ASYMPTOTIC CONVERGENCE OF NON-LINEAR, CONTINUOUS-TIME FILTERS

Thesis for the Degree of Ph. D.
MICHIGAN STATE UNIVERSITY
MICHAEL WESLEY BIRD
1969

LIBRARY
Michigan State
University

This is to certify that the

thesis entitled

ASYMPTOTIC CONVERGENCE OF
NON-LINEAR, CONTINUOUS-TIME FILTERS

presented by

Michael Wesley Bird

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in E.E.

[signature]
Major professor

Date [illegible]

O-169

BINDING BY
HOAG & SONS'
BOOK BINDERY INC.
LIBRARY BINDERS

ABSTRACT

ASYMPTOTIC CONVERGENCE OF
NON-LINEAR, CONTINUOUS-TIME FILTERS

by

Michael Wesley Bird

Modern non-linear, continuous-time filtering theory, characterized by the use of state variable techniques, is potentially applicable to many real-world problems. However, the filters suggested by the theory as solutions to the time-domain filtering problem have not been sufficiently analyzed to provide confidence in their behavior. One aspect of non-linear filter performance which has received no attention is asymptotic behavior. This thesis suggests a method for demonstrating the asymptotic performance of non-linear filters and applies this method to a class of continuous-time, non-linear filters in the scalar case.

For this investigation, the time-domain filtering problem consists of a message process, represented as the solution to a first-order differential equation with unknown initial condition, and an observation process, modeled as a signal containing the message to which white noise is added. Filters considered in this thesis are sequential, being represented by first-order differential equations which are identical to the model of the message process plus a correction term. The correction term consists of a gain function, which depends on the past of the filter output, multiplied by the difference between an estimate of the observation and the

observation itself. This structure is similar to the filter made popular by Kalman and Bucy for a related linear filtering problem. The non-linear, continuous-time filters considered here are not postulated to have any optimal properties; however, when the message process has a probability distribution, the gain function can be selected to make the filter approximately-optimal, providing nearly-minimum-variance estimates of the message.

In the discrete-time case, the stochastic approximation methods developed for parameter estimation led to asymptotic convergence theorems for sequential filters. These theorems show that the difference between the output of certain filters and the message process converges to zero, with probability one and in the mean square sense, as time increases. This thesis applies the same strategy to the continuous-time problem.

Stochastic approximation methods are developed for the continuous-time parameter estimation problem. The procedure developed uses sequential algorithms for estimating a parameter of a signal when the signal is observed in the presence of white Gaussian noise. Two theorems are proved which show that, when the signal models and the gain functions in theestimators satisfy certain conditions, the estimators converge asymptotically to the true parameter value. The proofs of these theorems rely on the properties of Ito calculus and super-martingales.

Relying on the concepts developed for continuous-time parameter estimation, two asymptotic convergence theorems are proved for a solution to the filtering problem. In these theorems, conditions are placed on the message process, observation process, and

gain functions in the filters that guarantee asymptotic convergence. For purposes of evaluation, these results are compared to a theorem which specifies the asymptotic behavior of the Kalman-Bucy filter which is a solution to a linear filtering problem.

The convergence theorems developed for filtering show the asymptotic convergence of a particular sequential filter which is an approximately-optimal filter. A filter is thus displayed whose output, under certain conditions, provides nearly-minimum-variance estimates of the message throughout the observation time interval and converges to the message process as time increases. Computer-simulated results show the behavior of this filter when applied to a non-linear filtering example.

ASYMPTOTIC CONVERGENCE OF NON-LINEAR, CONTINUOUS-TIME FILTERS

By

Michael Wesley Bird

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Electrical Engineering

1969

G 60135
1-21-70

ACKNOWLEDGEMENTS

For consistent confidence in the preliminary work and guidance in the actual writing of this thesis, the author wishes to thank Professor Richard Dubes, his teacher and advisor. His counsel and understanding are greatly appreciated.

For their consideration of this thesis, the author wishes to acknowledge the other members of his doctoral committee: H. G. Hedges, C. L. Park, R. B. Zemach, and H. Salehi.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

In the process of transmitting or gathering information, the signal that contains the information frequently becomes distorted. It is usually necessary to modify this signal in order to remove the distortion and recover the original message. This act of modifying the signal is called filtering. In other words, the essential purpose of a filter is to remove the distortion or noise from an observed signal.

The design of a filter requires mathematical models that represent the signal and observation processes. Wiener [1] was the first to treat these processes as random phenomena and described them in statistical terms. His treatment allows filters to be designed from criteria based on the statistics of these processes; he points out that his method of developing filters combines the techniques of random time-series and conventional electrical filter theory. More recently, Kushner [2], Bucy [3], Kalman [4], and Deutsch [5], still using the random descriptions, have developed filters based on fundamental statistical methods other than those of Wiener. The approaches of Wiener and Kalman are the most popular methods for designing filters.

The investigations in this thesis are based on the filtering formulation proposed by Kalman [4] and Kushner [2]. Section 1.1 outlines this formulation and discusses extensively criteria for designing filters. A survey of many solutions to the filtering problem is given in Sec. 1.2 pointing out some limitations of the filtering results presently available and indicating the need for studies into the asymptotic behavior of statistical filters. In Sec. 1.3, the objectives of the thesis are discussed and a brief summary of the main results of the thesis is given in Sec. 1.4.

1.1 Performance Criteria for Solutions of the Modern Filtering Problem

During the last eight years, a special form of the statistical filtering problem has received a vast amount of attention in engieering literature. This form of the problem uses modern modeling techniques and is distinguished from the traditional by its formulation, the message process being described by a stochastic differential equation. With $\underline{x}$ denoting the n-dimensional message process, or signal process, the message representation is:

$$\dot{\underline{x}} = \underline{f}(t,\underline{x}) + \underline{Q}(t)\underline{u}(t) \ ; \qquad \underline{x}(t_o) = \underline{c} \tag{1.1}$$

where $\underline{f}(t,\underline{x})$ is a set of n known functions of the n components of $\underline{x}$ and t, $\underline{Q}(t)$ is a known $n \times n$ matrix, $\underline{u}(t)$ is an n-dimensional, zero-mean, Gaussian, white noise process,[1] and $\underline{c}$ is an unknown constant vector. Equation (1.1) represents a Markov process and completely determines the probability distribution of $x(t)$. This distribution determines all moments and correlation properties of the process.

---

[1] An r-dimensional white noise process, say $\underline{w}(t)$, is a process with the property that $Ew_i(t)w_j(s) = \delta(t-s)\delta_{ij}$ for $1 \le i,j, \le r$, where δ_{ij} is the Kronecker delta and $\delta(t-s)$ is the generalized delta function.

The remainder of the filtering formulation defines the m-dimensional observation process, $\underline{y}(t)$, which is assumed to have the conventional model:

$$\underline{y}(t) = \underline{h}(t,\underline{x}(t)) + \underline{R}(t)^{\frac{1}{2}}\underline{v}(t) \tag{1.2}$$

where $\underline{h}(t,\underline{x})$ is a set of m known functions of the n vector $\underline{x}$ and t, $\underline{R}(t)^{\frac{1}{2}}$ is the square root of the positive definite $m \times m$ matrix $\underline{R}(t)$, and $\underline{v}(t)$ is an m-dimensional, zero-mean, Gaussian, white-noise process.

The problem is to utilize this formulation to develop useful filters. At this time, no restrictions will be placed on the form of the filter, other than that its input must be the finite past of the observation process and its output must approximate the message process $x(t)$. The type of approximation provided by the output depends on the performance criteria used in selecting the filtering solution. The performance criteria or, stated in different words, the requirements for a good solution, are selected according to the requirements of each particular filtering application. A few such applications and standards for judging the quality of a filter will be discussed to enumerate features of useful filters.

The communication model, for example, has recently been placed into the framework of the modern filtering formulation described above [16]. The solutions to the filtering problem are used as receivers at the output of noisy channels and are intended to demodulate the message which has been contaminated by channel noise. This modern approach to filtering is applicable to the communication problem because the resulting filters can give reasonably accurate estimates during the transient phase of the filter's response. The transient response has long been a difficulty with classical Wiener filters since they are designed from steady-state error criteria [1]. However, even though the modern filter has this advantage, its performance as a communication receiver is still judged on the quality of its steady-state behavior.

A second example, the aerospace guidance problem, has been one of the most studied applications of filtering theory [7]. The time-varying parameters describing the position of a space vehicle are designated as the message. Noisy measurements of these parameters are filtered to give estimates of the parameters from which the vehicle's location can be established. The filter must be analyzed in terms of three design criteria: 1) the asymptotic behavior of the error variance; 2) the amount of time required to converge into the asymptotic state; 3) the computational feasibility of the algorithm used in synthesizing the filter.

Since guidance systems are operated on a real-time basis for extended periods, the third criteria becomes very important in the implementation of the filter. Intuitively, it appears that the first two criteria can be satisfied if the filter is designed to give an accurate estimate of the message during the transient period, in which case both the asymptotic error variance and the length of the transient period should be reduced. However, excessively stringent accuracy requirements increase the complexity of the filter, so a trade-off may be necessary to satisfy all three criteria.

