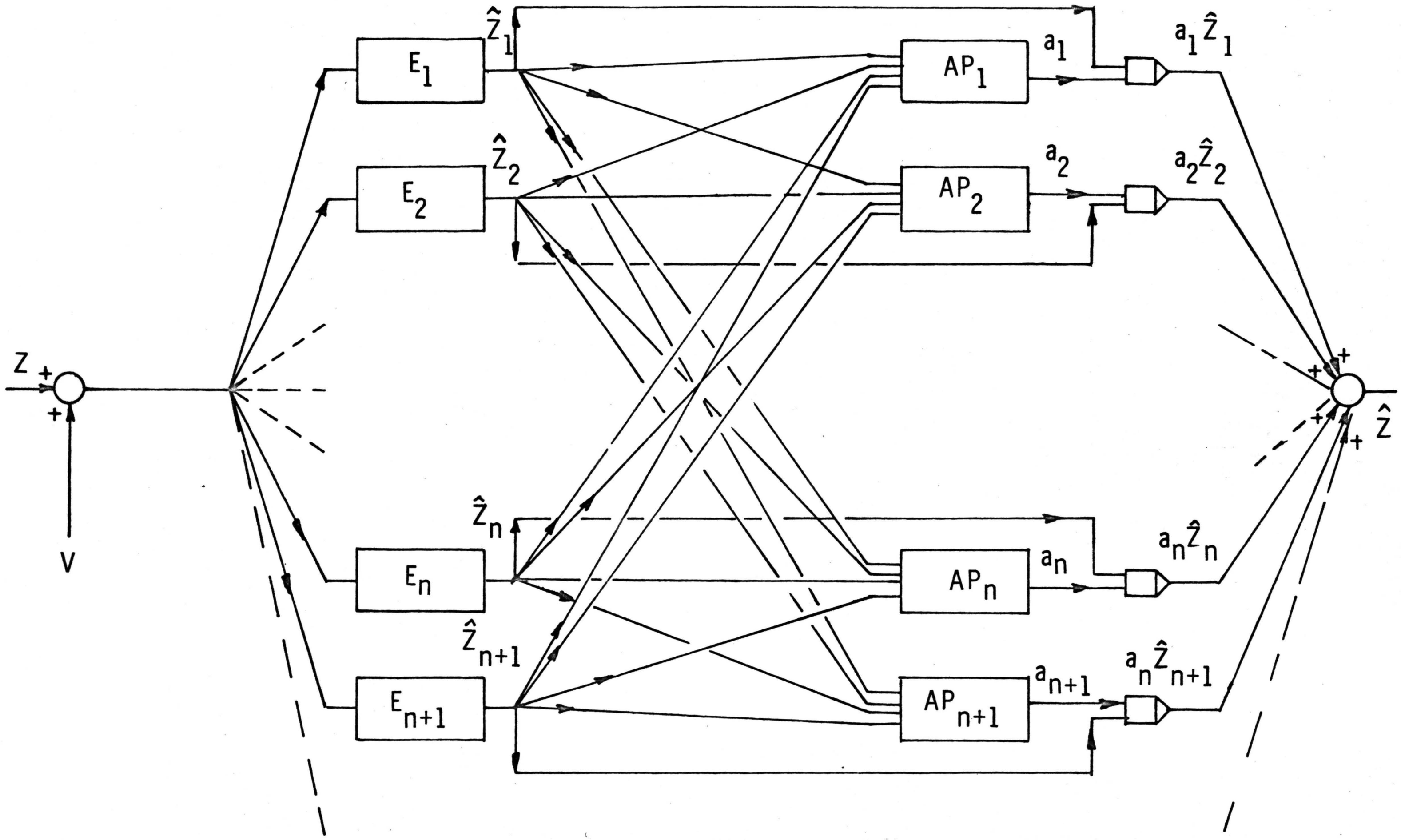

Fig. 4.1 Structure of the Parameter Adaptive Estimator (E_n is the estimator conditioned on the n^{th} hypothesis; AP_n is the a posteriori probability computer for the n^{th} hypothesis).

$$h_i: \quad Y(t) = \int_a^t X_i(s)ds + V(t)$$

$$h_o: \quad Y(t) = V(t)$$

The least squares estimate of Z(t) is the weighted average of the hypothesis conditioned estimates by the a posteriori probabilities.

Remark: It is interesting to note that almost all sample functions of the a posteriori probability $a_n(t)$ will be continuous except at a set of P-measure zero. This follows from 4.28 and the property of the Ito integral which states that the points of continuity of the random process defined by the stochastic integral are the points of continuity of the Wiener process with respect to which the integral is defined [8].

We shall now find the stochastic differential equations that the coefficients $a_n(t)$ satisfy in the case when $\Omega_1 = \{w_1, w_2, \ldots, w_M\}$ is finite. These equations can be used in the implementation of the filter. We will also find an expression for the conditional error covariance in terms of the conditioned error covariance for each hypothesis, that can be used as a measure of performance of the filter. We now give these results in the next theorem.

Theorem 4.11. Consider the joint detection and estimation problem of Theorem 4.10 with $\Omega_1 = \{w_1, w_2, \ldots, w_M\}$, i.e.,

$$Y(t) = \int_a^t Z(s)ds + V(t) \qquad t \in [a,b]$$

where $Z(t) \in \{X_1(t), X_2(t), \ldots, X_M(t)\}$ and the rest of the hypothesis remains the same.

The least squares estimate of Z(t) is given by

$$\hat{Z}(t) = \sum_{n=1}^{M} a_n(t)\hat{Z}_n(t) \tag{4.31}$$

where $\hat{Z}_n(t) = E(z(t)|Y_t, w=w_n)$.

The coefficient a_n in Equation 4.31 is the unique solution of the following stochastic differential equation with $a_n(t) = P_1(w_n)$ at t = a.

$$da_n(t) = a_n(t)(\hat{Z}_n(t) - \sum_{i=1}^{M} a_i(t)\hat{Z}_i(t))'R^{-1}(t)(dY - \sum_{i=1}^{M} a_i(t)\hat{Z}_i(t)dt)$$

$$n = 1,2,\ldots,M \tag{4.32}$$

The conditional error covariance $P(t) = [(Z(t)-\hat{Z}(t))(Z(t)-\hat{Z}(t))'|Y_t]$ is given by

$$P(t) = \sum_{n=1}^{M} a_n(t)(P_n(t) + \hat{Z}_n(t)\hat{Z}_n'(t) - \hat{Z}(t)\hat{Z}'(t)) \tag{4.33}$$

where $P_n(t)$ is the conditional error covariance conditioned on the n^{th} hypothesis, i.e.,

$$P_n(t) = [E(\hat{Z}(t) - \hat{Z}_n(t))(\hat{Z}(t) - \hat{Z}_n(t))'|Y_t, w = w_n]$$

<u>Proof</u>: From Theorem 4.10, with $\Omega_1 = \{w_1,\ldots,w_M\}$, we have

$$\hat{Z}(t) = \sum_{n=1}^{M} a_n(t)Z_n(t) \qquad t \in [a,b] \tag{4.34}$$

$$a_n(t) = \frac{P_1(w_n)\ \exp(\int_a^t \hat{Z}_n'(s)R^{-1}(s)dY(s) - \frac{1}{2}\int_a^t \hat{Z}_n'(s)R^{-1}(s)\hat{Z}_n'(s)ds)}{\sum_{i=1}^{M} P_1(w_i)\exp(\int_a^t \hat{Z}_i'(s)R^{-1}(s)dY(s) - \frac{1}{2}\int_a^t \hat{Z}_i'(s)R^{-1}(s)\hat{Z}_i'(s)ds} \qquad (4.35)$$

We will first prove 4.33:

$$P(t) = E((Z-\hat{Z})(Z-\hat{Z})'|Y_t)$$

$$= E(ZZ'|Y_t) - E(Z\hat{Z}'|Y_t) - E(\hat{Z}Z'|Y_t) + E(\hat{Z}\hat{Z}'|Y_t)$$

Since $\hat{Z}$ is Y_t-measurable

$$= E(ZZ'|Y_t) - \hat{Z}\hat{Z}'$$

From Lemma 4.8

$$= \sum_{n=1}^{M} a_n\ E(ZZ'|Y_t,\ w=w_n) - \hat{Z}\hat{Z}'$$

$$= \sum_{n=1}^{M} a_n\ E[(Z-\hat{Z}_n)(Z-\hat{Z}_n)'|Y_t,\ w=w_n]$$

$$+ \sum_{n=1}^{M} a_n\ E(Z\hat{Z}_n'+\hat{Z}_nZ'-\hat{Z}_n\hat{Z}_n'|Y_t,\ w=w_n) - \hat{Z}\hat{Z}$$

Since $\hat{Z}_n$ is $(Y_t,\ w=w_n)$-measurable

$$= \sum_{n=1}^{M} a_nP_n + \sum_{n=1}^{M} a_n\hat{Z}_n\hat{Z}_n' - \hat{Z}\hat{Z}'$$

Since $\hat{Z}\hat{Z}' = \sum_{n=1}^{M} a_n\hat{Z}\hat{Z}'$ (because $\sum_{n=1}^{M} a_n = 1$)

$$P(t) = \sum_{n=1}^{M} a_n(P_n+\hat{Z}_n\hat{Z}_n'-\hat{Z}\hat{Z}')$$

which verifies 4.33.

To prove Equation 4.32, we could directly apply Ito's differentiation Lemma on Equation 4.35 which contains Mm stochastic integrals. However, we shall not do so since the direct method gets exceedingly complex. Instead, we note that 4.35 is the quotient of two stochastic integrals

$$a_n = s_1/s_2 \quad \text{with } s_2 > 0$$

where $s_1 = P_1(w_n)\Lambda_n$ and

$$s_2 = \sum_{i=1}^{M} P_1(w_i)\Lambda_i$$

with

$$\Lambda_j = \exp\left(\int_a^t \hat{Z}_j'(s)R^{-1}(s)dY(s) - \frac{1}{2}\int_a^t \hat{Z}_j'(s)R^{-1}(s)\hat{Z}_j(s)ds\right)$$

The fact that s_1 and s_2 are stochastic integrals follow from Lemma 3.2. Since the indicated partials exist and are continuous, by Ito's differentiation Lemma we have

$$da_n = \sum_{i=1}^{2} \frac{\partial a_n}{\partial s_i} ds_i + \frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2} \frac{\partial^2 a_n}{\partial s_i \partial s_j} ds_i ds_j$$

So it follows that

$$d(s_1/s_2) = \frac{1}{s_2} ds_1 - \frac{S_1}{S_2^{\,2}} ds_2 + \frac{1}{2}[- \frac{1}{s_2^{\,2}} ds_2 ds_1 - \frac{1}{s_2^{\,2}} ds_1 ds_2 + \frac{2s_1}{s_2^{\,3}} ds_2 ds_2] \tag{4.36}$$

To simplify notation let $P_1(w_j) = p_j$. By Lemma 3.2 we have

$$ds_1 = p_n \Lambda_n \hat{Z}_n' R^{-1} dY$$

$$ds_2 = \sum_{i=1}^{M} p_i \Lambda_i \hat{Z}_i' R^{-1} dY$$

Let us now evaluate each of the terms in Equation 4.36:

$$ds_1/s_2^{\,2} = p_n \Lambda_n \hat{Z}_n' R^{-1} dY / \sum_{i=1}^{M} p_i \Lambda_i = a_n \hat{Z}_n' R^{-1} dY \tag{4.37}$$

$$(s_1/s_2^{\,2}) ds_2 = p_n \Lambda_n \sum_{i=1}^{M} p_i \Lambda_i \hat{Z}_i' R^{-1} dY / (\sum_{j=1}^{M} p_j \Lambda_j)^2$$

$$= a_n \sum_{i=1}^{M} a_i \hat{Z}_i' R^{-1} dY \tag{4.38}$$

$$\frac{ds_1 ds_2}{s_2^{\,2}} = \frac{ds_2 ds_1}{s_2^{\,2}} = p_n \Lambda_n \hat{Z}_n' R^{-1} dY \sum_{i=1}^{M} p_i \Lambda_i \hat{Z}_i' R^{-1} dY / (\sum_{j=1}^{M} p_j \Lambda_j)^2$$

$$= a_n \hat{Z}_n' R^{-1} \sum_{i=1}^{M} a_i \hat{Z}_i dt \tag{4.39}$$

$$\frac{s_1}{s_2^{\,3}} ds_2 ds_2 = p_n \Lambda_n (\sum_{i=1}^{M} p_i \Lambda_i \hat{Z}_i' R^{-1} dY)(\sum_{j=1}^{M} p_j \Lambda_j \hat{Z}_j R^{-1} dY)/$$

$$(\sum_{i=1}^{M} p_i \Lambda_i)^3 = a_n (\sum_{i=1}^{M} a_i \hat{Z}_i)' R^{-1} (\sum_{i=1}^{M} a_i \hat{Z}_i) dt \tag{4.40}$$

Equations 4.39 and 4.40 follow from the fact that

$$(\hat{Z}_i' dY)(\hat{Z}_j' dY) = (\sum_{k=1}^{m} \hat{Z}_i(k(dY(d)) \sum_{k=1}^{M} \hat{Z}_j(k)dY(k)$$

where $\hat{Z}_i$ is i^{th} component of $\hat{Z}$ and $Y_{(i)}$ is the component of Y. Since $dY = X_q dt + dV$ for some X_q

$$(\hat{Z}_i' dY)(\hat{Z}_j' dY) = [\sum_{k=1}^{M} \hat{Z}_i(k)(X_q(k)dt+dV(k))][\sum_{k=1}^{m} \hat{Z}_j(k)(X_r(k)dt+dV(k))]$$

$$= \sum_{k=1}^{m} \hat{Z}_i(k)\hat{Z}_j(k)dt = \hat{Z}_i'\hat{Z}_j dt$$

Combining 4.36, 4.37, 4.38, 4.39, and 4.40 we get 4.32. So we have proved that 4.35 is a solution of 4.32. To show that 4.35 is the only solution of 4.32 with the initial condition $P_1(w_n)$, we shall first show that it is the only solution of the following stochastic differential equation with this initial condition:

$$da_n(t) = a_n(t)(\hat{Z}_n(t))' R^{-1}(t)(dY-\hat{Z}(t)dt) \tag{4.41}$$

Note that we have just replaced $\sum_{i=1}^{M} a_i \hat{Z}_i$ by $\hat{Z}$ in 4.32. Now let x_1 and x_2 be two solutions of 4.32; from 4.36 we have

$$d(x_1/x_2) = \frac{x_2 dx_1 - x_1 dx_2}{x_2^{\,2}} + \frac{x_1 dx_2 dx_2 - x_2 dx_1 dx_2}{x_2^{\,3}} \tag{4.42}$$