A final example, system identification, may be placed in the filtering formulation [8], [9]. The typical identification problem demands a filter which produces estimates of system parameters from noisy measurements of the system's input and output. The unknown parameters are either constants or varyslowly with time. For any reasonable identification procedure, the filter's output must converge in some probabilistic sense to the true parameter as time increases. A secondary, but still important, consideration is the amount of time required for convergence. The convergence time may be minimized if the estimates are as accurate as possible during the entire time interval. The identification scheme is frequently a small segment of a very large and complicated system so constraints must be placed on the complexity of the algorithm. Consequently, a trade-off between complexity and accuracy may be necessary.

These three applications demonstrate the type of real-world problems which may be formulated using the time-domain representations defined in (1.1) and (1.2). Since the filter characteristics cited in these examples are typical of real problems, the following three requirements for a good and useful filter are proposed.

R.1 The output of a good filter should accurately estimate the message throughout the observation interval. For the filtering problem in this thesis, there are a few statistically-based methods which provide a filter satisfying this requirement. The most commonly used method selects the filter which minimizes the mean-square difference between the message process and filter output for each time instant in the observation interval. The filter minimizing this performance index is represented by the conditional mean functional. This procedure is widely used and, since the filter is a solution to an optimization problem, it is commonly called the optimal filter.

R.2 Analytical results should be available that describe the asymptotic behavior of a good filter as $t \to \infty$. Examining the solution to time-domain filtering problems can bring out the type of analysis needed. The earlier discussion indicates that the filter has two phases of operation, the transient-response phase and a type of asymptotic phase. For a wide variety of real problems, this second phase will determine the filter's usefulness. For these problems, a good transient response is, of course, ideal, but not at the expense of asymptotic behavior. Any filtering algorithm that guarantees the filter's large-time behavior will place restrictions on the message and observation models. These restrictions assure that the fluctuations of the estimate of the message will be confined by a calculable bound.

R.3 A useful filter must not be too complex, so a sequential algorithm is required. An algorithm is sequential when the filter's output is defined by the solution to a finite-dimensional system of first order differential equations, where the derivatives are functions of the present value of the observation process, present value of the system solution and, possibly, a specified functional of the past solution. This might seem to be an unreasonable limitation for an optimal filter. However, the time-domain approach to filtering yields optimal or nearly-optimal solutions which are sequential. In fact, many present-day problems are being formulated in terms of this modern approach in order to obtain sequential algorithms with nearly-optimal properties. Optimal sequential filters, plus the computer technology for efficiently implementing them, are providing useful solutions to many previously-unmanageable problems.

Any filter should attempt, in some sense, to satisfy these three criteria. However, the approaches to filtering in the literature are concentrated on deriving approximately-optimal filters, with little or no study of their asymptotic behavior. Analytical investigations of asymptotic behavior are very difficult and appear to be a major barrier in the utilization of useful filters. Considering the difficulties inherent in the analysis of the optimal filter, it is logical to study other filters, without stringent specifications on transient response, but which are more amenable to mathematical treatment than the optimal algorithm. These algorithms will be sequential and, along with well-analyzed asymptotic behavior, attempt to satisfy the requirements enumerated in R.2 and R.3. This thesis will be primarily concerned with initiating an investigation into this type of filter.

1.2 Literature Review

To better understand the state of filtering theory and to further reinforce the previous discussions, a summary of existing results is presented in this section. This synopsis will not only include available solutions to the filtering problem discussed in Sec. 1.1, but will also summarize the publications on the discrete-time filtering problem.

At this point, it is necessary to make the distinction between continuous-time and discrete-time processes. Equations (1.1) and (1.2) define waveforms specified on the entire time axis, called continuous-time processes. In contrast are the discrete-time processes which are sequences of variables defined on only a countable number of time instants.

The following continuous-time investigations are surveyed in Secs. 1.2.1, 1.2.2, and 1.2.3: the well-established linear theory, the optimal approach to non-linear filtering, and the completely unexplored realm of asymptotically stable, sequential filters. A review of the optimal and sub-optimal approaches to discrete-time filtering are contained in Secs. 1.2.4 through 1.2.7. All these discussions place a heavy emphasis on reporting studies, or the need for studies, into the asymptotic behavior of filters. The discrete-time survey will suggest exploring the method of stochastic approximation as an approach to investigating the large-time performance of continuous-time filtering solutions, so this section will conclude with an outline of the few publications on continuous-time stochastic approximation.

1.2.1 Linear Filtering Theory - the Kalman-Bucy Filter

Assuming $\underline{f}(t,\underline{x}) = \underline{F}(t)\underline{x}$ and $\underline{h}(t,\underline{x}) = \underline{H}(t)\underline{x}$, the filtering formulation defined in (1.1) and (1.2) becomes linear. It is also assumed that $\underline{c}$ is Gaussian with zero mean and variance $\underline{\Gamma}$. Kalman and Bucy orginated this formulation in 1961, and developed a very effective and useful solution to the filtering problem [4]. They derived an optimal filter which produced the conditional mean, the minimum-variance estimator. Their solution has been shown to have the following properties:

A. The filter's output is determined from the solution to two differential equations. Denoting $\underline{x}(t)$ as the n-dimensional output at time t, the equations are written as:

$$\dot{\hat{\underline{x}}} = \underline{F}(t)\hat{\underline{x}} + \underline{P}(t)\underline{H}(t)^T[\underline{y}(t) - \underline{H}(t)\hat{\underline{x}}] \; : \; \hat{\underline{x}}(t_o) = 0 \qquad (1.3)$$

$$\dot{\underline{P}} = \underline{F}(t)\underline{P} + \underline{PF}(t)^T - \underline{PH}(t)^T\underline{R}(t)^{-1}\underline{H}(t)\underline{P} + \underline{QQ}^T \; : \; \underline{P}(t_o) = \underline{\Gamma} \qquad (1.4)$$

This optimal algorithm is sequential.

B. The solution to (1.4), commonly called a Riccati equation, is the error variance, since $\underline{P}(t) = E(\hat{\underline{x}}(t) - \underline{x}(t))(\hat{\underline{x}}(t) - \underline{x}(t))^T$. Equation (1.4) does not depend on observations and can be solved prior to actual implementation of the filter.

C. Kalman and Bucy formulated observability and controllability conditions which insure adequate asymptotic behavior of this optimal filter [10].

The optimal solution (1.3), commonly called the Kalman-Bucy filter, satisfies R.1, R.2, and R.3, the postulates for useful filters. This optimal, sequential, asymptotically stable filter found immediate use in the fields of guidance and control. More recently, its popularity has spread to the fields of system identification, pattern recognition, and differential encoding of television signals [11].

1.2.2 The Optimal Non-linear Filter Solution

Significant work on optimal non-linear filtering solutions began in 1964, with important contributions from Kushner [2], Bucy [3], Stratonovich [12], and Kallianpur [13]. Their primary aim was establishing the conditional mean functional and, because of the non-linear nature of the problem, their derivations required a rigorous treatment of detail. Initially, the mathematical model for the Gaussian white noise disturbance was interpreted as the formal derivative of an independent-increment process, Brownian motion, which requires an understanding of Ito's stochastic differential equations [3]. Next, the message process, observation process, and conditional mean process were represented as elements in an abstract functional space. Finally, a stochastic differential equation was derived for the conditional mean using Ito's special differentiation rule.

The stochastic differential equation representing the solution for the conditional mean appears initially to be neat and concise. However, careful examination shows that the optimal filter is described by an infinite-dimensional system of first-order stochastic differential equations, each with the observation process as a driving term. Furthermore, the entire system requires a simultaneous solution. These two facts make the optimal filter impossible to realize.

The present emphasis in optimal, non-linear filtering research is centered upon deriving algorithms that approximate the optimal filter. The standard approach is to represent the non-linear functions in (1.1) and (1.2) as series expansions and to retain the first few terms. The infinite system of differential equations describing the optimal filter then reduces to a finite set and a realizable algorithm is possible. The approximations made on the non-linear functions restrict the applications of the filter so a filtering solution must be developed for each class of problems. Kushner [14] discusses the shortcomings of a number of such approximation schemes.

Almost all approximately-optimal filters are sequential. An examination of the system of stochastic differential equations defining the optimal filter shows that most finite approximations will be sequential. The demand for a filter with a sequential structure instigated the now-prodigious amount of research on the time-domain approach.

While many investigators have been successful in deriving approximately-optimal filters, analytical studies of the error characteristics, or asymptotic behavior, for such filters are not available. All work devoted to the properties of such filters involve lengthy computer simulations for very specific problems. A recent survey paper has pointed out both this void in the literature and the reluctance of users to apply the present non-linear filter theory [15].

1.2.3 Sub-Optimal Sequential Filters with Known Asymptotic Behavior

Without general analysis of the behavior of optimal, non-linear filters, the real-world user does not have confidence in their behavior. However, the demand for sequential, non-linear filters exhibits a need for sub-optimal schemes whose performances are understood and analyzed. As Sec. 1.1 indicated, there are numerous applications where large strings of data must be processed using a minimal amount of equipment; these applications require filters with little storage and good large-time behavior. The present state of optimal filtering theory cannot supply algorithms with satisfactory properties for such tasks.

1.2.4 The Discrete-Time Filtering Problem

Presently, there are no published investigations of the asymptotic performance of continuous-time filters. Insight into this area may be gained from a review of the discrete-time filtering results. A brief outline of the discrete-time filtering approaches is presented next, emphasizing investigations into asymptotic behavior.

The structure of the discrete-time filtering problem is similar to that of the continuous-time problem in Sec. 1.1. The n-dimensional message process is:

$$\underline{x}_{k+1} = \underline{f}_k(\underline{x}_k) + \underline{Q}_k\underline{u}_k, \qquad k = 0,1,2,\ldots \tag{1.5}$$

where $\{\underline{u}_k\}$ is an r-dimensional, zero-mean noise sequence with $E\, u_k^i u_m^j = \delta_{km}\delta_{ij}$, $1 \le i,j \le r$, $\underline{f}_k(\underline{x}_k)$ is a known n-vector of functions, and $\underline{Q}_n$ is an $n \times n$ matrix. The m-dimensional observation equation is

$$\underline{y}_k = \underline{h}_k(\underline{x}_k) + \underline{R}_k^{\frac{1}{2}}\underline{v}_k \tag{1.6}$$

where $\{\underline{v}_k\}$ is an m-dimensional, zero-mean noise sequence, $E\, v_\ell^i v_k^j = \delta_{\ell k}\delta_{ij}$, $1 \le i,j \le m$, $E\, v_\ell^i u_k^j = 0$, and $\underline{R}_k^{\frac{1}{2}}$ is the square root of a positive definite matrix, $\underline{R}_k$.

The filtering problem is to determine a function of the past observations which will furnish a good estimate of the present message value. The function transforming the observed data into a sequence of estimates represents a filter.

The discrete-time filtering problem has received a great deal more attention than its continuous-time counterpart. This may be attributed to the availability of results in classical statistics which concern random phenomenon described by sequences of random variables and which have been applied directly to the discrete-time problem. Investigations into the filtering problem are split into two groups. When the noises in (1.5) and (1.6) are Gaussian, the

Bayes-optimal approach concentrates on finding a function representing a conditional mean. The other group of investigations are studies of asymptotically-stable, recursive, but non-optimal algorithms. An algorithm is recursive when its output is the solution to a difference equation with the (n+1)st output depending on the nth output and nth or (n+1)st observation.

1.2.5 Optimum Discrete-Time Filtering Theory

Many investigators have suggested a representation for the conditional mean of the message, given the observations, as a solution to the discrete-time filtering problem [16], [17], [18], [19]. This optimal solution is only employed when the noises $\{\underline{u}_k\}$ and $\{\underline{v}_k\}$ are Gaussian. As could be expected from the non-linear nature of the problem, a closed-form solution for the general optimal filter does not exist. The research effort has been channeled into derivations of approximate algorithms. Aoki [16] discussed the following recursive algorithm for an approximately-optimal filter solution:

$$\underline{\hat{x}}_{n+1} = \underline{f}_n(\underline{\hat{x}}_n) + \underline{K}_{n+1}[\underline{y}_{n+1} - \underline{h}_n(\underline{f}_n(\hat{x}_n))] \tag{1.7}$$

where $\underline{\hat{x}}_n$ is the filter's output and $\{\underline{K}_n\}$, the gain sequence, is determined from a set of recursive equations. This filter has had numerous applications and is included here for reference.

The development of discrete-time, non-linear filters is hindered by the lack of asymptotic (large n) results. Albert and Gardner [20] and Pearson [21] both emphasize the need for work in this area and indicate the necessity of studying convergent, recursive filtering schemes. These authors rely heavily on the methods of stochastic approximation as means for generating sub-optimum filtering algorithms as outlined in the next section.

1.2.6 The Method of Stochastic Approximation in Parameter Estimation

Stochastic approximation algorithms are used for estimating parameters of signals which are observed in the presence of noise. Albert and Gardner [20] have applied the method of stochastic approximation to problems where the observations have the form

$$y_n = h_n(\theta) + v_n \qquad (1.8)$$

where θ is an unknown parameter, $\{v_n\}$ is a zero-mean process with $E\, v_k v_n = \delta_{nk}$, and $h_n(\theta)$ is a known function of n and θ. The noise distribution need not be known. If θ is considered a message, the problem of estimating θ can be considered a special case of the filtering problem. The scalar case is treated here to simplify the notation.

In parameter estimation, as in filtering theory, the problem is to process the observation sequence and estimate the value of the unknown constant. The method of stochastic approximation attempts to find a zero of a regression function and results in the following recursive algorithm which solve the estimation problem.

$$\bar{x}_{n+1} = \bar{x}_n + a_n[y_n - h_n(\bar{x}_n)] \; : \qquad \bar{x}_o \text{ is arbitrary} \qquad (1.9)$$

where $\bar{x}_n$ is the nth estimate of θ.

The next step is the formulation of conditions which guarantee that $\bar{x}_n - \theta$ converges to zero as $n \to \infty$ with probability one or in the mean-square sense. These conditions are illustrated in the statement of the following stochastic approximation convergence theorem [20]. The hypotheses are:

T1. $h_n(x)$ is assumed monotone in x and differentiable;

T2. For each n, the sign of a_n is equal to the sign of $\dot{h}_n(x)$ where $\dot{h}_n(x) = \frac{\partial}{\partial x} h_n(x)$;

T3. $\sum_{n=1}^{\infty} b_n a_n = \infty$ with $b_n = \inf_x |\dot{h}_n(x)|$;

T4. $\sum_{n=1}^{\infty} a_n^2 < \infty$.

Then $x_n \to \theta$ as $n \to \infty$ with probability one and in mean-square.

Stochastic approximation algorithms are traditionally associated with parameter estimation problems where noise models are unknown, large amounts of data are processed, and simple computations are required.

1.2.7 Recursive Filters

Generalizing on the method of stochastic approximation, Pearson [21] and Albert and Gardner have suggested the recursive algorithms

$$\bar{x}_{n+1} = f_n(\bar{x}_n) + a_n[y_n - h_n(\bar{x}_n)] \tag{1.10}$$

with $\bar{x}_n$ the filter output as solutions to the general filtering problem defined in Sec. 1.2.4 with $m = n = r = 1$. Equation (1.10) is similar to the approximately-optimal filter in (1.7). These authors reported on the results of initial investigations into the application of stochastic approximation methods to the determination of recursive filters with desirable asymptotic characteristics.

Pearson considered a special case of the problem, assuming $u_n = 0$ and $h_n(x) = hx$. Then, with the hypotheses $|f(x) - f(y)| \le |x - y|$, $\sum_{n=1}^{\infty} a_n = \infty$, $\sum_{n=1}^{\infty} a_n^2 < \infty$, a stochastic approximation convergence theorem was applied showing that $E(\bar{x}_n - x_n)^2 \to 0$ as $n \to \infty$. The last two hypotheses are identical with T3 and T4 of Sec. 1.2.6. The Lipschitz condition on $f(x)$

was a sufficient restriction on the message dynamics to allow use of the convergence theorem from stochastic approximation. Recently Wolverton [22], using the same hypotheses, proved convergence with probability one.

Albert and Gardner [20] considered the general filtering problem. Their hypotheses are lengthy and provide no motivation for this discussion. The important point is the complete reliance of their convergence proofs on techniques developed in stochastic approximation convergence proofs. Their main conclusion was: $\limsup E(\bar{x}_n - x_n)^2 \leq$ known bound.

Two recent theses, one in pattern recognition, the other in system identification, illustrate the possibilities of using stochastic approximation theory in developing approximately-optimal convergent algorithms. Blaydon [23] is concerned with a pattern recognition problem in which a probability density function, specified in terms of unknown parameters, is to be learned. He derives a recursive estimation algorithm via a minimum-mean-square criteria and shows convergence to the true parameter value with a stochastic approximation theorem. In system identification, Donoghue [24] attempts to formulate an algorithm for learning unknown system parameters from noisy observations of the input and output. He approaches the estimation problem from a Bayes-optimal point of view and eventually derives an approximate solution. A stochastic approximation result is applied to show convergence of the algorithm.

1.2.8 Continuous-Time Stochastic Approximation Methods

The stochastic approximation method has proved useful in studying the asymptotic behavior of discrete-time filters. The extensions of the method to continuous-time processes available in the literature are examined in this section. Only a few investigations have been made into convergent, sequential algorithms for continuous-time parameter estimation problems. The major difference between the formulation of the observation process in Sec. 1.1 and the models available in the literature is that Sec. 1.1 places more stringent restrictions on the correlation properties of the observation noise than the models in the literature. Driml and Nedoma [25] and others [26] only assume the noise is stationary and ergodic, while Sakrison [27] adds a complicated condition on the spread of the correlation function. In contrast is the white noise assumption in (1.2).

Initially, the investigations in the literature appear applicable to the parameter estimation study needed for the continuous-time filtering problem. However, detailed analyses of the theorems indicate that the proposed algorithms and methods of proof are very specialized and not applicable to the general filtering problem. Sakrison [27] and others [26] postulate a structure for the observation process which does not resemble (1.2) and their entire development requires this special structure. Only Driml and Nedoma have a formulation compatible with (1.2). They assume very restrictive estimation algorithms and these special forms fit neatly into a convergence proof relying on the Law of Large Numbers. This method of proof cannot be extended to the algorithms needed for filtering.