It is clear that $x_2 dx_1 - x_1 dx_2 = 0$. From the reasoning used on 4.39 and 4.40 it follows that

$$x_1 dx_2 dx_2 = x_1 x_2 (\hat{Z}_n - \hat{Z})' R^{-1} (dY - \hat{Z}dt) x_2 (\hat{Z}_n - \hat{Z})' R^{-1} (dY - \hat{Z}dt)$$
$$= x_1 x_2 x_2 (\hat{Z}_n - \hat{Z})' R^{-1} (\hat{Z}_n - \hat{Z}) dt$$

Similarly, $x_2 dx_1 dx_2 = x_2 x_1 x_2 (\hat{Z}_n - \hat{Z})' R^{-1} (\hat{Z}_n - \hat{Z}) dt$. We, therefore, have

$$d(x_1/x_2) = 0 \rightarrow \frac{x_1(t)}{x_2(t)} - \frac{x_1(a)}{x_2(a)} = 0 \qquad \text{for all } t \in [a,b]$$

$$x_2(t) = \frac{x_2(a)}{x_1(a)} x_1(t) \qquad \text{for all } t \in [a,b]$$

So if $x_1(a) = x_2(a)$, we must have $x_1(t) = x_2(t)$ on $[a,b]$. Now the next question to answer is whether we can replace $a_n(t)$ in the expression for $\hat{Z}(t)$ equation 4.31 by $\bar{a}_n(t)$ where $\bar{a}_n$ is the solution of the differential equation 4.41 with the prescribed a priori probability as the initial condition or not. The answer is affirmative. This can be proved by induction. It is certainly true for $N = 1$. Suppose the claim is true for $N = M$. Consider

$$\hat{Z} = \sum_{n=1}^{M} a_n \hat{Z}_n + a_{M+1} \hat{Z}_{M+1}$$

Using the fact that $a_k \hat{Z}_k$ is Y_t measurable, we have

$$\hat{Z} - \sum_{n=1}^{M} a_n \hat{Z}_n = E(\hat{Z} - \sum_{n=1}^{M} a_n \hat{Z}_n | Y_t) = a_{M+1} \hat{Z}_{M+1} = \bar{a}_{M+1} \hat{Z}_{M+1}$$

$$\hat{Z} - a_{M+1} \hat{Z}_{M+1} = E(\hat{Z} - a_{M+1} \hat{Z}_{M+1}) = \sum_{n=1}^{M} a_n \hat{Z}_n = \sum_{n=1}^{M} \bar{a}_n \hat{Z}_n$$

So that

$$Z = \sum_{n=1}^{M+1} \bar{a}_n \hat{Z}_n$$

The proof is complete.*

Remark: As pointed out in the proof of the theorem, the *a posteriori* probability is the unique solution of

$$da_n(t) = a_n(t)(\hat{Z}_n(t)-\hat{Z}(t))'R^{-1}(t)(dY(t)-\hat{Z}(t)dt) \qquad (4.43)$$

with $a_n(t) = P_1(w_n)$ at $t = a$ for $n = 1,2,\ldots,M$. The form of the solution as in 4.43 gives an innovation interpretation of the results by noting that $dY - \hat{Z}dt$ is the innovation of the process Y [14].

We shall now apply these results to linear continuous stochastic dynamical systems. We summarize the results in the next corollary.

Corollary 4.12. Consider the detection and estimation problem posed by Theorem 4.11 with $X_i(t) = C_i(t)x(t)$ where $x(t)$ satisfies Equation 3.1. The least squares estimate of $x(t)$ is given by

$$\hat{x}(t) = \sum_{i=1}^{M} a_i(t)\hat{x}_i(t) \qquad (4.44)$$

where $\hat{x}_i$ satisfies Equation 3.17 and the coefficient $a_i(t)$ is the unique solution of the following stochastic differential equation

$$da_i(t) = a_i(t)(C_i(t)\hat{x}_i(t) - \sum_{j=1}^{M} a_j(t)C_j(t)\hat{x}_j(t))'$$

$$R^{-1}(t)(dY - \sum_{j=1}^{M} a_j(t)C_j(t)\hat{x}_j(t)dt) \qquad (4.45)$$

with $a_i(t) = P_1(w_i)$ at $t = a$. The closed form solution of the coefficient a_i is given by

$$a_i(t) = \frac{P_1(w_i)\Lambda_i(t)}{\sum_{j=1}^{M} P_1(w_j)\Lambda_j(t)} \tag{4.46}$$

where

$$\Lambda_j(t) = \exp \int_a^t \hat{x}_j'(s)C_j'(s)R^{-1}(s)dY(s) - \frac{1}{2}\int_a^t \hat{x}_j'(s)C_j'(s)R^{-1}(s)C_j(s)\hat{x}_j(s)ds$$

Proof: The proof follows from Theorem 4.16.*

Remark: Comparing the results in the above corollary and the results obtained by nonlinear filtering earlier (Lemma 3.4 and Theorem 3.6), we see that we have, in fact, found the solution for the stochastic differential equation of 3.29 that the indicator variables satisfy.

So the optimal parameter adaptive estimation problem for general second order vector random processes is completely solved in this chapter. The results are applied to the joint detection and estimation problem in linear continuous stochastic dynamic systems. In the next chapter, we shall apply the results of this chapter to find some analogous results for linear discrete time stochastic dynamic systems.

Chapter V

JOINT DETECTION AND ESTIMATION IN LINEAR DISCRETE STOCHASTIC DYNAMICAL SYSTEMS

The discrete stochastic systems are simpler to deal with than continuous stochastic dynamic systems. In this chapter, we shall derive the discrete version of some of the results of Chapter III. These results are fairly well-known in the literature [35], [24], [26]. Our treatment will be concise. We start with the following Lemma.

Lemma 5.1. Let $v(k)$ be a zero mean white Gaussian sequence of random vectors with $Ev(k)v'(j) = R\delta_{kj}$ for some positive definite matrix R and let $z(k)$ be a Gaussian (necessarily second order) sequence of random vectors independent of v; δ_{kj} is the Kronecker delta. Consider the sequence of random m-dimensional vectors $y(k)$ defined by

$$y(k) = z(k) + v(k) \qquad k = 1,2,\ldots \tag{5.1}$$

Let $Y(k) = (y(1),y(2),\ldots,y(k))$. The probability density of $Y(k)$ is given by

$$P_{Y(k)}(y(1),\ldots,y(k)) = \prod_{i=1}^{k} \frac{1}{(2\pi)^{m/2}|Q(i)|^{1/2}}$$

$$\exp\left\{-\frac{1}{2}\sum_{i=1}^{k}(r'(i)Q^{-1}(i)r(i)\right\}$$

where $|Q(i)| = \det(Q(i))$, $r(i) = y(i) - \hat{z}(i|i-1)$, (5.2)

$$Q(i) = E(\tilde{z}(i|i-1)\tilde{z}'(i|i-1)|Y_{i-1}) + R$$

with $\tilde{z}(i|i-1) = z(i) - \hat{z}(i|i-1)$ and $\hat{z}(i|i-1) = E(z(i)|Y_{i-1})$ where Y_{i-1} is the σ-algebra generated by $Y(i-1)$.

Proof: From the chain rule for probability densities

$$p_{Y(k)}(.) = p_{y(1)}(.) \sum_{i=2}^{k} p_{y(i)|Y(i-1)}(.) \tag{5.3}$$

Now

$$E(y(i)|Y_{i-1}) = E(z(i)|Y_{i-1}) + E(v(i)|Y_{i-1}) = \hat{z}(i|i-1) \tag{5.4}$$

Also

$$\begin{aligned} &E([y(i)-\hat{y}(i|i-1)]\ [\hat{y}(i|i-1)]'|Y_{i-1}) \\ &\quad = E(\tilde{z}(i|i-1)\tilde{z}'(i|i-1)|Y_{i-1}) + E(v(i)v'(i)|Y_{i-1}) \\ &\qquad + E((z(i)-\hat{z}(i|i-1))v'(i)|Y_{i-1}) \\ &\qquad + E(v(i)(z(i)-\hat{z}(i|i-1))'|Y_{i-1}) \end{aligned} \tag{5.5}$$

Now

$$\begin{aligned} E((z(i)-\hat{z}(i|i-1))v'(i)|Y_{i-1} &= E(z(i)v'(i)|Y_{i-1}) \\ &\quad - E(\hat{z}(i|i-1)v'(i)\ Y_{i-1}) = 0 \end{aligned}$$

The first term is zero because z and v are independent and v has zero mean; the second term is zero since $\hat{z}(i|i-1)$ is Y_{i-1} measurable and v has zero mean. So 5.5 becomes

$$\begin{aligned} &E([y(i)-\hat{y}(i|i-1)]\ [y(i)-\hat{y}(i|i-1)]'|Y_{i-1}) \\ &\quad = E(\tilde{z}(i|i-1)\tilde{z}'(i|i-1))|Y_{i-1}) + R \end{aligned} \tag{5.6}$$

The proof follows from 5.3, 5.4, 5.6, and the Gaussian assumption.*

Theorem 5.4. Consider the joint detection and estimation problem posed by:

$$y(k) = z(k) + v(k) \qquad k = 1,2,\ldots,N$$

where $v(k)$ is as in Lemma 5.1, $z(k) \in \{x_1(k),\ldots,x_M(k)\}$ where each $x_j(k)$ is a Gaussian random sequence that in independent of $v(k)$. The a priori probability P_j of the event which represents $z(t) = x_j(t)$ is given such that

$$\sum_{j=1}^{M} P_j = 1.$$

The least-squares estimate of $z(t)$ is given by

$$\hat{z}(k) = \sum_{n=1}^{M} d_n(k)\hat{z}_n(k) \tag{5.7}$$

where $z_n(k) = E(z(k)|Y_k, w=w_n)$. The coefficient $d_n(k)$ is the a posteriori probability of the hypothesis, h_n being true given the observations $Y(k)$ and it is given by:

$$d_n(k) = \frac{P_n\Lambda_n(k)}{\sum_{i=1}^{M} P_i\Lambda_i(k)} \tag{5.8}$$

where

$$\Lambda_n(k) = \prod_{i=1}^{k} \frac{1}{|Q_n(i)|^{1/2}} \exp\{-\frac{1}{2}\sum_{i=1}^{k} r_n'(i)Q_n^{-1}(i)r_n(i)\} \tag{5.9}$$

with

$$Q_n(i) = E(\tilde{z}_n(i|i-1)\tilde{z}_n'(i|i-1)|Y_{i-1}) + R$$

$$r_n(i) = y(i) - \hat{z}_n(i|i-1)$$

Also, the coefficients $d_n(k)$ satisfy the difference equations

$$d_n(k) = \frac{L_n(k)}{\sum_{i=1}^{M} L_i(k)d_i(k-1)} d_n(k-1) \qquad \begin{matrix} n = 1,2,\ldots,M \\ k = 1,2,\ldots \end{matrix} \tag{5.10}$$

where

$$L_i(k+1) = \frac{1}{|Q_i(k+1)|^{1/2}} \exp\{-\frac{1}{2} r_i'(k+1)Q_i^{-1}(k+1)r_i(k+1)\}$$

with $d_n(0) = P_n$.

Proof: Equation 5.7 follows from Lemma 4.13 with $\Omega_2 = (R^m)^k$. Equations 5.8 and 5.9 follow from Lemma 4.12 with μ_0 as the Lebesgue measure on R^{mk} and Lemma 5.1. To prove 5.10

$$d_n(k) = \frac{P_n\Lambda_n(k)}{\sum_{i=1}^{M} P_i\Lambda_i(k)} = \frac{L_n(k)P_n\Lambda_n(k-1)}{\sum_{i=1}^{M} P_i\Lambda_i(k)}$$

$$= \frac{L_n(k)}{\sum_{i=1}^{M} P_i\Lambda_i(k)} \frac{P_n\Lambda_n(k-1)}{\sum_{j=1}^{M} P_j\Lambda_j(k-1)} \sum_{j=1}^{M} P_j\Lambda_j(k-1) \tag{5.11}$$

Since

$$\frac{\sum_{i=1}^{M} P_i\Lambda_i(k)}{\sum_{j=1}^{M} P_j\Lambda_j(k-1)} = \sum_{i=1}^{M} \frac{P_i\Lambda_i(k-1)L_i(k)}{\sum_{j=1}^{M} P_j\Lambda_j(k-1)} = \sum_{i=1}^{M} d_i(k-1)L_i(k) \tag{5.12}$$

Combining 5.11 and 5.12 we get

$$d_n(k) = \frac{L_n(k)}{\sum_{i=1}^{M} d_i(k-1)L_i(k)} d_n(k-1)$$

which verifies 5.10.*

These results can be applied to the case where x is generated by one of the following systems (with x being generated by the i^{th} system, having a probability P_i):

$$\begin{aligned} x(k+1) &= A_i(k)x(k) + B_i(k)w_i(k) \\ y(k+1) &= C_i(k)x(k) + v_i(k) \end{aligned} \qquad i = 1,2,\ldots,M \tag{5.13}$$

where w_i is a zero mean white Gaussian sequence of random vectors of dimension m with E $w_i(k)w_i'(j) = Q_i(k)\delta_{kj}$, v_i is a zero-mean Gaussian sequence of random vectors of dimension m independent of w_i with E $v_i(k)v_i'(j) = R_i(k)\delta_{kj}$, and A_i, B_i, C_i are matrices of appropriate dimension. In this case, the least squares estimate of x will be given by

$$\hat{x}(k) = \sum_{j=1}^{M} d_i(k)\hat{x}_i(k) \tag{5.14}$$

where $\hat{x}_i(k) = E(x(k)|Y_k, h=h_i)$. The coefficient $d_i(k)$ is given by

$$d_i(k) = \frac{P_n\Lambda_n(k)}{\sum_{i=1}^{M} P_i\Lambda_i(k)} \tag{5.15}$$

where now

$$\Lambda_n(k) = \sum_{i=1}^{k} \frac{1}{|Q_n(i)|^{1/2}} \exp \{-\frac{1}{2} \sum_{i=1}^{k} r_n'(i)Q_n^{-1}(i)r_n(i)\} \quad (5.16)$$

with

$$r_n(i) = y(i) - C_n(i)A_n(i-1)\tilde{x}_n(i-1) \quad (5.17)$$

$$Q_n(i) = C_n(i)V_{\tilde{x}_n}(i|i-1)C_n'(i) + R_n(i) \quad (5.18)$$

where

$$V_{\tilde{x}_n}(i|i-1) = E((x(i) - \tilde{x}_n(i\ i-1))$$
$$(x(i) - \tilde{x}_n(i\ i-1))'|Y_{i-1},\ h=h_n) \quad (5.19)$$

where $V_{\tilde{x}_n}$ is the prediction error covariance of n^{th} filter that generates $\tilde{x}_n$. The equations that generate $\tilde{x}_n$ can be found in [25].