All the continuous-time, stochastic approximation papers mentioned above rely principally on the Law of Large Numbers for their convergence proofs. This appears to be the only method available for treating general, non-white, observation noise; to extend this approach to more meaningful estimation algorithms may require generalizations or extensions of ergodic theory.

Figure 1 summarizes approaches to discrete-time and continuous-time filtering.

1.3 Thesis Objectives

The emphasis of the research reported in this thesis is on formulating an approach to the non-linear filtering problem of (1.2) by which asymptotic convergence of filtering algorithms can be exhibited and on demonstrating useful analysis techniques. In a sense, the discrete-time approach to asymptotically-stable, recursive algorithms will be generalized to the continuous-time filtering problem.

Investigators of discrete-time problems relied upon stochastic approximation methods to prove their convergence results but stochastic approximation methods have not been developed for the continuous-time filtering problem. The first major objective of this thesis will be to state and prove convergence theorems analogous to theorems from the discrete-time stochastic approximation literature.

After developing an approach to continuous-time stochastic approximation, the second objective will be to examine a class of sequential solutions for the filtering problem defined in Sec. 1.1. Since the general problem is very difficult to study, this thesis will concentrate on the scalar, noiseless-message case; i.e. $\underline{u}(t) = 0$, $m = n = 1$ in (1.1) and (1.2). Under these assumptions, the message may be considered to be a time-varying parameter

Filtering Problem / Investigations	Discrete-Time		Continuous-Time	
	Noise: Independent and Gaussian	Noise: Independent, non-Gaussian	Noise: white and Gaussian	Noise: Non-white
Approximately-optimal filters	Many results, Refs: 12, 16, 18, 19	One study, Ref: 18	Many results Refs: 3, 12, 14, 29, 30, 18	None
Convergent algorithms for parameter estimation (stochastic approximation)	Independent noise case applicable	Well studied Refs: 20, 28	None	A few results available, Refs: 25, 27, 26
Algorithms with asymptotic properties for general filtering problems	Independent noise case applicable	A few studies available, Refs: 20, 21	None	None
Approximately-optimal convergent algorithms for parameter estimation	A few studies available Refs: 24, 23	None	None	None
Approximately-optimal convergent filters for general filtering problems	None	None	None	None

Figure 1. Summary of Literature Review

specified by the solution to a deterministic differential equation excited by an unknown initial condition. The ideas developed for the stochastic approximation section will be utilized to suggest filtering algorithms whose error variances converge to zero.

1.4 Thesis Outline

The discrete-time results for recursive filters with known asymptotic properties cited in Secs. 1.2.6 and 1.2.7 were developed for problems where the noise distributions were unknown. The analogous situation in the continuous-time filtering problem is to assume only white noise in (1.1) and (1.2). However, the more general non-Gaussian case of filtering cannot be treated because the mathematical techniques available for handling non-linear stochastic differential equations are restricted to problems involving Gaussian, white noise. This restriction has not been effectively explained in engineering literature, so Appendix A is devoted to representations of white noise and their relation to stochastic differential equations.

The non-linear filtering problem treated in this thesis is defined rigorously in Chapter 2. To supplement later discussions in Chapters 3 and 4, an approximately-optimal filter is derived based on a representation for an optimum filter. The class of sequential filters, whose asymptotic convergence will be investigated in later chapters, is also defined and discussed.

The original contributions of this thesis begin with the statement and proof of two theorems in Chapter 3 similar to theorems from discrete-time stochastic approximation. The filtering problem is attacked in Chapter 4 and the techniques developed in Chapter 3 are utilized to demonstrate convergence for the class of sequential filters defined in Chapter 2. This thesis marks the first time that stochastic approximation ideas have been applied to the problem of estimating a time-varying parameter described by a differential equation. Accordingly, every aspect of this work is original. The last section of Chapter 4 specifies conditions on (1.1) and (1.2) which guarantee the asymptotic convergence of the approximately-optimal filter defined in Chapter 2. These conditions indicate the existence, under certain conditions, of nearly-optimal filters satisfying the requirements of good and useful filters.

Chapter 5 examines an example of a non-linear filtering problem. The performance of the approximately-optimal filter documented in Chapter 2 is displayed by means of error profiles generated from a digital computer simulation.

The results of the thesis are reviewed in Chapter 6 and conclusions are drawn concerning the application of this approach in demonstrating asymptotic convergence. A number of extensions of these investigations are proposed.

CHAPTER 2

A FILTERING PROBLEM AND A CLASS OF SEQUENTIAL SOLUTIONS

Chapter 1 outlined the general time-domain filtering problem. This chapter is devoted to rigorously defining the particular problem of concern in this thesis. In subsequent chapters, the emphasis will be on investigations into the asymptotic convergence of a class of sub-optimal, sequential filters; this class is defined and discussed in Sec. 2.3. Chapters 3 and 4 will indicate that this class of sub-optimal filters satisfies two of the three goodness criteria discussed in Chapter 1. In hopes of displaying filters meeting all three performance criteria, Sec. 2.2 derives a nearly-optimal filter to which the convergence results will be applicable.

2.1 The Non-linear Filtering Formulation

This thesis examines solutions for a particular case of the general filtering problem defined in Sec. 1.1. Although the signal and observation processes were defined using conventional engineering models, the type of analysis performed in later chapters necessitates a rigorous mathematical representation for these processes. The reason for the special care being taken to define the models of these processes is to avoid the technical controversies which have plagued investigations of optimal, continuous-time, non-linear filters [31], [32].

The following definition of the scalar "noiseless-message-filtering" problem is a specialization of the general definition given in Sec. 1.1. The signal process is assumed to be the solution to a first-order differential equation:

$$\frac{dx(t)}{dt} = f(t,x(t)) \; : \qquad x(t_o) = b, \text{ an unknown constant} \tag{2.1}$$

and $f(\cdot,\cdot)$ is a specified function. The absence of a noise driver in this equation explains the "noiseless" adjective. The function $f(t,\cdot)$ is assumed to satisfy all sufficient conditions which guarantee the existence, uniqueness, and continuity of the solution to (2.1). A complete discussion of these conditions is given in Coddington and Levinson [33].

The observed process $y(t)$ retains the structure of Sec. 1.1;

$$y(t) = h(t,x(t)) + v(t) \tag{2.2}$$

where $h(\cdot,\cdot)$ is a known function and $v(t)$ is a zero-mean, Gaussian, white noise process with correlation function $Ev(t)v(s) = \delta(t-s)$. The structure of (2.2) is general enough to make $y(t)$ a model of many physical processes. Kailath [34] discusses when (2.2) is a reasonable representation for real-world processes.

Appendix A discusses the mathematical complications of defining a white noise representation which will provide valid solutions for non-linear stochastic differential equations involving this white noise. The appendix indicates the necessity for restating the observed process, (2.2), in terms of an independent-increment, Brownian motion process $B(t)$. Applying Ito's stochastic integral when manipulating equations having the form (2.2) eliminates the use of delta functions; Ito calculus provides consistent meanings for all equations. Specifically, $z(t)$ is defined by (2.3) and (2.2) is taken as equivalent to (2.4).

$$z(t) = \int_{t_o}^{t} y(s)ds \tag{2.3}$$

$$dz(t) = h(t,x(t))dt + dB(t) \tag{2.4}$$

Equation (2.4) is equivalent to (2.2) if (2.4) is divided formally by dt and $\frac{dB(t)}{dt}$ is interpreted as the Gaussian white noise, $v(t)$. Equation (2.4) is interpreted as being equivalent to (2.5), a random integral equation.

$$z(t) = z(t_o) + \int_{t_o}^{t} h(s,x(s))ds + \int_{t_o}^{t} dB(s) \tag{2.5}$$

The first integral in (2.5) is the ordinary Riemann integral, while the second is Ito's stochastic integral. In addition to Appendix A, a recent paper by Wonham [35] briefly surveys the properties of (2.4).

Equations (2.1) and (2.4) define models of the signal and observation processes. The formulation of the noiseless message problem may be completed by summarizing the discussion of filter solutions in Sec. 1.1. The observed waveform is denoted by $\{y(s) = \frac{dz(s)}{ds};\ t_o \le s \le t\} = y_{t_o,t}$. A filter may be defined as a functional mapping of $y_{t_o,t}$ into a real variable for each $t \ge t_o$. The filtering problem is considered solved when a particular mapping has characteristics which satisfy the three goodness criteria of optimality, asymptotic convergence, and sequential structure.

2.2 An Approximately-Optimal Filter

This section develops a filter which is approximately-optimal and sequential. A proper evaluation of any approximation requires the optimal solution to the filtering problem. It is important to note that, when considering an optimal solution to the filtering problem, a probability distribution will be assigned to the initial condition on the message process, $x(t_o) = b$. Since optimal filters are concerned, at least in part, with accurate estimates of the message during the transient portion of the filters' time responses, it is mandatory that all moments of the initial condition be known.

A precise development of an optimal filter is contained in the papers of Kushner [2] and Kallianpur [13]. The stochastic differential equation for the expectation of any function, $g(t,x(t))$, of the message, $x(t)$, and the time, t, conditioned on the observed waveform $z_{t_o,t}$ is developed in these papers; conditioning on $z_{t_o,t}$ is equivalent to conditioning on $y_{t_o,t}$ by (2.3). Letting $Lg(t,x(t)) = f(t,x(t))\dot{g}(t,x(t))$, where $\dot{g}(t,x) = \frac{\partial}{\partial x} g(t,x)$, this stochastic differential equation is

$$d\hat{g}(t) = \widehat{Lg}(t)dt + [\widehat{gh}(t) - \hat{g}(t)\hat{h}(t)][dz(t) - \hat{h}(t)dt] \qquad (2.6)$$

where

$$\hat{g}(t) = E[g(t,x(t))|z_{t_o,t}]$$

$$\widehat{Lg}(t) = E[f(t,x(t))\dot{g}(t,x(t))|z_{t_o,t}]$$

$$\widehat{gh}(t) = E[g(t,x(t))h(t,x(t))|z_{t_o,t}]$$

$$\hat{h}(t) = E[h(t,x(t))|z_{t_o,t}]$$

$$\hat{x}(t) = E[x(t)|z_{t_o,t}] .$$

When $g(t,x) = x$, (2.6) becomes

$$d\hat{x}(t) = \hat{f}(t)dt + [\widehat{xh}(t) - \hat{x}(t)\hat{h}(t)][dz(t) - \hat{h}(t)dt] \qquad (2.7)$$

with $\hat{x}(t)$ the conditional mean. This thesis will follow the convention of calling the differential equation (2.7) the representation of the optimal filter, even though there are other legitimate choices for an "optimal" filter.

In general, the optimal filter (2.7) can never be realized since an infinite-dimensional system of stochastic differential equations must be solved to determine the random functions $\hat{f}(t)$, $\widehat{xh}(t)$, and $\hat{h}(t)$. Appendix C displays this infinite-dimensional system.

Approximations must be made to produce a finite set of equations. An approximation scheme is now described. The structure of the resulting filter is very similar to that of the linear Kalman-Bucy filter (Sec. 1.2.1). The conditional mean $\hat{x}(t)$ is assumed to be in the neighborhood of $x(t)$ so that

$$f(t,x(t)) \doteq f(t,\hat{x}(t)) + \dot{f}(t,\hat{x}(t))[x(t) - \hat{x}(t)] \qquad (2.8)$$

$$h(t,x(t)) \doteq h(t,\hat{x}(t)) + \dot{h}(t,\hat{x}(t))[x(t) - \hat{x}(t)] \; . \qquad (2.9)$$

It is also assumed that the third conditional moment,

$$E([\hat{x}(t) - x(t)]^3 | z_{t_o,t}) \doteq 0. \qquad (2.10)$$

An examination of these assumptions should provide insight into how well the approximately-optimal filter described in (2.13) and (2.14) will satisfy the goodness criteria R.1, optimality. This approximate filter, to be studied in Chapter 4, is now derived using (2.6) - (2.10). Equation (2.8) shows that $\hat{f}(t) \doteq f(t,\hat{x}(t))$ (2.11) so that (2.7) becomes

$$d\hat{x}(t) \doteq f(t,\hat{x}(t))dt + P(t)\dot{h}(t,\hat{x}(t))[dz(t) - h(t,\hat{x}(t))dt] \qquad (2.12)$$

where $P(t) = \widehat{x^2}(t) - \hat{x}(t)^2$.

The function $P(t)$ can be approximated from (2.6) and Ito's Lemma (Appendix B). Letting $g(t,x) = x^2$ and using (2.6) along with (2.8) - (2.10),

$$d\widehat{x^2}(t) \doteq 2\,\hat{x}(t)f(t,\hat{x}(t))dt + 2\,\dot{f}(t,\hat{x}(t))P(t)dt$$
$$+ \dot{h}(t,\hat{x}(t))[\widehat{x^3}(t) - \widehat{x^2}(t)\hat{x}(t)][dz(t) - h(t,\hat{x}(t))dt].$$

Applying Ito's lemma, Appendix B, to (2.12) gives

$$d\hat{x}(t)^2 \doteq 2\,\hat{x}(t)f(t,\hat{x}(t))dt + 2\,\hat{x}(t)\dot{h}(t,\hat{x}(t))P(t)[dz(t) - h(t,\hat{x}(t))dt]$$
$$+ [\dot{h}(t,\hat{x}(t))P(t)]^2 dt.$$

The differential of $P(t)$ is now approximated by

$$dP(t) = d[\widehat{x^2}(t) - \hat{x}(t)^2] \doteq 2\,\dot{f}(t,\hat{x}(t))P(t)dt - [\dot{h}(t,\hat{x}(t))P(t)]^2 dt$$
$$+ \dot{h}(t,\hat{x}(t))[\widehat{x^3}(t) - \widehat{x^2}(t)\hat{x}(t) - 2\,\hat{x}(t)P(t)][dz(t)-h(t,\hat{x}(t))dt$$

The condition (2.10) can be rewritten as

$$\widehat{x^3} - 3\,\widehat{x^2}\hat{x} + 2\,\hat{x}^3 = \widehat{x^3} - \widehat{x^2}\hat{x} - 2\,\hat{x}P \doteq 0.$$

The simultaneous integration of the following stochastic differential equations provides an approximately-optimal filter; $\hat{x}(t)$ is the filter output.

$$d\hat{x}(t) = f(t,\hat{x}(t))dt + P(t)\dot{h}(t,\hat{x}(t))[dz(t) - h(t,\hat{x}(t))dt];\ \hat{x}(t_o) = E(b) \qquad (2.13)$$

$$\frac{dP(t)}{dt} = 2\,\dot{f}(t,\hat{x}(t))P(t) - [\dot{h}(t,\hat{x}(t))P(t)]^2\ ;\quad P(t_o) = \mathrm{Var}\ (b) \qquad (2.14)$$

Recalling the equivalence of (2.2) and (2.4) and dividing by dt permits (2.13) to be interpreted as

$$\frac{d\hat{x}(t)}{dt} = f(t,\hat{x}(t)) + P(t)\dot{h}(t,\hat{x}(t))[y(t) - h(t,\hat{x}(t))]$$

Retaining (2.13) in its special differential form will emphasis the necessity of using Ito calculus for its solution. Equation (2.14) does not contain any white noise terms and may be treated as an ordinary differential equation. This development is a simplified version of work performed by Bass, Norum, and Schwartz [29].

The approximately-optimal filter (2.13) - (2.14) was derived from a solution which minimized the error variance at each instant in the observation interval. Bellman obtained the same equations by using a least-squares criteria and the theory of invariant imbedding [30]. Friedland and Bernstein also derived this filter as an approximate solution to the filtering problem analyzed from the maximum likelihood approach [18].

This filter, besides being a common academic filtering solution, has been applied to numerous real-world problems [19], [36]. Its performance was analyzed in a recent paper by Stear and Schwartz; their computer simulation compared several approximately-optimal filters and showed no distinction between this approximation scheme and other, more nearly optimum filters [37]. It may thus be concluded that this approximate filter is a realistic solution to the optimal filtering problem.

2.3 Sub-Optimal, Sequential Filters

The following sequential filters are suggested as solutions to the filtering problem defined in Sec. 2.1:

wh

$\bar{x}$(

ou

is

(2

di

wh

as

me

is

a(

ve

wi

wo

ex

th

in

th

st

2.

ma

al

$$d\bar{x}(t) = f(t,\bar{x}(t))dt + a(t,\bar{x}_{t_o,t})[dz(t) - h(t,\bar{x}(t))dt] \qquad (2.15)$$

where $\bar{x}(t)$ is the filter output with an arbitrary initial value $\bar{x}(t_o)$. The gain term, $a(t,\bar{x}_{t_o,t})$, is a time-varying function of the output waveform $\bar{x}_{t_o,t} = \{\bar{x}(s): t_o \le s \le t\}$. The structure of (2.15) is very similar to that of the approximately-optimal filter in (2.13) - (2.14). In fact, if $a(t,\bar{x}_{t_o,t}) = P(t)\dot{h}(t,\bar{x}(t))$ and the initial conditions are set equal, the filters are identical.

The filtering solutions proposed in (2.15) provide filters which will fulfill two of the requirements for a useful filter; R.2, asymptotic convergence, and R.3, sequential structure. These requirements can be met even if $\bar{x}(t_o)$ is not specified so no prior knowledge is required of the initial condition $x(t_o)$ in (2.1).

The outstanding feature of (2.15) is the random gain term, $a(t,\cdot)$. This gain term must satisfy conditions guaranteeing convergence but, otherwise, has an arbitrary form. It may be possible, within these convergence limitations, to specify a gain function which would allow the filter to fulfill the optimality requirement R.1. An example of this idea would be to set $a(t,\bar{x}_{t_o,t}) = P(t)\dot{h}(t,\bar{x}(t))$. On the other hand, the discussions in Secs. 1.1 and 1.2.3 indicate that in many applications one may prefer that $a(t,\cdot)$ be deterministic, thus providing the filter with a simple, asymptotically convergent structure.

2.4 Summary

Equation (2.15) introduces a class of sequential filters which may include an approximately-optimal filter. The convergence of algorithms from this class will be the main concern of the remaining chapters.

This chapter also introduces the concepts of Ito calculus and stochastic differential equations into the non-linear filtering problem. An understanding of these concepts has enabled investigators to develop an acceptable optimal filter theory. This knowledge and a familiarity with the discrete-time stochastic approximation methods lead to asymptotic convergence studies.