In the next chapter, we will apply the results of this and the previous chapters to the design of a digital flight control system which is optimally tolerant of sensor failures.

Chapter VI

APPLICATIONS TO FLIGHT CONTROL SYSTEMS

We shall now show applications of some of the results developed in the previous chapters to flight control systems. Namely, we shall present a design method for digital self-reorganizing control system which is optimally tolerant of failures in aircraft sensors, reported by Caglayan and Montgomery in [34]. The design process will be applied to the design of self-reorganizing control system for a current configuration of the space shuttle orbiter at Mach 5 and 120,000 feet. The failure detection capabilities of the system are demonstrated using a real-time simulation of the system with noisy sensors.

Before we go on, we would like to remark that the problem is one of joint detection, estimation, and control. Separation between estimation and control has not been shown to be valid in this problem up to now, and we have not addressed ourselves to this problem in the text. In the simulations, we have used the steady-state regulator gains and the estimate corresponding to the most likely hypothesis in the Bayesian sense.

Consider the equations of motion of an aircraft to be represented by

$$x(t) = A \int_a^t x(s)ds + B \int_a^t U(s)ds + W(t) \qquad t \in [a,b] \tag{6.1}$$

where x is the n-dimensional state vector, U is the m-dimensional control vector, W(t) is a zero mean Wiener process with W(a) = 0 and

$E\,W(t)W'(s) = \min\{(t-a),(s-a)\}\,Q_1$. In equation 6.1, the matrices A and B are determined from the aircraft's stability and control derivatives, its mass and inertia characteristics, and its geometric characteristics. The variable W may, but need not, represent turbulence. It may represent uncertainty in the designer's knowledge of the aircraft's characteristics. We shall be interested in the digital control of the plant which constrains the control to be piecewise constant with a sampling period T; i.e.,

$$U(t) = U(kT) \qquad \text{for } t \in [kT,(k+1)T) \tag{6.2}$$

So that we get the discrete equations of motion of the aircraft

$$x(k+1) = \Phi x(k) + \Gamma U(k) + w(k) \tag{6.3}$$

where $\Phi = \exp(AT)$, $\Gamma = \int_0^T \exp(As)ds\,B$, and $w(k)$ is a zero mean white Gaussian sequence of random vectors with

$$E\,w(k)w'(j) = [\int_0^T \exp(As)Q_1 \exp(A's)ds]\,\delta_{kj} = Q\delta_{kj}$$

(δ_{kj} is the Kronecker delta)

We shall hypothesize the following failure-mode observation models for the aircraft.

$$y(k) = C_i x(k) + v_i(k) \qquad i = 1,2,\ldots,M-1 \tag{6.4}$$

where $v_i(k)$ is a white Gaussian sequence of random vectors with $E(v_i(k)) = (0,0,\ldots,m_i',\ldots)'$ and $E(v_i(k)v_i'(j)) = R_i\delta_{kj}$ for some positive definite matrix R_i. The quantity m_i is the unknown parameter vector of the distribution. The approach that we will take is one of approximation. We shall solve the problem as if m_i were known and then use the maximum likelihood estimate of m_i under the i^{th} hypothesis. This procedure is known as a generalized likelihood ratio approach in the communications literature [12]. For the normal unfailed operation we will assume

$$y(k) = C_0 x(k) + v_0(k) \tag{6.5}$$

where $v_0(k)$ is a white Gaussian sequence of random vectors with $Ev_0(k) = 0$ and $Ev_0(k)v_0'(j) = R_0 \delta_{kj}$. As usual, we shall introduce an extra dummy hypothesis h_M by

$$h_M: \;\; y(k) = v_M(k) \tag{6.6}$$

where $v_M(k)$ is a white Gaussian sequence of random vectors with $Ev_M(k) = 0$ and $Ev_M(k)v_M'(j) = R_M \delta_{kj}$.

We will be concerned with the selection of the most probable hypothesis based on a finite set of measurements $\{y(1),y(2),\ldots,y(k)\} = Y(k)$. We shall use the detection theory results of Chapter II and choose the hypothesis that will result in minimum Bayesian risk. To this end, we need to evaluate the likelihood ratios $\{\Lambda_i,\ i = 0,.,\ldots,M-1\}$ where Λ_i is the ratio of the probability densities

$$\Lambda_i = p_{Y(k)}(y(1),\ldots,y(k)|\, h_i)/p_{Y(k)}(y(1),\ldots,y(k)|h_M)$$

For a fixed K number of observations, from the results of the last chapter, we have for the likelihood ratios

$$\ln \Lambda_i = \ln\left(\sum_{j=1}^{K} |R_M|^{1/2}/|Q_i(j)|^{1/2}\right) - \frac{1}{2}\sum_{j=1}^{K} r_i'(j)Q_i^{-1}(j)r_i(j) \tag{6.7}$$

where $Q_i(j) = C_i V_{\tilde{x}_i}(j\ j-1)C_i' + R_i$ with $V_{\tilde{x}_i}$ being the prediction error covariance of the i^{th} filter, $r_i(j)$ is the innovation sequence of the i^{th} filter given by

$$r_i(j) = y(j) - C_i(\Phi \hat{x}_i(j-1) + \Gamma u(j-1)) - \hat{m}_i(j) \tag{6.8}$$

where $\hat{x}_i(j) = E(x(j)|Y(j),h_i)$ and $\hat{m}_i$ is the maximum likelihood estimate of m_i under hypothesis h_i. Actually, the true value of $m_i(j)$ should be used in 6.8. When the maximum likelihood estimate, $\hat{m}_i(j)$ is used instead of $m_i(j)$, then 6.7 becomes a generalized likelihood ratio.

We also assign the a priori probabilities $\{P(h_i) = P_i,\ i = 0,.,\ldots,M\}$. Of course, $P_M = 0$. When the cost of making a right decision is zero and the cost of making a wrong one are all equal, from the results in Chapter II it follows that we have to choose the maximum of

$$\{P_i\Lambda_i,\ i = 0,1,\ldots,M-1\} \tag{6.9}$$

Combining equations 6.7 and 6.9, it follows that we have to choose the maximum of

$$\ln P_i - \ln \sum_{j=1}^{K} |Q_i(j)|^{1/2} - \frac{1}{2}\sum_{j=1}^{K} r_i'(j)Q_i^{-1}(j)r_i(j) \qquad i = 0,.,\ldots,M-1 \tag{6.10}$$

When the steady-state filters are used and the a priori probabilities are all equal, then an equivalent criterion that would result in the minimum Bayesian risk is to choose the minimum of the set $\{\tau_i, i = 0,1,\ldots,M-1\}$ where

$$\tau_i = \frac{K}{2}\ln|Q_i| + \frac{1}{2}\sum_{j=1}^{K} r_i'(j)Q_i^{-1}r_i(j) \tag{6.11}$$

The above considerations are implemented on a real time hybrid simulation using the lateral dynamics for a current space shuttle orbiter with A and B matrices:

$$A = \begin{bmatrix} -0.058 & 0 & 0.017 & -5.791 \\ 1.0 & 0 & 0.5773 & 0 \\ -0.0029 & 0 & -0.0085 & -0.7438 \\ -0.5 & .0055 & -0.8660 & -0.0009 \end{bmatrix}$$

$$B = \begin{bmatrix} 2.256 \\ 0 \\ 0.0553 \\ 0 \end{bmatrix}$$

The state vector $x = (p,\emptyset,r,\beta)'$ and the control vector $v = \delta_a$ where p is the roll-rate in degrees per second, $\emptyset$=is the bank angle in degrees, r is the yaw rate in degrees per second, β is the sideslip angle in degrees, and δ_a is the aileron input angle in degrees. The vehicle in consideration has a roll-rate gyro, a yaw-rate gyro, and a sideslip indicator. So we have the following observation models:

$$h_0:\ y = [p\ r\ \beta]' + v_0$$

$$h_1:\ y = [0\ r\ \beta]' + v_1$$

$$h_2: \ y = [p \ 0 \ \beta]' + v_2$$

$$h_3: \ y = [p \ r \ 0]' + v_3$$

The variances Q_1, R_o, R_1, R_2, R_3 are chosen as

$$Q_1 = \text{diag } (2(.05)^2, 0, 2(1.01)^2, 3(.001)^2)$$

$$R_o = (.0025, .0001, .0001)$$

$$R_1 = \text{diag } (.025, .0001, .0001)$$

$$R_2 = \text{diag } (.0025, .001, .0001)$$

$$R_3 = \text{diag } (.0025, .0001, .01)$$

That is, h_o represents the normal operation under no sensor failures, h_1 represents the failure of the roll-rate gyro and so on. For each hypothesis, the system is observable, but, given the measurement errors and uncertainties in the vehicle equations of motion, each sensor group has a different capability of estimating the aircraft's state. In this example, we have used steady-state filters and simplified Bayesian risk weights so that Equation 6.11 is valid. The sampling period used is .1 seconds. A moving window of innovations sequence of length 5 is implemented. Using the innovation sequence in each moving window, Equation 6.11 is evaluated. The estimate of the filter corresponding to the most likely hypothesis is used. The structure of the system is shown in Figure 6.1. Figure 6.2 illustrates the unaugmented and the augmented step response of the non-minimum phase vehicle to an aileron input. Figure 6.3 shows the step response of the closed-loop aircraft with feedback gains (-4.9, 0.4, 14.5, -6.0) using state variable feedback with noisy measurements under hypothesis h_o. In Figure 6.4, the responses of the system are indicated for the case when h_2 is true but for each hypothesis being accepted at different times. The failure mode in Figure 6.4 is an increase in measurement noise. Figures 6.5 and 6.6 illustrate the actual performance of the system under increased noise type failure. By looking at the (p,r,β) measurements, it can be seen that the following sensor failure mode sequence is simulated: $\{h_o,h_1,h_2,h_3,h_2,h_1,h_o\}$. The plot, showing the hypothesis decided, indicates the superior performance of the detector logic. Figures 6.7 and 6.8 deal with the detection of hardover failures. In the first one, a hardover failure in the beta sensor is simulated and a zero hardover failure of the yaw-rate gyro in the second figure.

Note that the steady-state filter gains and the estimate corresponding to the most likely hypothesis is used in this simulation. The results may be improved by using time varying filter gains and by using optimum least squares estimate of the state which is the average of conditioned estimates by the a posteriori probabilities. The overall system performance is pretty good in the sense that the system is capable of identifying increased noise and hardover type sensor failures quite rapidly in the order of one or two sampling periods.

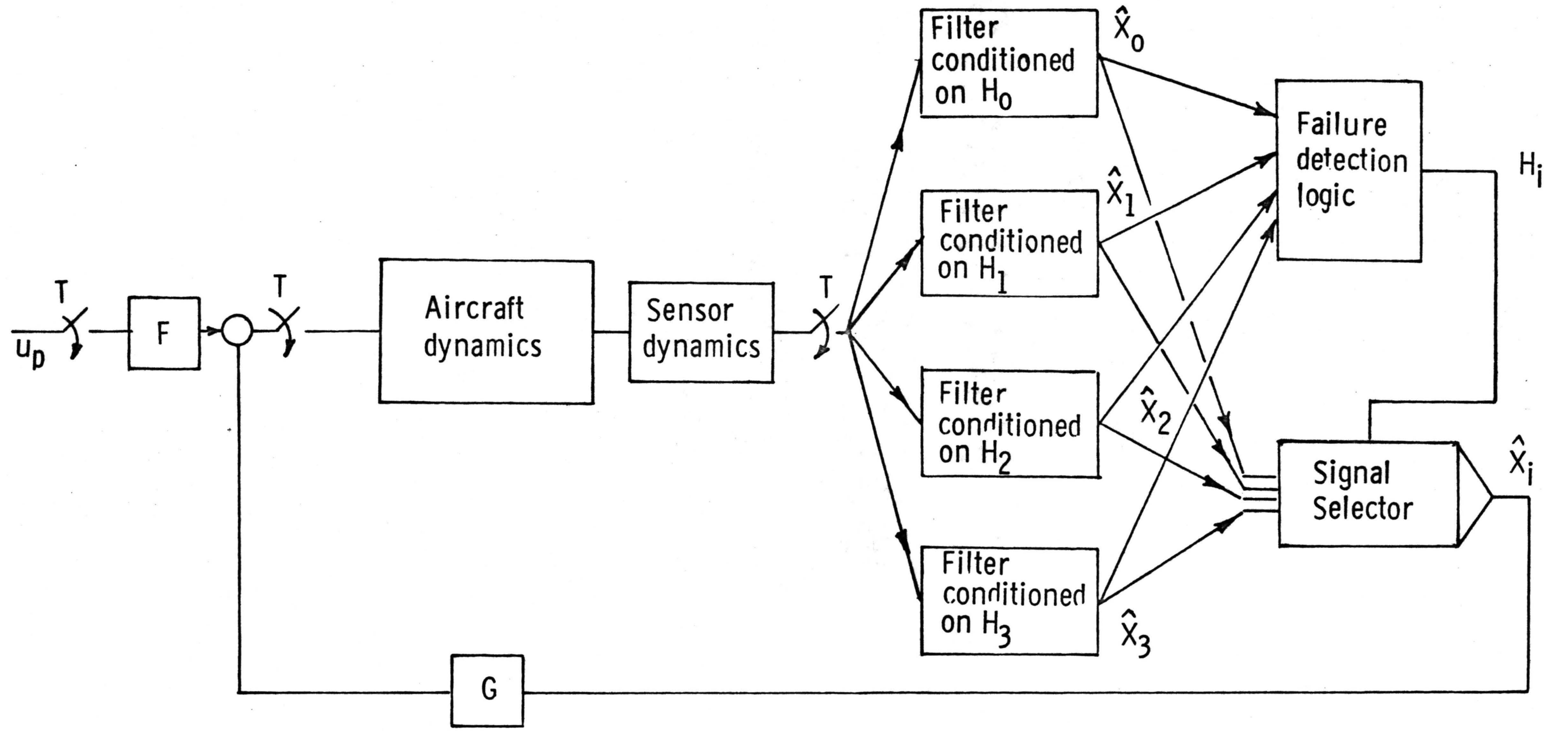

Figure 6.1 Self-Reorganizing System Structure

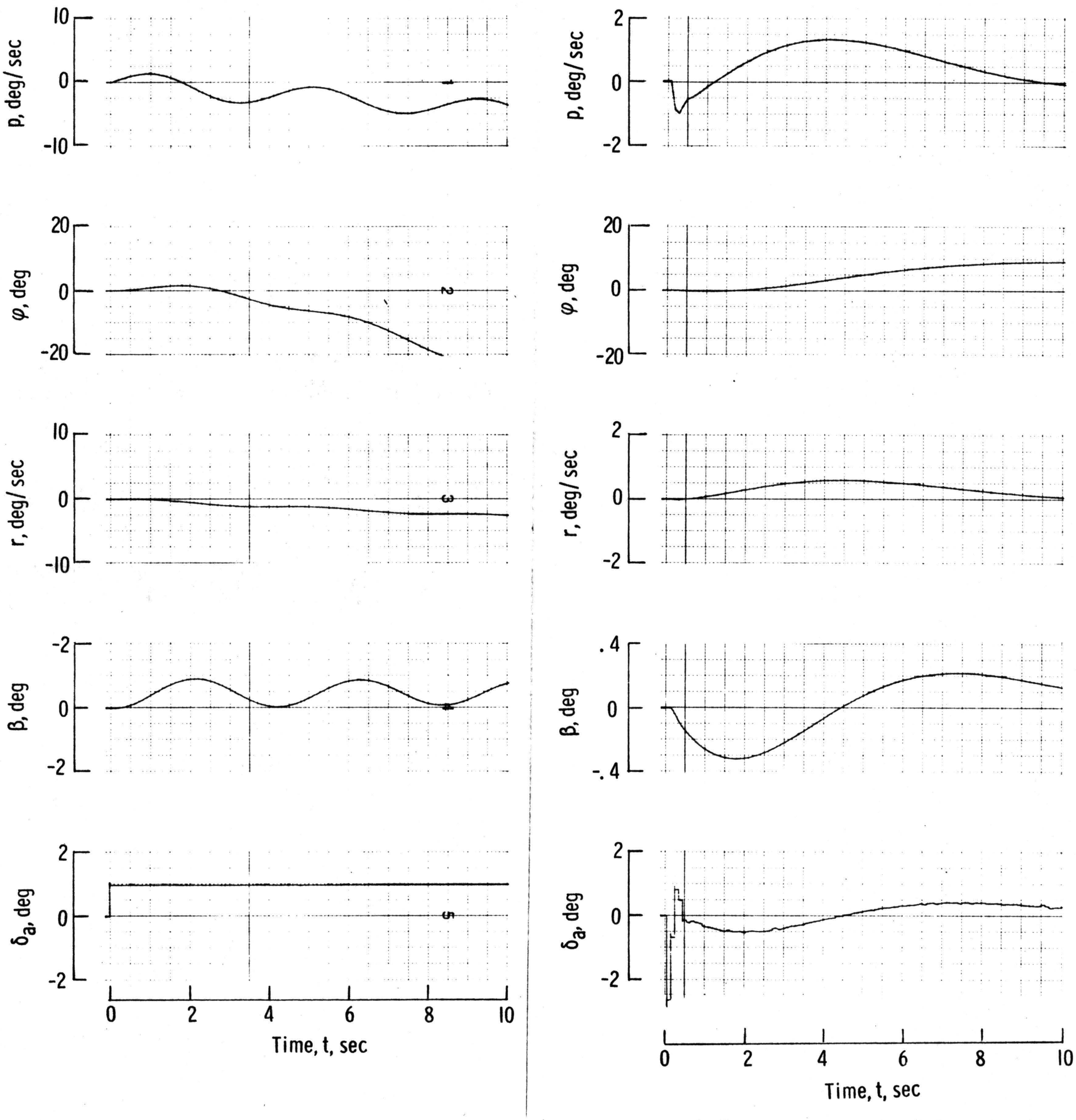

Figure 6.2 Response of the Aircraft to an Aileron Step

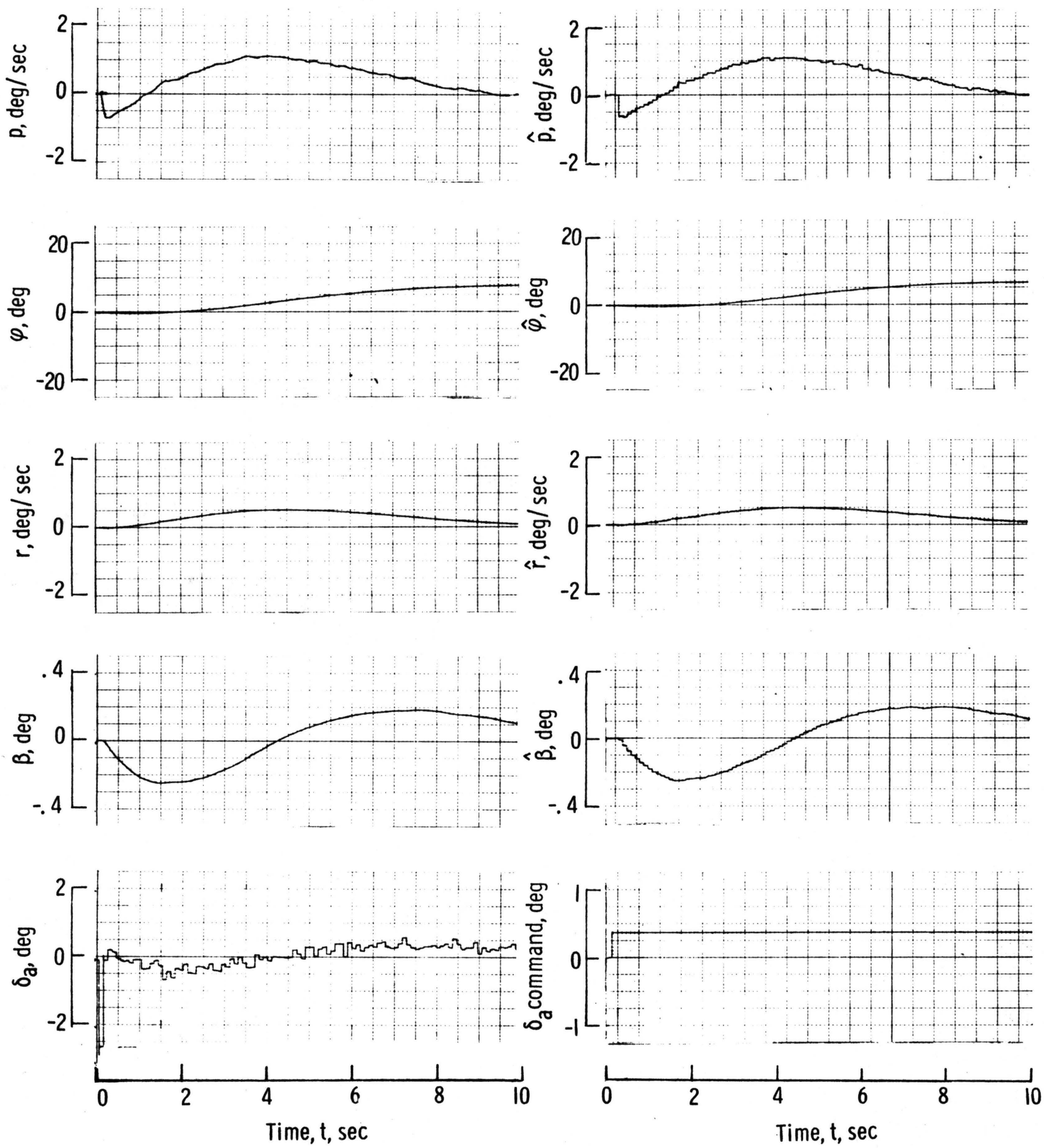

Figure 6.3 Closed-Loop System Under Normal Operation

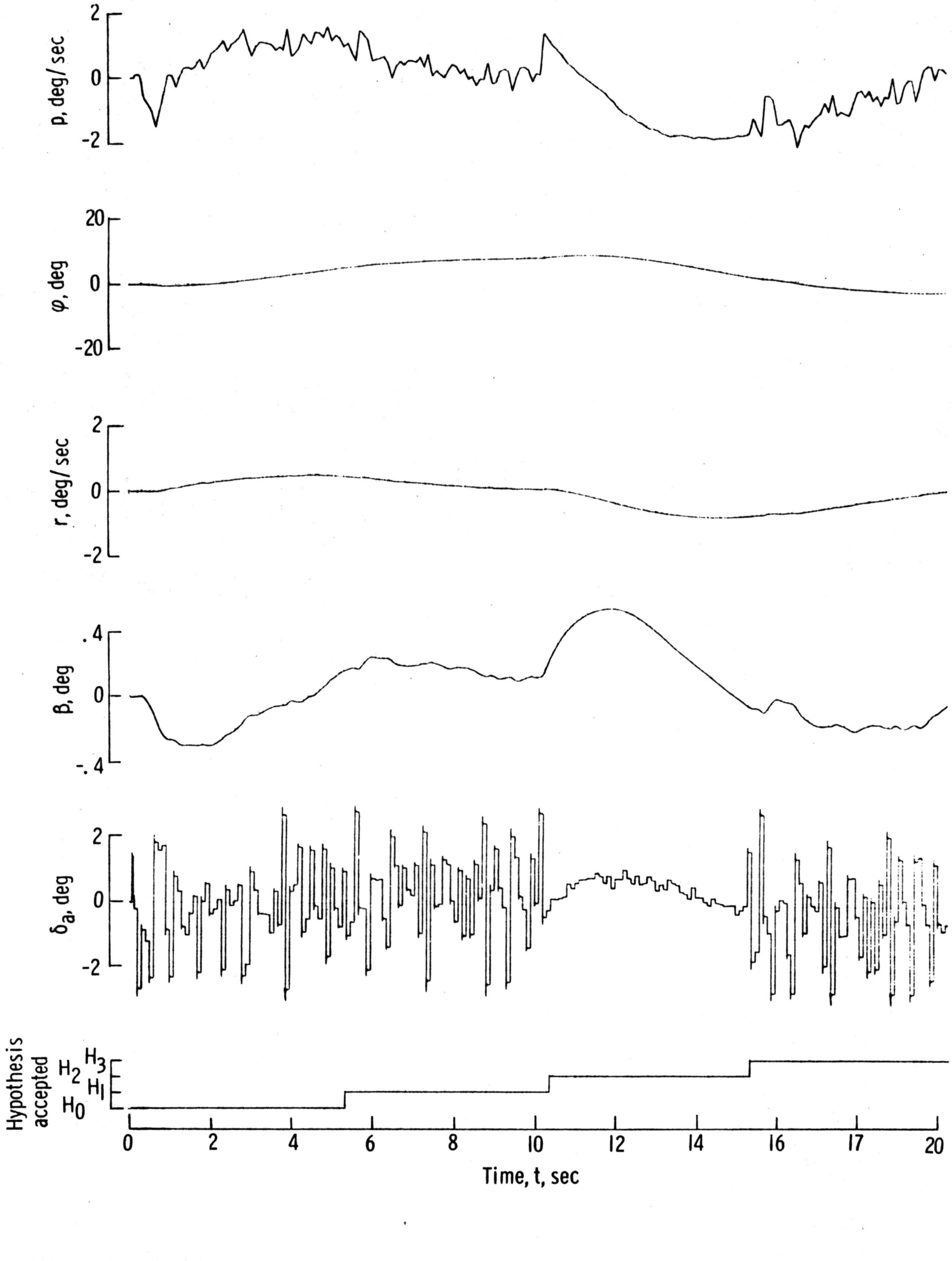

Figure 6.4 Effects of Accepting Various Hypotheses When h_2 is True

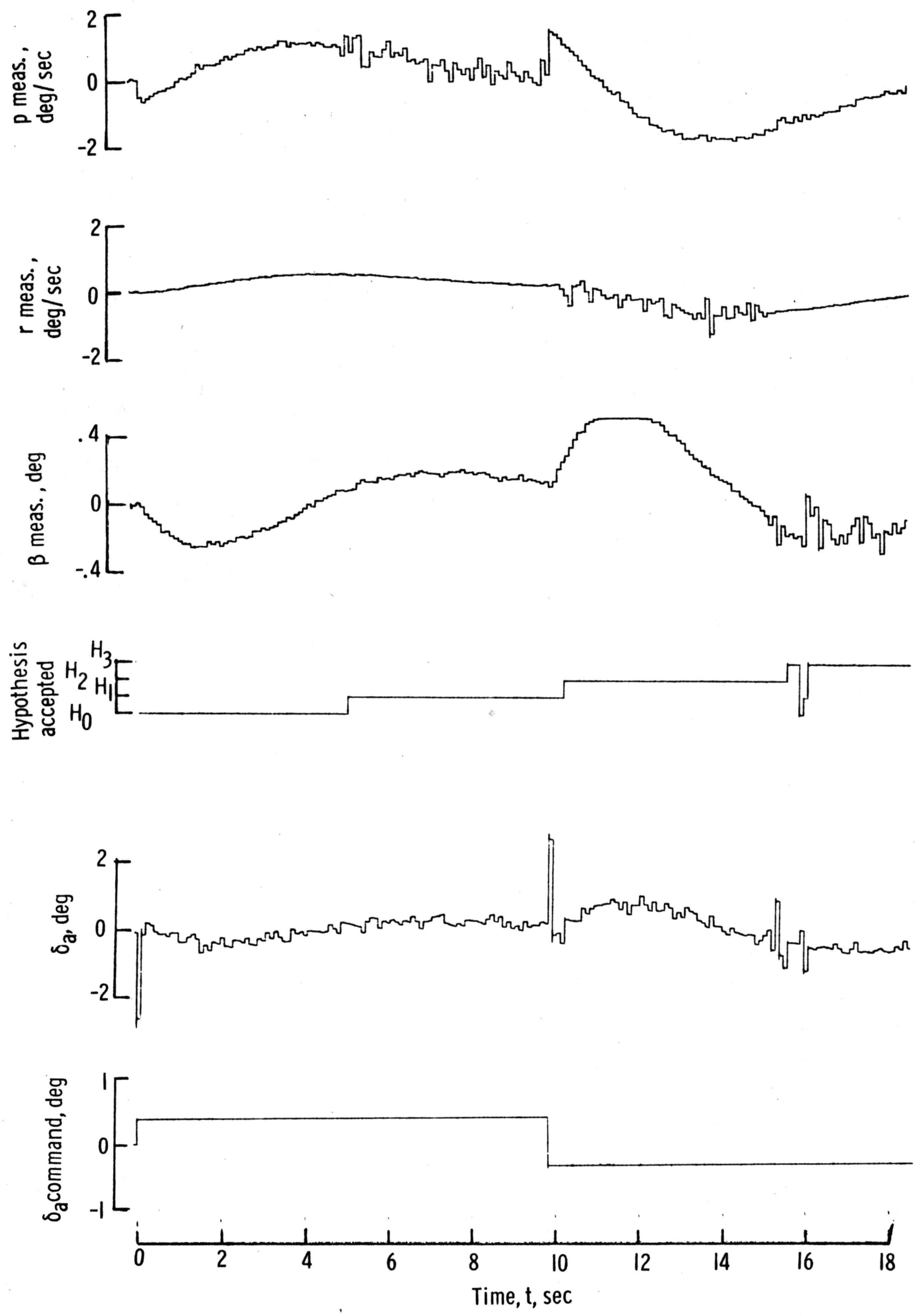

Figure 6.5 Self-Reorganization in Operating During the Failure Sequence (h_o,h_1,h_2,h_3)

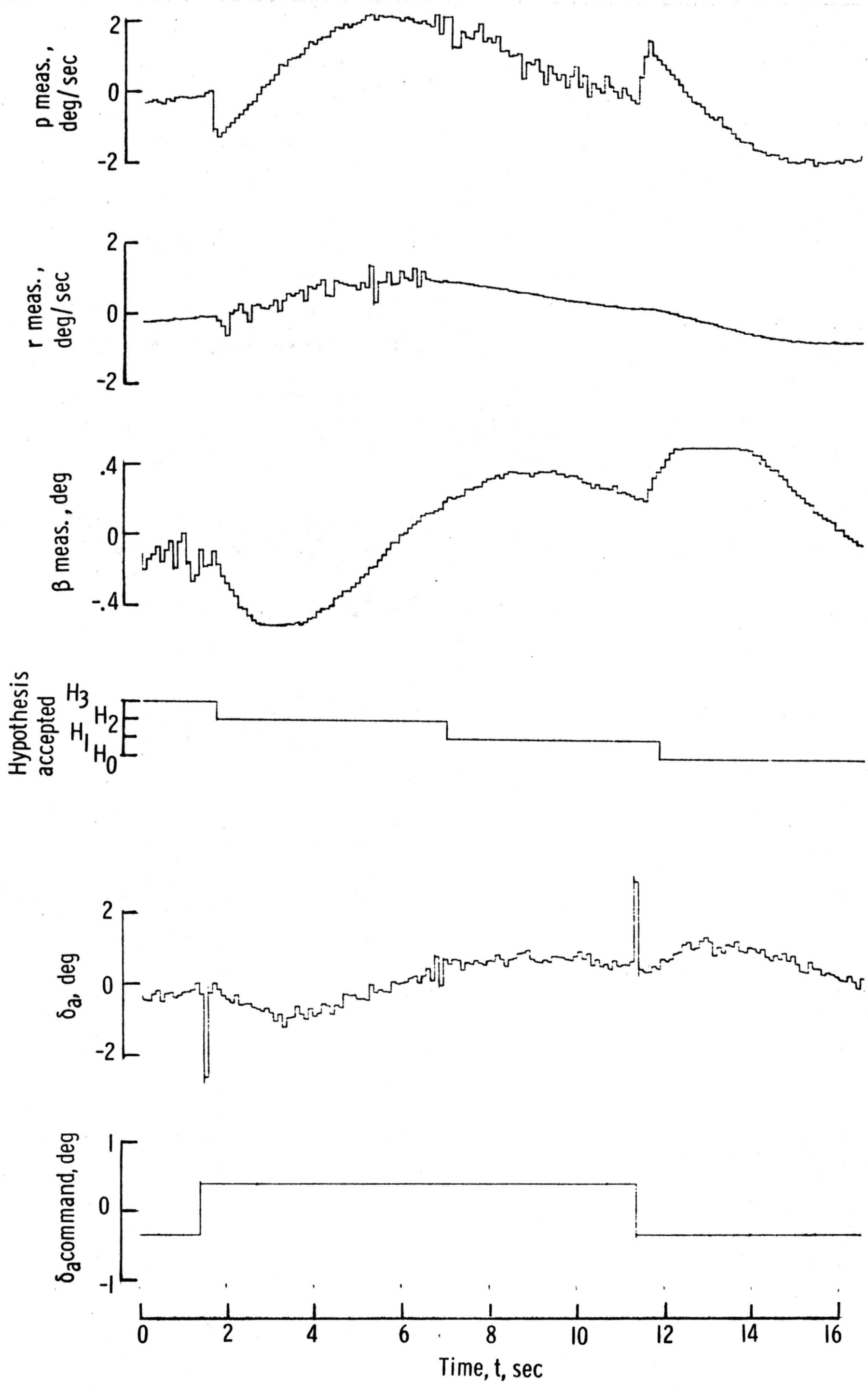

Figure 6.6 Self-Reorganization in Operation During the Failure Sequence (h_3, h_2, h_1, h_o)

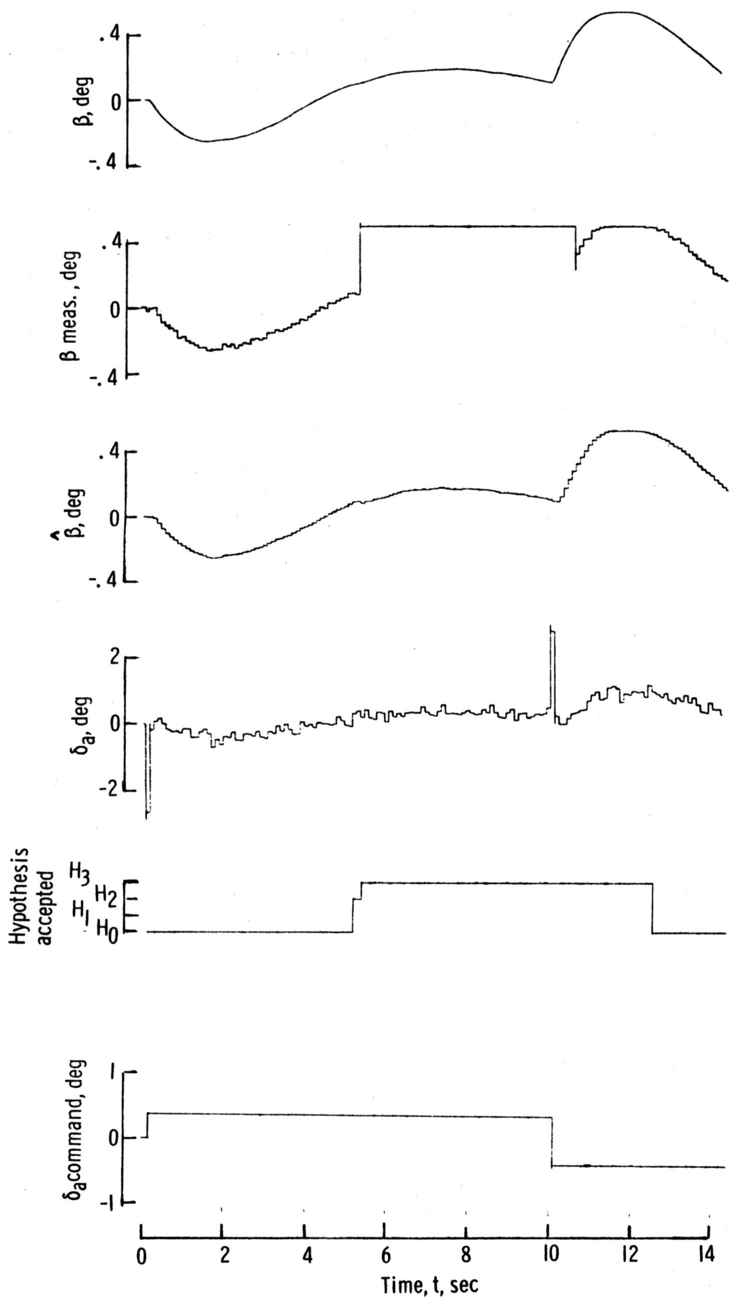

Figure 6.7 Self-Reorganization Operation Under Saturation Hardover Failure

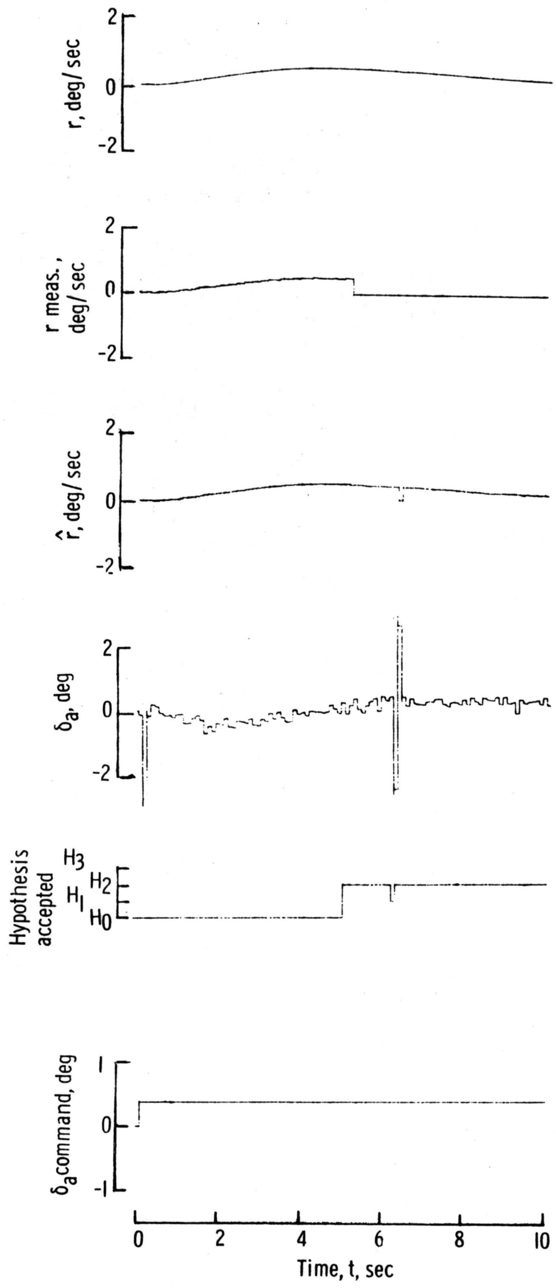

Figure 6.8 Self-Reorganization Operation Under Zero Hardover Failure

Chapter VII

CONCLUSIONS AND RECOMMENDATIONS

A. Conclusions

In this dissertation, we have considered the simultaneous detection and least squares estimation of vector random processes corrupted with Gaussian noise. The previous results in this field have been extended to include random processes which are the solutions of nonlinear stochastic differential equations and general vector random processes which are not necessarily the solutions of stochastic differential equations. In Chapter IV, we have proved an extended version of "the Partition Theorem" of joint detection and estimation. The least squares estimator for this problem is shown to be composed of two parts: A non-adaptive part in which estimates conditioned on each hypothesis are calculated. An adaptive part in which the *a posteriori* probability of each hypothesis conditioned on the observations is calculated using the estimates computed the non-adaptive part. In obtaining this result, we have taken a basic, rigorous measure theoretic approach. By using an extended version of the product measure theorem we have obtained a Radon-Nikodym representation for the *a posteriori* probability. This Representation Theorem (Theorem 4.5 and Lemma 4.7) can be considered as an extension of the classical Bayes Theorem. For the case where there is a finite number of parameters, the stochastic differential equation, that each *a posteriori* probability satisfies, is found. It is also shown that the *a posteriori* probability is the unique solution of the stochastic differential equation that it satisfies with the a priori probability as the initial condition.

In Chapter III, we solved the joint detection and estimation problem for linear continuous dynamic systems through a nonlinear filtering theory approach. Comparing the results of this chapter with the more general results of Chapter IV, it is shown that the nonlinear filtering approach and the measure theoretic approach give the same results in the case of linear continuous dynamic systems.