CHAPTER 3

CONVERGENT ALGORITHMS FOR PARAMETER ESTIMATION

Many of the statistical methods which have been developed for the classic problem of estimating unknown parameters have provided approaches to filtering problems. The literature review in Chapter 1 showed that stochastic approximation schemes, developed for discrete-time parameter estimation, suggest ideas for analyzing the asymptotic behavior of the continuous-time filtering algorithms proposed in Chapter 2. This chapter concentrates on the problem of continuous-time parameter estimation by developing two convergence theorems analogous to theorems from discrete-time stochastic approximation.

The filtering problem discussed in Sec. 1.1 may be considered equivalent to a problem in estimation theory where the filter's output provides an estimate of the present message value. For this reason, the problem of determining, or estimating, parameters is a special case of the filtering problem with $f(t,x) = 0$ in (2.1). Optimum solutions for both filtering and parameter estimation problems are based on similar theorems and procedures [12]. There is, however, a conceptual difference between the two problems. A filter's output attempts to follow the time variations of the message process; when the message is random, the output can never follow the message exactly. The parameter estimation problem, where the message is time-invariant, requires an estimation algorithm which eventually determines the <u>true</u> value of the unknown constant. It is for this reason that any study of a practical parameter estimation scheme must include an analysis of asymptotic behavior.

This chapter analyzes the large-time behavior of the estimator represented by (2.15) with $f(t,x) = 0$. This sub-optimal estimator is shown to converge asymptotically to the true value of the parameter. The chapter is broken into three parts. Section 3.1 contains a list of assumptions regulating the behavior of the observations, (2.4), and the gain function in (2.15). Also, two lemmas are developed for later use. In Sec. 3.2 and 3.3, the assumptions are used as hypotheses for two convergence theorems. These theorems show that certain estimators converge to the true parameter with probability one and in the mean-square sense. The theorems are a major result and indicate that the sequential estimators defined in (2.15) behave asymptotically as practical estimators.

The convergence theorems do not require any structure for the gain functions, so in the last part of this chapter a specific class of gain functions is proposed. Corollary 3.3 shows that this class of functions satisfies the assumptions of the main convergence theorems and that the approximately-optimal filter derived in Chapter 2 and specialized in this chapter has a gain function which is a member of this class of functions. So, under certain assumptions on the observation process, an approximately-optimal algorithm is shown to converge to the true parameter value. This is a major result since it implies an estimation scheme which satisfies the three requirements of a useful filter.

All aspects of this chapter are new and original. Convergence of the estimator algorithm (2.15), in which the information supplied by the past observations is contained in $a(t,\bar{x}_{t_o,t})$, has never been proved for a continuous-time problem. The assumptions given in Sec. 3.1 are directly analogous to the discrete-time assumptions of Albert and Gardner [20] but the proofs of the convergence theorems do not follow in a straight-forward manner.

3.1 Preliminary Considerations

This section provides the groundwork necessary for deriving the convergence theorems in Sec. 3.2. This basic material includes the formulation of the parameter estimation problem, a description of the estimation schemes being investigated in this chapter, two useful lemmas, and the hypotheses for the convergence theorems.

The parameter estimation problem and possible solutions considered in this chapter are defined using equations equivalent to those in Chapter 2. The parameter to be estimates is θ and if $x(t) = x(t_o) = \theta$, (2.4) represents noisy observations of this unknown parameter. This observed process is modeled by

$$dz(t) = h(t,\theta)dt + dB(t) \qquad (2.4)$$

with $\frac{dz(t)}{dt} = y(t)$ the observed process and $B(t)$ a Brownian motion process. In the first three sections of this chapter, θ is not assigned a prior distribution. The class of sub-optimal estimation algorithms proposed in (2.15) with $f(t,x) = 0$ are considered as the possible estimators for θ:

$$d\bar{x}(t) = a(t,\bar{x}_{t_o,t})[dz(t) - h(t,\bar{x}(t))dt] : \quad \bar{x}(t_o) \text{ arbitrary} \qquad (3.1)$$

where $\bar{x}(t)$ is the estimate of θ at time t.

The estimator error is denoted as $e(t) = \bar{x}(t) - \theta$ and, according to (2.4) and (3.1), has the following differential representation.

$$de(t) = a(t,\bar{x}_{t_o,t})[h(t,\theta) - h(t,\bar{x}(t))]dt + a(t,\bar{x}_{t_o,t})dB(t) \quad (3.2)$$

An examination of (3.2) brings out many of the difficulties inherent in a study of the asymptotic convergence of $e(t)$. First of all, (3.2) is a stochastic differential equation containing the differential of Brownian motion. Strict attention must be paid to the properties of Ito calculus in all further derivations. Since (3.2) is non-linear and since $a(t,\bar{x}_{t_o,t})$, the gain function, is unspecified, an analytical solution for $e(t)$ is impossible. Some description of the behavior of the gain function is mandatory and assumptions must be placed on $h(t,x)$ before any inferences can be made about $e(t)$ and/or $e(t)^2$.

The development of convergence theorems for equations such as (3.2) is not obvious. In a survey of the discrete-time stochastic approximation literature, Martingale theory was often used to obtain convergence with probability one [38], [39], [40]. Martingale theory and Ito calculus seem very compatible and both are used in later proofs. A survey of stochastic approximation methods does suggest certain assumptions about $h(t,x)$ and $a(t,\bar{x}_{t_o,t})$ which can be used effectively to specify the asymptotic behavior of the estimation error. This background of discrete-time investigations provides a fruitful source of ideas for analyzing (3.2).

Before listing the assumptions sufficient for convergence, two lemmas are stated and proved. These lemmas are applied in later proofs both in this chapter and in Chapter 4.

Lemma 3.1 Let $w(t)$ be a stochastic process satisfying

$$w(t) \le w(r) + \int_r^t g(s)dB(s) \quad \text{for} \quad t,r \ge t_o \qquad (3.3)$$

where $B(s)$ is a Browian motion process. Assume that A_t is a σ-algebra, that $w(t)$ is measurable on A_t, and that $A_t \subset F\{[t_o,t], B(s)\}$, the σ-algebra generated by $B(s)$, $t_o \le s \le t$. Also assume that $g(t)$, a random function, is measurable on $F\{[t_o,t], B(s)\}$. Then $w(t)$ is a positive super-martingale.[2]

Proof. Apply the expectation relative to A_r to each side of (3.3).

$$E\{w(t)|A_r\} \le w(r) + E[\int_r^t g(s)dB(s)|A_r] \qquad (3.4)$$

For $w(t)$ to be a super-martingale the second term on the right side of the inequality must be zero. A property of conditional expectations[3] shows that

$$E[\int_r^t g(s)dB(s)|A_r] = E\{E[\int_r^t g(s)dB(s)|F\{[t_o,r], B(u)\}]|A_r\}$$

Since $g(s)$ is measurable on $F\{[t_o,s], B(u)\}$ the conditional expectation satisfies property B.2; i.e.

$$E[\int_r^t g(s)dB(s)|F\{[t_o,r], B(u)\}] = 0 \qquad (3.4a)$$

This completes Lemma 3.1.

Lemma 3.2 If $m(t) > 0$ and $M(t) = c + \int_{t_o}^t m(s)ds$ with $c < \infty$ arbitrary and if $M(t) \to \infty$ as $t \to \infty$ then

$$\int_{t_o}^t \frac{m(s)}{M(s)^\alpha} ds \quad \begin{cases} \text{diverges if } \alpha \le 1 \\ \text{converges if } \alpha > 1 \end{cases}.$$

---

[2] Consult Appendix B for discussion of σ-algebras and martingales.

[3] Doob [41], p. 37.

Proof. Assume $\alpha \le 1$; there exists a $T_o > t_o$ such that for $t > T_o$, $M(t) \ge 1$. Now for $a \ge T_o$ and $b \ge 0$ and because of the monotone behavior of $M(t)$,

$$\int_a^{a+b} \frac{m(s)ds}{M(s)^\alpha} \ge \frac{1}{M(a+b)^\alpha} \int_a^{a+b} m(s)ds \ge \frac{1}{M(a+b)} \int_a^{a+b} m(s)ds$$

The right hand side can be integrated to obtain

$$\int_a^{a+b} \frac{m(s)ds}{M(s)^\alpha} \ge \frac{M(a+b) - M(a)}{M(a+b)} = 1 - \frac{M(a)}{M(a+b)}$$

The monotone property of $M(t)$ shows that for every $a \ge T_o$, there exists a $T \ge T_o$ such that for $a+b > T$, $1 - \frac{M(a)}{M(a+b)} > \frac{1}{2}$

This contradicts the necessary Cauchy condition [42] for the convergence of

$$\int_{t_o}^{t} \frac{m(s)ds}{M(s)^\alpha}$$

Assume $\alpha > 1$, let $\alpha = 1 + \delta$ where $\delta > 0$, examine the integral

$$\int_{t_o}^{t} \frac{d}{ds} [M(s)]^{-\delta} ds = M(t)^{-\delta} - M(t_o)^{-\delta} \xrightarrow[t\to\infty]{} - M(t_o)^{-\delta} < \infty$$

Evaluate the derivative in the integral.

$$\frac{d}{ds} M(s)^{-\delta} = \frac{d}{ds}\left[\frac{1}{[c + \int_{t_o}^{s} m(u)du]^{\delta}}\right]$$

$$= - \frac{\delta[c + \int_{t_o}^{s} m(u)du]^{\delta-1} m(s)}{[c + \int_{t_o}^{s} m(u)du]^{2\delta}}$$

$$= - \frac{\delta\, m(s)}{[c + \int_{t_o}^{s} m(u)du]^{1+\delta}} = - \frac{\delta\, m(s)}{M(s)^\alpha}$$

So

$$\int_{t_o}^{t} \frac{m(s)}{M(s)^{\alpha}}\, ds \rightarrow \frac{M(t_o)^{-\delta}}{\delta} \quad \text{as} \quad t \rightarrow \infty .$$

Lemma 3.2 is complete.