Some analogous results are derived for discrete dynamic systems in Chapter V. Chapter VI includes an application of the theory to a digital flight control system which is tolerant of sensor failures. Real-time hybrid computer simulations show that this approach has engineering significance.

B. Recommendations for Further Study

This work has been limited to the parameter adaptive estimation in the case when the parameter has a discrete distribution. At the time of the printing of this thesis, the work has been extended by the author to the case where the parameter has a continuous distribution.

Another limitation of this work is that the parameter was considered to be a random variable. That is, the parameter had a constant distribution at all times. Physical implication of this assumption is that one and only one random process is being observed at all times. An interesting extension would be the case where the parameter is a random process. Work along this line, where the parameter is a jump Markov process, has been reported by Wonham [46] and Sworder [47].

The other logical step for further research is to test the feasibility of on line implementation of the stochastic differential equations that the coefficients of the hypothesis conditioned estimates satisfy on the computer. Different discrete numerical algorithms can be derived to approximate these stochastic differential equations and they can be compared by computer simulations.

Appendix I

LINEAR DYNAMIC SYSTEMS AND A HILBERT SPACE APPROACH TO THE REGULATOR PROBLEM

In this appendix we shall review the linear matrix differential equations and study the optimization of linear deterministic dynamic systems in a Hilbert space setting. The main result will be to obtain the solution of the linear quadratic regulator problem without any variational arguments. We shall denote the set of all real valued measurable functions on a closed interval [a,b] of the real line by $\mathcal{F}(a,b)$. Measurable functions constitute a very broad space, e.g., all continuous, piecewise continuous functions and the limits of sequences of measurable functions are measurable functions [1]. Although Lebesgue measurable functions need only be defined almost everywhere (denoted by a.e.; that is, everywhere except at a set of Lebesque measure zero), this non-physical concept introduced is outweighed by the extra class of control problems that can be analyzed in this context.

We shall denote the subset of $\mathcal{F}(a,b)$ consisting of all Lebesgue integrable functions on [a,b] by $L_1(a,b)$. Similarly, let $L_2(a,b)$ be the subset of all square integrable functions on [a,b]; i.e.,

$$L_2(a,b) = \{f \in \mathcal{F}(a,b): \int_a^b |f(t)|^2\, dt < \infty \} \tag{A.1}$$

It is well-known [2] that the set $L_2(a,b)$ is a separable infinite-dimensional Hilbert space over the field of real numbers under pointwise addition and scalar multiplication. The inner product on $L(a,b)$ is defined by:

$$<f,g> = \int_a^b f(t)g(t)dt \qquad \text{for } f \in L_2(a,b) \text{ and } g \in L_2(a,b) \qquad (A.2)$$

The space L_2 is deservedly the most extensively studied of the function spaces, because it is the most fruitful generalization of the finite dimensional Euclidean space.

We shall be mainly interested in the optimization of the linear dynamical system $(A(t),B(t),C(t))$ governed by the differential equation:

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) \qquad t \in [a,b],\ x(a) \text{ known} \qquad (A.3)$$
$$y(t) = C(t)x(t)$$

where $x(t)$ is the n-dimensional state vector, $u(t)$ is the m-dimensional control vector, and $y(t)$ is the r-dimensional measurement vector. $A(t)$, $B(t)$, and $C(t)$ are time-varying (not necessarily continuous) matrices of appropriate dimension. We shall assume that the elements of $A(t)$, $B(t)$, and $C(t)$ are real-valued measurable functions that belong to $L_2(a,b)$.

<u>Remark A.1</u>. If the optimization interval is finite, then the above assumptions imply that the elements of $A(t)$, $B(t)$, and $C(t)$ are integrable. This follows since if $f \in L_2(a,b)$, by Hölder's Inequality

$$\int_a^b |f(t)\cdot 1|dt \leq \left[\int_a^b |f(t)|^2\, dt\right]^{\frac{1}{2}} \left[\int_a^b |1|^2\, dt\right]^{\frac{1}{2}} = \sqrt{b-a}\ ||f||_2$$

So $L_2(a,b) \subset L_1(a,b)$ for $-\infty < a < b < +\infty$.*

In the sequel we shall primarily deal with the integral form of Equation A.3. The solution can be easily obtained from the following two theorems which are a summary of results on the classical theory for system of linear equations.

Theorem A.1. Consider the homogeneous part of equation A.3

$$\dot{x}(t) = A(t)x(t) \qquad t \in [a,b] \tag{A.5a}$$

Given any initial condition x(a) of the elements of A(t) are real-valued measurable functions in $L_1(a,b)$, then there exists a unique solution of equation A.5. The solution is given by

$$x(t) = \psi(t)\psi^{-1}(a)x(a) \tag{A.5b}$$

where $\psi(t)$ is any nxn matrix whose n-columns are n linearly independent solutions of equation A.5a. $\psi(t)$ is called a fundamental matrix of equation A.5a.

The fundamental matrix has the following properties:

1. $\psi(t)$ is absolutely continuous on [a,b] and satisfies the matrix differential equation

$$\dot{\psi}(t) = A(t)\psi(t) \qquad \text{a.e. on } [a,b] \tag{A.5c}$$

2. A necessary and sufficient condition that $\psi(t)$, which satisfies equation A.5c, be a fundamental matrix of equation A.5, is that det $\psi(t) \neq 0 \quad t \in [a,b]$.

3. Every solution of equation A.5 is a unique linear combination of the columns of $\psi(t)$.

4. The fundamental matrix is not unique. Any two fundamental matrices are related to each other by $\psi_2(t) = \psi_1(t)C$ for some non-singular constant C matrix.

5. $\psi(t)$ determines A(t) uniquely a.e. on [a,b].

Proof: (See Thm. 2.1, 2.2, and 2.3 in Chapter 3 of Ref. 3.)

It follows from the property 3 of $\psi(t)$ that the solution x(t) should be of the form $x(t) = \psi(t)c$, where c is an nx1 column vector. Since $x(a) = \psi(a)c$ and $\psi(t)$ is always non-singular, $c = \psi^{-1}(a)x(a)$.*

Theorem A.2. Consider the linear non-homogeneous differential system

$$\dot{x}(t) = A(t)x(t) + b(t) \qquad t \in [a,b] \tag{A.6}$$

where A(t) is an nxn matrix, b(t) is an nx1 column vector. If the elements of A(t) and b(t) are in $L_1(a,b)$, then, for any given initial condition x(a), there exists a unique solution of equation A.6 with initial condition x(a) which is given by

$$x(t) = \psi(t)\psi^{-1}(a)x(a) + \psi(t)\int_a^t \psi^{-1}(s)b(s)ds$$

where $\psi(t)$ is any fundamental matrix of equation A.5a.

Proof: (See Thm. 3.1 in Chapter 3 of Ref. 3.)*

In system theory, the fundamental matrix with $\psi(a) = I$ (where I is the nxn identity matrix) is given the name state-transition matrix because it represents the evolution of the state vector in time for zero-input.

Definition A.1. Let $\psi(t)$ be any fundamental matrix of $\dot{x}(t) = A(t)x(t)$, $t \in [a,b]$, the state-transition matrix of this differential system is defined by

$$\Phi(t,s) = \psi(t)\psi^{-1}(s) \qquad t,s \text{ on } [a,b]$$

The properties of the state-transition matrix can be summarized as follows:

1. The state-transition matrix is unique. This follows since ψ_1, ψ_2 are two different fundamental matrices, they are related by $\psi_2(t) = \psi_1(t)C$ where C is some non-singular constant matrix C.

$$\psi_2(t)\psi_2^{-1}(s) = (\psi_1(t)C)\ (\psi_1(s)C)^{-1}$$
$$= \psi_1(t)CC^{-1}\psi_1^{-1}(s) = \psi_1(t)\psi_1^{-1}(s)$$

2. $\Phi(t,t) = I$
3. $\Phi^{-1}(t,s) = \Phi(s,t)$
4. $\Phi(t,s) = \Phi(t,\tau)\Phi(\tau,s) \quad \tau \in [a,b]$
5. The state transition matrix can be interpreted as the unique fundamental matrix of $\dot{x} = A(t)x$ with $\psi(a) = I$. Therefore, the state-transition matrix satisfies

$$\Phi(t,a) = A(t)\Phi(t,a) \qquad \text{a.e. on } [a,b]$$

with $\Phi(a,a) = I$. The solution of the linear dynamical system described by equation A.3 can now be easily derived from the preceding two theorems.

Corollary A.2. Consider the linear dynamical system governed by the differential equations given in equation A.6.

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) \qquad t \in [a,b] \tag{A.6}$$

If the elements of $A(t)$, $B(t)$, and $u(t)$ are in $L_2(a,b)$, then, for any initial condition $x(a)$, there exists a unique solution of equation A.3 which is given by

$$x(t) = \Phi(t,a)x(a) + \int_a^t \Phi(t,s)B(s)u(s)ds \qquad t \in [a,b]$$

where $\Phi(t,s)$ is the state-transition matrix for $\dot{x} = A(t)x$.

Proof: By Remark A.1, the elements of $A(t)$ are in $L_1(a,b)$. Now consider the column vector $c(t)$ defined by

$$c(t) = B(t)u(t) \to c_i(t) = \sum_{j=1}^{m} b_{ij}(t)u_j(t)$$

By Hölder's Inequality $b_{ij}(t) \in L_2(a,b)$ and $u_j(t) \in L_2(a,b)$ implies that $b_{ij}(t)u_j(t) \in L_1(a,b)$. Since $L_1(a,b)$ is a linear vector space

$$\sum_{j=1}^{m} b_{ij}(t)u_j(t) \in L_1(a,b).$$

Now Thm. A.2 can be applied with $b(t) = c(t)$ and the corollary follows.*

We shall now formulate the quadratic linear regulator problem in a Hilbert space setting. We will show that, in this formulation, the concept of observability arises naturally in defining an inner product. A global optimal control law will be derived without going through any variational arguments. In order to proceed we need the definition for the Cartesian product of vector spaces:

<u>Definition A.2</u>. Let X_1, X_2, ..., X_n be vector spaces over the same field of scalars, then the Cartesian product of these vector spaces is defined to be

$$X_1 x X_2 x \ldots x X_n = \{(x_1, x_2, \ldots, x_n) \colon x_k \in X_k,\ k = 1, \ldots, n\}$$

Addition and scalar multiplication are defined on this Cartesian product as follows

$$(x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n)$$

$$\alpha(x_1, x_2, \ldots, x_n) = (\alpha x_1, \ldots, \alpha x_n)$$

We shall denote the Cartesian product of a vector space with itself n-times by X^n.

<u>Definition A.3</u>. Let Q(t) be an nxn matrix and x(t) be an nx1 vector both of whose elements are real-valued measurable functions on (a,b). Q(t) is said to be positive definite on (a,b) if

$$x'(t)Q(t)x(t) > 0 \qquad \text{for all } x(t) \neq 0$$

$$x'(t)Q(t)x(t) = 0 \qquad \text{for } x(t) = 0$$

Similarly, Q(t) is positive semidefinite on (a,b) if

$$x'(t)Q(t)x(t) \geq 0 \qquad \text{for all } x(t).$$

Notation $Q(t) > 0$, $Q(t) \geq 0$.

Now consider equation A.3. We can interpret $x(t) = (x_1(t), \ldots, x_n(t))'$ with $x_k(t) \in L_2(a,b)$ for $k = 1, \ldots, n$; i.e., $x(t) \in L_2^n(a,b)$. Similarly, let $u(t) \in L_2^m(a,b)$. It is well known that for each $x_1(t)$, $x_2(t)$ in $L_2^n(a,b)$

$$<x_1(t), x_2(t)> = \int_a^b x_1'(t)C'(t)C(t)x_2(t)dt \tag{A.7}$$

defines an inner product on $L_2(a,b)$ whenever $C'(t)C(t)$ is a positive-definite nxn matrix. Of course, equality in the inner-product axioms [2] should now be interpreted as equality almost everywhere; i.e., we say $x_1(t)$ is equal to $x_2(t)$ whenever

$$x_1(t) = x_2(t) \qquad \text{a.e. on } [a,b]$$

Similarly, we can define an inner product on $L_2^m(a,b)$ by

$$<u_1(t), u_2(t)> = \int_a^b u_1'(t)D'(t)D(t)u_2(t)dt \tag{A.8}$$

whenever $D'(t)D(t)$ is an mxm positive-definite matrix on [a,b].

Remark A.2. All the axioms of inner-product except

$$<(x,u)(x,u)> = 0 \rightarrow (x,u) = 0$$

are satisfied whether or not $C'(t)C(t)$ and $D'(t)D(t)$ are positive definite.

Now let us consider the Cartesian product of L_2^n and L_2^m; i.e.,

$$L_2^n(a,b) x L_2^m(a,b) = \{x(t), u(t): x(t) \in L_2^n(a,b) \text{ and } u(t) \in L_2^m(a,b)\}$$

An inner product on $L_2^n(a,b) x L_2^m(a,b)$ can be defined by

$$<(x_1,u_1),(x_2,u_2)> = <x_1,x_2> + <u_1,u_2> \tag{A.9}$$

In the linear quadratic regulator problem we are concerned with the ordered pairs (x,u) in $L_2^n x L_2^m$ that satisfy the integral constraint.

$$V = \{(x,u) \varepsilon L_2^n x L_2^m : x(t) = \Phi(t,a)x(a) + \int_a^t \Phi(t,s)B(s)u(s)ds\} \tag{A.10}$$

where x(a) is some finite initial condition.

We will now investigate whether this extra dynamic information can be used to relax the positive-definiteness of C'(t)C(t) and D'(t)D(t) in the inner product or not. It turns out that the observability of the pair (C(t),A(t)) on [a,b] [5] and positive definiteness of D'(t)D(t) is a sufficient condition for equation A.9 to be an inner product on V. We state this fact in the following theorem.

Theorem A.3. Let V denote the set of ordered pairs (x(t),u(t)) defined by equation A.10. If the pair (C(t),A(t)) is observable on [a,b] and D'(t)D(t) is positive definite on [a,b], then the following bilinear form defined for every (x(t),u(t)) ε V by equation A.9.

$$\begin{aligned} <(x_1(t),u_1(t)),(x_2(t),u_2(t))> = \int_a^b & (x_1'(t)C'(t)C(t)x_2(t) \\ & + u_1'(t)D'(t)D(t)u_2(t))dt \end{aligned} \tag{A.11}$$

defines an inner product on V.

Proof: By Remark A.2 it is sufficient to show that if (C(t),A(t)) is observable and D'(t)D(t) is positive definite on [a,b], then $<(x(t),u(t)), (x(t),u(t))> = 0$ implies $(x(t),u(t)) = 0$. Let $(C(t),A(t))$ be observable and $D'(t)D(t)$ be positive definite on $[a,b]$.

$$<(x,u),(x,u)> = 0 \rightarrow \int_a^b x'(t)C'(t)C(t)x(t)dt = 0$$

and

$$\int_a^b u'(t)D'(t)D(t)u(t)dt = 0$$

Since $D'(t)D(t) > 0$ on $[a,b]$, we have $u(t) = 0$ on $[a,b]$. And with $u(t) = 0$, we have

$$\int_a^b x'(t)c'(t)C(t)x(t)dt = x'(a) \int_a^b \Phi'(t,a)C'(t)C(t)\Phi(t,a)dtx(a)$$

Define the observability Gramian $M(a,b)$ by

$$M(a,b) = \int_a^b \Phi'(t,a)C'(t)C(t)\Phi(t,a)dt$$

In general, $M(a,b)$ is a positive semidefinite matrix. Since $(C(t),A(t))$ is observable on $[a,b]$, $M(a,b)$ is positive definite [5] $\rightarrow x(a) = 0$. Since $x(a) = 0$, $u(t) = 0$ on $[a,b] \rightarrow x(t) = 0$ on $[a,b]$. It follows then $(x(t),u(t)) = 0$ on $[a,b]$.*

We shall now formulate the quadratic regulator problem and find the unique global optimal solution without variational arguments.

Lemma A.4. The set

$$M = \{(y(t),u(t)) \in L_2^r x L_2^m\colon y(t) = \int_a^t C(t)\Phi(t,s)B(s)u(s)ds$$

where $u(t) \in L_2^m(a,b)\}$, is a linear manifold.

Proof: Let (y_1,u_1) and (y_2,u_2) be in M. This implies that

$$y_1(t) = \int_a^t C(t)\Phi(t,s)B(s)u_1(s)ds$$

$$y_2(t) = \int_a^t C(t)\Phi(t,s)B(s)u_2(s)ds$$

where u_1, u_2 are in $L_2^m(a,b)$. For a_1, a_2 in R:

$$a_1y_1(t) + a_2y_2(t) = \int_a^t C(t)\Phi(t,s)B(s)(a_1u_1(s) + a_2u_2(s))ds$$

Since $a_1u_1 + a_2u_2 \in L_2^m(a,b) \rightarrow a_1(y_1,u_1) + a_2(y_2,u_2) \in M$, M is a linear manifold.*

In order to show that M is closed, we need to show that $\int_a^t \Phi(t,s)x(s)ds$ defines a continuous linear operator.

Lemma 2.6. The operator K: $L_2^n(a,b) \rightarrow L_2^n(a,b)$ defined by Kx = $\int_a^t \Phi(t,s)x(s)ds$ $x(t) \in L_2^n(a,b)$ where $\Phi(t,s)$ is the fundamental matrix defined by Definition A.1 is a bounded linear transformation on $L_2^n(a,b)$ and its adjoint K* is given by

$$K^*x^* = \int_t^b \Phi'(s,t)x^*(s)ds$$

(The inner product is the standard inner product on $L_2^n(a,b)$; i.e.,

$$\langle x_1x_2\rangle = \int_a^b x_1'(t)x_2(t)dt.)$$

Proof: Clearly, K is a linear operator.

$$\langle Kx,x^*\rangle = \int_a^b \left(\int_a^t \Phi(t,s)x(s)ds\right)'x^*(t)dt$$

$$= \int_a^b \int_a^t \sum_{i=1}^n \sum_{j=1}^n \phi_{ij}(t,s)x_j(s)x_i^*(t)ds\ dt$$

$$= \sum_{i=1}^n \sum_{j=1}^n \int_a^b \int_a^t \phi_{ij}(t,s)x_j(s)x_i^*(t)ds\ dt$$

Since $\phi_{ij}(t,s)$ is continuous on the closed and bounded rectangle [a,b] x [a,b], it is bounded. Since $-\infty < a < b < +\infty$,

$$\int_a^b |\phi_{ij}(t,s)|^2\ d(txs) < \infty$$

So $\int_a^t \phi_{ij}(t,s)x_j(s)ds$ is a Hilbert-Schmidt operator with kernel $\phi_{ij}(t,s)$ on $L_2(a,b)$. It is bounded and its adjoint is given by [2]

$$< \int_a^t \phi_{ij}(t,s)x_j(s)ds,\quad x_i^*(t)> = <x_j(t),\ \int_t^b \phi_{ij}(t,s)x_i^*(s)ds>$$

It follows then

$$\langle Kx,x^*\rangle = \sum_{i=1}^n \sum_{j=1}^n \int_a^b \int_t^b \phi_{ij}(s,t)x_j(t)x_i^*(s)ds\ dt$$

$$= \int_a^b x'(t) \int_t^b \Phi'(s,t)x^*(s)ds\ dt$$

$$= <x(t),\ \int_t^b \Phi'(s,t)x^*(s)ds\ dt$$

Defining $K^*x^* = \int_t^b \Phi'(s,t)x^*(s)ds$

$$\langle Kx, x^* \rangle = \langle x, K^*x^* \rangle \qquad x, x^* \in L_2^n(a,b)$$

By Theorem 5 in [6] on page 48, K is bounded and K* is, obviously, the adjoint of K.*

Lemma A.7. If the elements of $B(t)$ are bounded and measurable on (a,b), then $N = \{B(t)u(t): u(t) \in L_2^m(a,b)\}$ is a subspace of $L_2^n(a,b)$.

Proof: $B(t)u(t) = (\sum_{j=1}^{m} b_{ij}(t)u_j(t))$, $i = 1,\ldots,n$

$$|b_{ij}(t)u_j(t)|^2 \leq |b_{ij}(t)|^2 \; |u_j(t)|^2$$

Since $|b_{ij}(t)|^2$ is bounded and $|u_j(t)|^2$ is integrable on (a,b), it follows then their product is integrable. So

$$b_{ij}(t)u_j(t) \in L_2(a,b) \rightarrow \sum_{j=1}^{m} b_{ij}(t)u_j(t) \in L_2(a,b) \rightarrow B(t)u(t) \in L_2^n(a,b)$$

So N $L_2^n(a,b)$. N is a linear manifold; to show that it is closed we will show that B is a bounded linear transformation.

$$||B(t)u(t)||^2 = \int_a^b (B(t)u(t))'(B(t)u(t))dt$$

$$\leq \int_a^b tr(B'(t)B(t))u'(t)u(t)dt$$

$$\leq \int_a^b (tr(B'(t)B(t)))^2 dt ||u||$$

for all $u \in L_2^m(a,b)$ with $||u|| \leq 1$. It follows then that B(t) is a bounded linear transformation.

Now if $\{x_n\}_{n=1}^{\infty} \subset N$ is a fundamental sequence, since $L_2^n(a,b)$ is complete $\lim x_n$ exists and $\lim x_n \in L_2^n(a,b)$. Also since $x_n \in N$, $n = 1,2,\ldots$, there must exist $\{u_n\}_{n=1}^{\infty} \subset L_2^m(a,b)$ such that

$$x_n(t) = B(t)u_n(t) \qquad n = 1,2,\ldots$$
$$\lim x_n(t) = \lim B(t)u_n(t)$$

Since B(t) is a bounded linear transformation, it is also continuous so that

$$\lim B(t)u_n(t) = B(t) \lim u_n(t)$$

Since $\lim u_n(t) \in L_2^m(a,b) \rightarrow \lim x_n(t) \in N \rightarrow N$ is closed.*

<u>Lemma A.8</u>. If the elements of C(t) are bounded and measurable on (a,b), then $C(t) \int_a^t \Phi(t,s)B(s)u(s)ds$ where $u(t) \in L_2^m(a,b)$ is a continuous transformation.

<u>Proof</u>: By Lemma A.7, B(t) maps $L_2^m(a,b)$ into a subspace of $L_2^n(a,b)$. By Lemma A.6, $\int_a^t \Phi(t,s)(B(s)u(s))ds$ is a bounded linear transformation, and, therefore, continuous. Since the elements of C(t) are bounded and measurable, C(t) is a bounded linear transformation (see Lemma A.7). It is continuous; i.e.,

$$\lim C(t) \int_a^t \Phi(t,s)B(s)u_n(s)ds = C(t) \int_a^t \Phi(t,s)B(s)\lim_n u_n(s)ds$$

So $\int_a^t C(t)\Phi(t,s)B(s)u(s)ds$ is a continuous transformation.*

Lemma A.9. Consider $\{(y(t),u(t)\}\subset L_2^r x L_2^m$ defined by

$$M = \{(y(t),u(t)) \varepsilon L_2^r x L_2^m;\ y(t) = \int_a^t C(t)\Phi(t,s)B(s)u(s)ds$$

where

$$u(t) \varepsilon L_2^m(a,b)\}$$

M is a closed linear manifold under the norm induced by inner product on $L_2^r x L_2^m$:

$$<(y_1,u_1),(y_2,u_2)> = \int_a^b (y_1'(t)y_2(t) + u_1'(t)D'(t)D(t)u_2(t)dt \quad (A.12)$$

Proof: M is a linear manifold by Lemma A.4. Let $\{x_n,u_n\}_{n=1}^{\infty}\subset M$ be a fundamental sequence in M; i.e.,

$$\lim_{n,m\to\infty} ||((x_n-x_m),(u_n-u_m))||^2 = 0 \text{ which implies}$$

$$\lim_{n,m\to\infty} \int_a^b (u_n(t)-u_m(t))'D'(t)(u_n(t)-u_m(t))dt = 0 \quad \text{and}$$

$$\lim_{n,m\to\infty} \int_a^b (\int_a^t C(t)\Phi(t,s)B(s)(u_n(s)-u_m(s))ds)'(\int_a^t C(t)\Phi(t,s)B(s)(u_n(s)$$
$$- u_m(s))ds)dt = 0$$

So $\{u_n\}_{n=1}^{\infty}$ is a fundamental sequence in $L_2^m(a,b)$. Since L_2^m is complete, $\lim_n u_n$ exists and $\lim_n u_n \varepsilon L_2^m(a,b)$. Also

$$\{\int_a^t C(t)\Phi(t,s)B(s)u_n(s)ds\}_{n=1}^{\infty}$$

is a fundamental sequence in $L_2^r(a,b)$ with the standard inner-product. Since $L_2^r(a,b)$ is complete, $\lim_n \int_a^t C(t)\Phi(t,s)B(s)u_n(s)ds$ exists and is in $L_2^r(a,b)$. By Lemma A.8

$$\lim_n \int_a^t C(t)\Phi(t,s)B(s)u_n(s)ds = \int_a^t C(t)\Phi(t,s)B(s)\lim_n u_n(s)ds$$

It follows that

$$\lim_n \left(\int_a^t C(t)\Phi(t,s)B(s)u_n(s)ds,\ u_n(s)\right)$$

$$= \left(\int_a^t C(t)\Phi(t,s)B(s)\lim_n u_n(s),\ \lim_n u_n(s) \in M\right)$$

Thus, M is a closed linear manifold.*

<u>Theorem A.10</u>: Consider the optimization of the linear dynamical system $(A(t),B(t),C(t))$ governed by the equations

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) \qquad t \in [a,b]\ x(a) \text{ is a finite initial condition}$$

$$y(t) = C(t)x(t)$$

with respect to the performance index.

$$J = \left(\int_a^b (y'(t)y(t) + u'(t)R(t)u(t)dt\right)^{\frac{1}{2}} \tag{A.13}$$

If $R(t)$ is positive definite on $[a,b]$, then the unique global optimal control law minimizing equation A.13 is given by $u^*(t) = -R^{-1}(t)B'(t)K(t)x(t)$ where $K(t)$ satisfies the Riccati equation

$$\dot{K}(t) = -A'(t)K(t) - K(t)A(t) + K(t)B(t)R^{-1}(t)B'(t)K(t) - C'(t)C(t)$$

with $K(b) = 0$.