The following assumptions are made for (2.4) and the gain function in (3.1). These assumptions are the hypotheses for the main convergence theorems proved in Secs. 3.2 and 3.3 and are analogous to the assumptions Albert and Gardner [20] formulated for their discrete-time observation equation and gain sequences.

Assumption 1: For each value of t, $h(t,x)$ is monotone and differentiable with respect to x.

Assumption 2: The function $a(t,\bar{x}_{t_o,t})$ is measurable on $F\{[t_o,t], \bar{x}(s)\}$, the σ-algebra generated by $\bar{x}(s)$, where $t_o \le s \le t$. The space of all real continuous functions defined on the time interval $[t_o,t]$ is denoted by $C_{t_o}^{t}$; $h(t,\bar{x})$ and $a(t,\bar{x}_{t_o,t})$ are assumed to satisfy all conditions necessary to guarantee the continuity of $\bar{x}(t)$. (Appendix D).

Assumption 3: For each t, the sign of $a(t,X)$ for all $X \in C_{t_o}^{t}$ is constant and equal to the sign of $\dot{h}(t,x) = \frac{\partial}{\partial x} h(t,x)$.

Assumption 4: $\int_{t_o}^{\infty} b(t) \inf|a(t)|\, dt = \infty$ where $b(t) = \inf_{x} |\dot{h}(t,x)|$ and $\inf|a(t)| = \inf_{X \in C_{t_o}^{t}} |a(t,X)|$

Assumption 5: $\int_{t_o}^{\infty} \sup|a(t)|^2 dt < \infty$ where $\sup|a(t)| = \sup_{X \in C_{t_o}^{t}} |a(t,X)|$

3.2 The Main Convergence Theorem

The theorem to be stated and proved in this section shows that the output of the estimation algorithm, $\bar{x}(t)$ in (3.1), converges, as t increases, to the value of the unknown parameter. A basic martingale theorem provides both convergence with probability one and in the mean-square sense.

Before the actual statement of Theorem 3.1, an equation is developed for the square of the estimation error using some of the assumptions listed in Sec. 3.1. The following simplified notation will be used for the remainder of this chapter: $a(t,\bar{x}_{t_o,t}) = a(t)$

Equation (3.2) exhibits the estimation error. Assumption 2 and the development in Appendix D show that $a(t)$ is measurable on $F\{[t_o,t], B(s)\}$, so, if $Z(t,e) = e^2$, Ito's lemma (Theorem B.1) can be applied to (3.2) producing a differential equation for $e(t)^2$.

$$de(t)^2 = 2\, e(t)a(t)[h(t,\theta) - h(t,\bar{x}(t))]dt + a(t)^2 dt + 2\, e(t)a(t)dB(t)$$

This equation is equivalent to (3.5).

$$e(t)^2 = e(r)^2 + \int_r^t 2\, e(s)a(s)[h(s,\theta) - h(s,\bar{x}(s))]ds + \int_r^t a(s)^2 ds$$
$$+ \int_r^t 2\, e(s)a(s)dB(s) \ ; \qquad t,r \ge t_o \tag{3.5}$$

The mean value theorem [42] gives:

$$h(s,\bar{x}(s)) = h(s,\theta) + \dot{h}(s,\phi(s))[\bar{x}(s) - \theta]$$

$$\text{where} \begin{cases} \theta \le \phi(s) \le \bar{x}(s) & \text{if } \bar{x}(s) \ge \theta \\ \bar{x}(s) \le \phi(s) \le \theta & \text{if } \bar{x}(s) \le \theta \end{cases}$$

Assumption 3 implies that:

$$a(s)\dot{h}(s,\phi(s)) = |a(s)||\dot{h}(s,\phi(s))| \ge 0$$

Combining these results with (3.5) gives

$$e(t)^2 \le e(r)^2 - 2\int_r^t |a(s)||\dot{h}(x,\phi(s))|e(s)^2 ds + \int_r^t \sup|a(s)|^2 ds$$

$$+ 2\int_r^t e(s)a(s)dB(s) \tag{3.6}$$

Equation (3.6) will be a basic equation both in the following proof and in the proof of Theorem 3.2 in Sec. 3.3.

Theorem 3.1 Under Assumptions 1-5, the process $\bar{x}(t)$ defined by (3.1) converges to θ with probability one and in the mean square sense:

$$\lim_{t\to\infty} \bar{x}(t) = \theta \quad \text{w.P.1.} \quad \text{and} \quad \lim_{t\to\infty} E(x(t) - \theta)^2 = 0.$$

Proof. Equation (3.6) provides an inequality for the squared error. The second term on the right hand side of (3.6) is negative, thus establishing a simple and very convenient inequality for $e(t)^2$.

$$e(t)^2 \le e(r)^2 + \int_r^t \sup|a(s)|^2 ds + 2\int_r^t e(s)a(s)dB(s) \ : \ r,t \ge t_o \tag{3.7}$$

Defining $w(t) = e(t)^2 + \int_t^\infty \sup|a(s)|^2 ds$ (3.8)

and substituting into (3.7),

$$w(t) \le w(r) + 2\int_r^t e(s)a(s)dB(s) \tag{3.9}$$

A review of the proof up to this point emphasizes the importance of Assumptions 1 through 3 in obtaining a transformation of $e(t)^2$ which satisfies an equation of the form (3.9). It has already been pointed out that $a(t)$ is measurable on $F\{[t_o,t], B(s)\}$ and since $\bar{x}(t)$ is measurable on this σ-algebra, $e(t)a(t)$ is also. The error squared is measurable on $F\{[t_o,t], \bar{x}(s)\}$ and it follows directly that $F\{[t_o,t], e(s)^2\} \subset F\{[t_o,t], B(s)\}$. So when A_t is defined equal to $F\{[t_o,t], e(s)^2\}$, Lemma 3.1 indicates that $w(t)$ is a positive super-martingale.

With the process $w(t)$ being a super-martingale, the function $Ew(t)$ is a non-increasing function of t and since it is bounded below it follows that[4]

$$\lim_{t\to\infty} Ew(t) = E\zeta = K < \infty \tag{3.10}$$

A combination of (3.8), Assumption 5, and (3.10) shows that

$$\lim_{t\to\infty} Ee(t)^2 = E\zeta \tag{3.11}$$

Returning to (3.6) and noting that $\inf|a(s)|b(s) \le |a(s)||\dot{h}(s,\phi))|$ allows another weakening of the inequality on $e(t)^2$.

$$e(t)^2 \le e(r)^2 - 2\int_r^t \inf|a(s)|b(s)e(s)^2 ds + \int_r^t \sup|a(s)|^2 ds$$
$$+ 2\int_r^t e(s)a(s)dB(s) \tag{3.12}$$

The expectation of each side of (3.12) is taken and Fubini's theorem is applied to the second term on the right hand side.

$$Ee(t)^2 \le Ee(r)^2 - 2\int_r^t \inf|a(s)|b(s)Ee(s)^2 ds + \int_r^t \sup|a(s)|^2 ds$$
$$+ 2\, E\left[\int_r^t e(s)a(s)dB(s)\right] \tag{3.13}$$

The proof of Lemma 3.1, specifically the part deriving (3.4a), can be reapplied to show $E\left[\int_r^t e(s)a(s)dB(s)\right] = 0.$ (3.14)

Applying (3.14), (3.13) becomes

$$Ee(t)^2 \le Ee(r)^2 - 2\int_r^t \inf|a(s)|b(s)Ee(s)^2 ds + \int_r^t \sup|a(s)|^2 ds \tag{3.15}$$

The application of (3.11) and Assumption 5 to (3.15) indicates the necessity of (3.16).

---

[4] Doob [41], chapter 7.

$$2 \int_r^{\infty} \inf|a(s)|b(s)Ee(s)^2 ds < \infty \quad \text{for} \quad t \geq t_o \tag{3.16}$$

If (3.16), (3.11), and Assumption 4 are all to be satisfied, it is necessary that

$\lim_{t\to\infty} Ee(t)^2 = 0$, so the mean-square convergence is assured.

The super-martingale inequality (Theorem B.3) shows convergence with probability one. The inequality gives

$$\text{Prob}\{\sup_{t\geq u\geq r} w(u) > \epsilon\} < [Ee(r)^2 + \int_r^{\infty} \sup|a(s)|^2 ds]\frac{1}{\epsilon} \quad \text{for each} \quad t \geq r. \tag{3.17}$$

The definition in (3.8) provides the following relationship

$$\text{Prob}\{\sup_{t\geq u\geq r} e(u)^2 > \epsilon\} < \text{Prob}\{\sup_{t\geq u\geq r} w(u) > \epsilon\}. \tag{3.18}$$

Mean-square convergence and Assumption 5 are now combined with (3.17) and (3.18). Given an $\epsilon > 0$ and $\delta > 0$, there exists a $R > 0$ such that for every $t > R$

$$\text{Prob}[e(u)^2 > \epsilon : t \geq u \geq R] < \delta$$

thus

$$\lim_{t\to\infty} e(t) = 0 \qquad \text{w.P.1.}$$

This completes the proof of Theorem 3.1.