Proof: The problem is equivalent to finding the element which has the minimum norm in the linear variety V of $L_2^r(a,b) \times L_2^m(a,b)$

$$V = (C(t)\Phi(t,a)x(a),0) + M \quad \text{where M is given by}$$

$$M = \{(y(t),u(t) \in L_2^r x L_2^m: y(t) = \int_a^t C(t)\Phi(t,s)B(s)u(s)ds$$

$$\text{for } u(t) \in L_2^m(a,b)\}$$

The fact that V is a linear variety follows from Lemma A.9. Now by the Projection Theorem there exists a unique element (y^*,u^*) of V, which is of minimum norm. A necessary and sufficient condition for (y^*,u^*) be of minimum norm is that $(y^*,u^*) \perp M$ [4].

$$(y^*,u^*) \perp (y,u) \qquad (y,u) \in M$$

which implies

$$\int_a^b y^{*\prime}(t)y(t)dt + \int_a^b u^{*\prime}(t)R(t)u(t)dt = 0 \qquad (y,u) \in M$$

It follows then

$$\int_a^b y^{*\prime}(t) \int_a^t C(t)\Phi(t,s)B(s)u(s)ds + \int_a^b u^{*\prime}(t)R(t)u(t)dt = 0$$

$$u \in L_2^m(a,b)$$

$$<C'(t)y^*(t), \int_a^t \Phi(t,s)(B(s)u(s))ds > R(t)u^*(t), u(t)> = 0$$

$$u \in L_2^m(a,b)$$

Now by Lemma A.6

$$< \int_t^b \Phi'(s,t)C'(s)y^*(s)ds, B(t)u(t)> + <R(t)u^*(t), u(t)> = 0$$

$$u \in L_2^m(a,b)$$

After some manipulation

$$<B'(t) \int_t^b \Phi'(s,t)C'(s)y^*(s)ds + R(t)u^*(t), u(t)> = 0$$

$$u \in L_2^m(a,b)$$

It follows that

$$B'(t) \int_t^b \Phi'(s,t)C'(s)y^*(s)ds + R(t)u^*(t) = 0$$

Since $y^*(t) = C(t)x^*(t)$

$$u^*(t) = -R^{-1}(t)B'(t) \int_t^b \Phi'(s,t)C'(s)C(s)x^*(s)ds$$

Now we seek a closed form solution

$$\int_t^b \Phi'(s,t)C'(s)x^*(s)ds \stackrel{?}{=} K(t)x^*(t)$$

That is, does there exist a matrix $K(t)$ satisfying the above equality? Now $K(b) = 0$; differentiating the above equation we get a differential equation for $K(t)$:

$$-\Phi'(t,t)C'(t)C(t)x^*(t) - A'(t)\int_t^b \Phi'(s,t)C'(s)C(s)x^*(s)ds$$

$$= \dot{K}(t)x^*(t) + K(t)\dot{x}^*(t)$$

Since $\dot{x}^*(t) = (A(t)-B(t)R^{-1}(t)B'(t)K(t))\ x^*(t)$

$$\dot{K}(t)x^*(t) = (-A'(t)K(t)-K(t)A(t) + K(t)B(t)R^{-1}(t)B'(t)K(t)-C'(t)C(t))x^*(t)$$

which implies the matrix differential equation

$$\dot{K}(t) = -A(t)K(t) - K(t)A(t) + K(t)B(t)R^{-1}(t)B'(t)K(t) - C'(t)C(t) \quad \text{with } K(b) = 0$$

which is, of course, the familiar Riccati equation of linear optimal control. It can be shown that this equation has a unique, symmetric, positive semidefinite solution on [a,b].*

Appendix II

BASIC DEFINITIONS OF PROBABILITY THEORY

A set function is a function whose domain of definition is a class of sets. An extended real valued set function μ defined on a class of sets, ξ, is countably additive if, for every disjoint sequence $\{E_n\}_{n=1}^{\infty}$ of sets in ξ whose union is also in ξ, we have

$$\mu\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mu(E_n)$$

A σ-ring is a non-empty class of sets which is closed under the formation of differences and countable unions.

A measure is an extended real valued, non-negative, countably additive set function μ, defined on a σ-ring, such that $\mu(\emptyset) = 0$.

A measurable space is a set X and a σ-ring S of subsets of X with $\cup S = X$. A subset E of X is said to be measurable if and only if it belongs to the σ-ring S.

A measure space is a measurable space (X,S) and a measure defined μ on S. A real valued function f on a measurable space (X,S) is called a measurable function if $\{x \varepsilon X\colon f(s) < c\}$ is measurable for every real number c.

A σ-algebra of sets is defined to be a non-empty class of sets closed under the formation of complements and countable unions.

A probability space is a measure space (Ω,A,P) for which A is a σ-algebra of subsets of Ω and $P(\Omega) = 1$.

If ζ is any class of sets, the <u>σ-algebra generated by ζ</u>, is the smallest σ-algebra containing the sets of ζ.

If Ω_1 and Ω_2 are any two sets, the Cartesian product of the sets <u>Ω_1 and Ω_2</u>, in symbols $\Omega_1 \times \Omega_2$, is the set of all ordered pairs (ω_1, ω_2), where $\omega_1 \varepsilon \Omega_1$ and $\omega_2 \varepsilon \Omega_2$.

If Ω_1 and Ω_2 are any two sets with σ-algebras S_1 and S_2 of subsets of Ω_1 and Ω_2, respectively, $S_1 \times S_2$ is defined to be the σ-algebra generated by the class of sets of the form $A_1 \times A_2$ where $A_1 \varepsilon S_1$ and $A_2 \varepsilon S_2$.

A <u>transformation</u> is a function T defined for every point of a set Ω_1 and taking values in a set Ω_2. For every subset A_1 of Ω_1, the <u>image</u> of A_1 under T, in symbols $T(A_1)$, is the set of th points of Ω_2 which are of the form $T(\omega_1)$ for some $\omega_1 \varepsilon S_1$. For every subset A_2 of Ω_2, the <u>inverse image</u> of A_2 under T, in symbols $T^{-1}(A_2)$, is defined to be the set of all those points of Ω_1 whose image is in A_2.

If (X_1, S_1) and (X_2, S_2) are two measurable spaces, T is said to be a measurable transformation, in symbols $T\colon (X_1, S_1) \rightarrow (X_2, S_2)$, if the inverse image of every measurable set is measurable.

REFERENCES

1. Halmos, P. R., "Measure Theory," D. Van Nostrand Company, Inc., Princeton, N. J., 1950.

2. Helmberg, G., "Introduction to Spectral Theory in Hilbert Space," North-Holland Publishing Company, Amsterdam, 1969.

3. Coddington, E. S. and Levinson, N., "Theory of Ordinary Differential Equations," McGraw-Hill Book Company, New York, 1955.

4. Luenberger, D. G., "Optimization by Vector Space Methods," John Wiley and Sons, Inc., New York, 1969.

5. Kalman, R. E., "Canonical Structure of Linear Dynamical Systems," Proc. National Acad. Sci. U. S., Vol. 48, pp. 596-600, 1962.

6. Akhiezer, N. I. and Glazman, I. M., "Theory of Linear Operators in Hilbert Space," Frederick Ungar Publishing Co., New York, 1961.

7. Loeve, M., "Probability Theory," D. Van Nostrand Company, Inc., Princeton, N. J., 1963.

8. Doob, J. L., "Stochastic Processes," John Wiley and Sons, Inc., New York, 1953.

9. Bucy, R. S. and Joseph, P. D., "Filtering for Stochastic Processes with Applications to Guidance," Interscience Publishers, 1968.

10. Wong, L., "Stochastic Processes in Information and Dynamical Systems," McGraw-Hill Book Company, 1971.

11. Ferguson, T. S., "Mathematical Statistics: A Decision Theoretic Approach," Academic Press, New York, 1967.

12. Weber, C. L., "Elements of Detection and Signal Design," McGraw-Hill Book Company, New York, 1968.

13. Kalman, R. E. and Bucy, R. S., "New Results in Linear Filtering and Prediction Theory," J. Basic Eng. ASME, Ser. D., pp. 95-108, 1961.

14. Kailath, T., "A Note on Least Squares Estimation by the Innovations Method," SIAM J. Control, Vol. 10, No. 3, August 1972.

15. Kailath, T., "A General Likelihood Ratio Formula for Random Signals in Gaussian Noise," IEEE Trans. Inform. Theory, Vol. IT-15, pp. 350-361, May 1969.

16. Girsanov, I. V., "On Transforming a Certain Class of Stochastic Processes by Absolutely Continuous Substitution of Measures," Theor. Probability Appl., Vol. 5, pp. 285-301, 1960.

17. Duncan, T. E., "Evaluation of Likelihood Functions," Inform. Contr., Vol. 13, pp. 62-74, 1968.

18. Duncan, T. E., "Likelihood Functions for Stochastic Signals in White Noise," Vol. 16, pp. 303-310, 1970.

19. McKean, H. P., "Stochastic Integrals," Academic Press, New York, 1969.

20. Balakrishnan, A. V., "Stochastic Differential Systems," Springer Verlag, New York, 1973.

21. Jazwinski, A. H., "Stochastic Processes and Filtering Theory," Academic Press, New York, 1970.

22. Wong, E. and Zakai, M., "On the Relation Between Ordinary and Stochastic Differential Equations," Int. J. Eng. Sci., Vol. 3, pp. 213-229, 1965.

23. Wong, E. and Zakai, M., "On the Convergence of Ordinary Integrals to Stochastic Integrals," Annals of Math. Stat., Vol. 36, pp. 1560-1564, 1965.

24. Magill, D. T., "Optimal Adaptive Estimation of Sampled Stochastic Processes," IEEE Trans. Automat. Control, Vol. AC-10, pp. 434-439, October 1965.

25. Kalman, R. E., "A New Approach to Linear Filtering and Prediction Problems," ASME, J. Basic Eng., 82, pp. 35-45, 1960.

26. Sims, F. L. and Lainiotis, "Recursive Algorithm for the Calculation of the Adaptive Kalman Filter Weighting Coefficients," IEEE Trans. Automat. Control, Vol. AC-14, pp. 215-218, April 1969.

27. Lainiotis, D. G., "Sequential Structure and Parameter Adaptive Pattern Recognition - Part I: Supervised Learning," Vol. IT-16, pp. 548-556, September 1970. IEEE Trans. Inform. Theory.

28. Lainiotis, D. G., "Optimal Adaptive Estimation Structure and Parameter Adaptation," IEEE Trans. on Automat. Control, Vol. AC-16, pp. 160-170, April 1972.

29. Kushner, H. J., "Dynamical Equations for Optimal Nonlinear Filtering," Journal of Differential Equations, Vol. 3, pp. 179-190, 1967.

30. Ash, R. B., "Real Analysis and Probability," Academic Press, New York, 1972.

31. Duncan, T. E., "On the Absolute Continuity of Measures," Annals. Math. Statist., Vol. 41, pp. 30-38, 1970.

32. Shepp, L. A., "Radon-Nikodym Derivatives of Gaussian Processes," Annals. Math. Statist., Vol. 37, pp. 321-354, 1966.

33. Kailath, T., "A Further Note on a General Likelihood Formula for Random Signals in Gaussian Noise," IEEE Trans. Inf. Theory, Vol. IT-16, July 1970.

34. Caglayan, K. A., and Montgomery, R. C., "A Self Reorganizing Digital Flight Control System for Aircraft," Presented at AIAA 12th Aerospace Sciences Meeting, Paper No. 74-21, Washington, D. C.

35. Kakutani, S., "On Equivalence of Infinite Product Measures," Ann. Math., Vol. 47, pp. 214-224, 1948.

36. Sosulin, Y., and Stratonovich, "Optimum Detection of a Diffusion Process in White Noise," Radio Eng. Electron. Phys. (USSR), Vol. 10, pp. 704-714, 1965.

37. Grenander, U., "Stochastic Processes and Statistical Inference," Arkio fur Mathematik, Vol. 1, pp. 195-277, 1950.

38. Kadota, T. T., "Optimum Reception of Gaussian Signals," Bell Systems Tech. J., Vol. 43, pp. 2767, 2810, 1964.

39. Kadota, T. T., "Optimum Reception of Gaussian Signals," Bell Systems Tech. J., Vol. 44, pp. 1621-1658, 1965.

40. Park, S. K. and Lainiotis, D. G., "Joint Detection-Estimation of Gaussian Signals in White Gaussian Noise," Inf. Sciences, Vol. 4, pp. 315-325, 1972.

41. Middleton, D. and Esposito, R., "Simultaneous Optimum Detection and Estimation of Signals in Noise," IEEE Trans. Inf. Theory, Vol. IT-14, pp. 434-444, 1968.

42. Nahi, N. E., "Optimal Recursive Estimation with Uncertain Observation," IEEE Trans. Inf. Theory, Vol. IT-15, pp. 457-462, 1969.

43. Stein, G. and Saridis, G. N., "A Parameter-Adaptive Control Technique," Automatica, Vol. 5, pp. 731-739, 1969.

44. Saridis, G. N. and Dao, T. K., "A Learning Approach to the Parameter Adaptive Self-Organizing Control Problem," Automatica, Vol. 8, pp. 589-597, 1972.

45. Deshpande, J. G., Upadhyay, T. N. and Lainiotis, D. G., "Adaptive Control of Linear Stochastic Systems," Automatica, Vol. 9, pp. 107-115, 1973.

46. Wonham, W. M., "Some Applications of Stochastic Differential Equations to Optimal Nonlinear Filtering," J. SIAM Control, Vol. 12, pp. 347-369, 1965.

47. Sworder, D. D., "Feedback Control of a Class of Linear Systems with Jump Parameters," IEEE Trans. on Automat. Control, Vol. AC-14, pp. 9-14, 1969.

The vita has been removed from the scanned document

OPTIMAL PARAMETER ADAPTIVE ESTIMATION OF STOCHASTIC PROCESSES

by

Alper K. Caglayan

(ABSTRACT)

This study is concerned with the simultaneous detection and least squares estimation of vector random processes. The problem is formulated in the following context: A random process, out of a countably infinite collection of (not necessarily Gaussian) vector random processes with known distributions, is observed with additive white Gaussian noise. The *a priori* probability, that a specific random process will be observed, is specified for each one in the collection. The least squares estimate of the random process that is being observed is to be found in terms of the hypothesis conditioned estimates.

It is shown that the best estimate is the linear combination of the hypothesis conditioned estimates weighted by the *a posteriori* probabilities of the hypotheses conditioned on the observations. A Radon-Nikodym derivative representation is derived for the *a posteriori* probability by using the specific structure of the product probability measure for this problem. It is shown that this Radon-Nikodym derivative can be expressed in terms of the Radon-Nikodym derivatives of measures induced by the random processes in the collection with respect to Wiener measure. By using the recent results on likelihood functions, an expression for the *a posteriori* probability is found in terms of the conditioned estimates. In this connection, an extended version of the partition theorem of parameter adaptive estimation

is proved. The unique stochastic differential equation, that each *a posteriori* probability satisfies with its associated *a priori* probability as the initial condition, is derived for the case of finitely many hypotheses along with an expression for the conditional error covariance in terms of the hypothesis conditioned error covariances.

The results are applied to the parameter adaptive estimation problem in linear continuous and discrete stochastic dynamic systems. In the continuous case, the solution is also obtained through an alternate approach using nonlinear filtering theory. An application of the theory to the design of a digital flight control system which is tolerant of sensor failures is presented with real-time hybrid computer simulation results. A review of random processes and statistical decision theory is also included.