3.3 An Alternate Convergence Proof

The proof of Theorem 3.1 was structured with convergence with probability one in mind; mean-square convergence was a by-product. The mean-square result was unexpected since, in all known stochastic approximation theories, demonstrations of mean-square and of probability-one convergence have required separate theorems. The theorem in this section demonstrates another way to prove the mean-square property. The techniques used in its proof will be applicable to part of the filtering investigation in Chapter 4.

Theorem 3.2 Under Assumptions 1-5 the process $\bar{x}(t)$ defined in (3.1) converges in the mean-square sense to θ; $\lim_{t\to\infty} E(\bar{x}(t) - \theta)^2 = 0$.

Proof. The inequality in (3.6) is weakened by applying the fact that $\inf|a(s)|b(s) \le |a(s)||\dot{h}(s,\phi(s))|$. The expectation of each side is taken to give

$$Ee^2(t) \le Ee^2(r) - 2\int_r^t \inf|a(s)|b(s)Ee^2(s)ds + \int_r^t \sup|a(s)|^2 ds. \quad (3.19)$$

The expectation of the last term in (3.6) is zero for the reasons given in Theorem 3.1. The remainder of this proof relys on the construction of a function which dominates $Ee^2(t)$. The function $c(t)$ is defined as the solution to

$$\frac{dc(t)}{dt} = -2\,b(t)\inf|a(t)|c(t) + \sup|a(t)|^2 \;:\; c(t_o) = Ee^2(t_o) \quad (3.20)$$

Equation (3.20) is a linear equation having the following solution:

$$c(t) = D(t,t_o)Ee^2(t_o) + \int_{t_o}^t D(t,s)\sup|a(s)|^2 ds \quad (3.21)$$

In (3.21), $$D(t,s) = \exp[-2\int_s^t b(u)\inf|a(u)|du] \quad (3.22)$$

Equation (3.20) is now written as an integral equation.

$$c(t) = c(r) - 2\int_r^t b(s)\inf|a(s)|c(s)ds + \int_r^t \sup|a(s)|^2 ds;\; t,r \ge t_o \quad (3.23)$$

Substracting (3.23) from (3.19),

$$[Ee^2(t) - c(t)] + 2\int_r^t b(s)\inf|a(s)|[Ee^2(s) - c(s)]ds \le [Ee^2(r) - c(r)]. \quad (3.24)$$

Equation (3.24) indicates that $c(t) \geq Ee^2(t)$. To verify this, suppose it is not true. Then there exists a time T_1 such that $Ee^2(T_1) - c(T_1) = 0$. If $r = T_1$ and $t = T_1 + \delta$ where δ is the amount of time $Ee^2(t) - c(t)$ is positive, then $\delta > 0$ since $Ee^2(t)$ and $c(t)$ are continuous and (3.24) provides a contradiction.

Equation (3.21) shows that

$$Ee^2(t) \leq D(t,t_o)Ee^2(t_o) + \int_{t_o}^{t} D(t,s)\sup|a(s)|^2 ds \tag{3.25}$$

The estimator defined in (3.1) converges to θ in the mean-square sense when the two terms on the right hand side of (3.25) converge to zero. The first term is easily handled by an application of Assumption 4 to $D(t,t_o)$ so that

$$D(t,t_o) \to 0 \quad \text{as} \quad t \to \infty. \tag{3.26}$$

Assumption 5 and (3.26) must both be used on the second term. Arbitrarily selecting $\varepsilon > 0$, a $T > 0$ can be found so that $\int_T^t \sup|a(s)|^2 ds < \varepsilon/2$ and $D(t,t_o) \leq 1$ for $t > T$. Writing the integral term under investigation as a sum shows that

$$\int_{t_o}^{t} D(t,s)\sup|a(s)|^2 ds \leq \int_T^t \sup|a(s)|^2 ds + \int_{t_o}^{T} D(t,s)\sup|a(s)|^2 ds \quad \text{for } t > T.$$

The constant ε_2 is defined as $\varepsilon_2 = \dfrac{\varepsilon}{2\int_{t_o}^{T} \sup|a(s)|^2 ds}$.

There exists a $T_2 > T$ such that $D(t,s) < \varepsilon_2$ for $t > T_2$ and $s < T$. Therefore, for every $\varepsilon > 0$ there exists a $T_2 > 0$ such that for $t > T_2$,

$$\int_{t_o}^{t} D(t,s)\sup|a(s)|^2 ds \leq \frac{\varepsilon}{2} + \frac{\varepsilon \int_{t_o}^{T} \sup|a(s)|^2 ds}{2\int_{t_o}^{T} \sup|a(s)|^2 ds} = \varepsilon,$$

which shows that $\lim_{t\to\infty} Ee^2(t) = 0$.

This completes the proof.

3.4 Particular Gain Functions

The hypotheses in Sec. 3.1 guarantee the convergence to θ of the estimator defined in (3.1). However, they are very general. This section lists a specific class of gain functions which satisfy the five assumptions of Sec. 3.1. Before discussing the gains, a few assumptions about $h(t,x)$ are listed which are a combination of one of the assumptions in Sec. 3.1 and two new ones. These new assumptions can replace those of Sec. 3.1 when the gains of this section are used in (3.1).

Assumption S1: The function $h(t,x)$ is monotone and differentiable in x for each t.

Assumption S2: $\sup_{t_o \le t < \infty} \frac{g(t)}{b(t)} \le K < \infty$ where $g(t) = \sup_x |\dot{h}(t,x)|$

Assumption S3: $\lim_{t\to\infty} B(t)^2 = \infty$ with $B(t)^2 = \int_{t_o}^{t} b(s)^2 ds$

Corollary 3.3 With Assumptions S1 - S3, gain functions $a(t,\bar{x}_{t_o},t)$ having the form of (3.27) satisfy Assumptions 1-5 of Sec. 3.1.

$$a(t,\bar{x}_{t_o},t) = \frac{\alpha(t,\bar{x}_{t_o},t)^2 \operatorname{sgn} \dot{h}(t,\bar{x}(t))}{\sigma(t,\bar{x}_{t_o},t)[c + \int_{t_o}^{t} \gamma(s,\bar{x}_{t_o},s)^2 ds]}, \tag{3.27}$$

where the three functions $\alpha(t,\bar{x}_{t_o},t)$, $\sigma(t,\bar{x}_{t_o},t)$, and $\gamma(t,\bar{x}_{t_o},t)$ are less than $g(t)$, are greater than, $b(t)$ and are measurable on $F\{[t_o,t], \bar{x}(s)\}$.

Proof. Assumptions 1-3 are satisfied by inspection. The integrand of Assumption 4 is now examined.

$$b(t)\inf|a(t)| \geq \frac{b(t)^3}{g(t)[c + \int_{t_o}^{t} g(s)^2 ds]} \geq \frac{b(t)^3}{K^3\ b(t)[c' + \int_{t_o}^{t} b(s)^2 ds]}$$

This shows that

$$\int_{t_o}^{t} b(s)\inf|a(s)|ds \geq \frac{1}{K^3}\int_{t_o}^{t} \frac{b(s)^2}{[c' + \int_{t_o}^{s} b(u)^2 du]}\, ds\ .$$

Lemma 3.2 and Assumption S3 indicate that Assumption 4 is satisfied. Similarily,

$$\sup|a(t)|^2 \leq \left[\frac{g(t)^2}{b(t)[c + \int_{t_o}^{t} b(s)^2 ds]}\right]^2 \leq \frac{K^4 b(t)^4}{b(t)^2[c + \int_{t_o}^{t} b(s)^2 ds]^2}$$

which shows that

$$\int_{t_o}^{t} \sup|a(s)|^2 ds \leq K^4 \int_{t_o}^{t} \frac{b(s)^2}{[c + \int_{t_o}^{s} b(u)^2 du]^2}\, ds\ .$$

Assumption 5 also follows from Assumption S3 and Lemma 3.2.

One gain function which satisfies (3.27) is of unusual interest. This function is

$$a(t,\bar{x}_{t_o},t) = \frac{\dot{h}(t,\bar{x}(t))}{[1/P + \int_{t_o}^{t} \dot{h}(s,\bar{x}(s))^2 ds]} \tag{3.28}$$

where $P = \text{var}\ \theta$. For the discussion of this gain and the corresponding estimator, θ, the unknown constant, is assigned a prior distribution and the requirement $x(t_o) = E(\theta)$ is made for (3.1). To better understand the significance of (3.28) $P(t)$ is introduced.

$$P(t) = \frac{1}{[P^{-1} + \int_{t_o}^{t} \dot{h}(s,\bar{x}(s))^2 ds]}$$

Differentiating,

$$\frac{dP(t)}{dt} = \frac{d}{dt}\left[\frac{1}{[P^{-1} + \int_{t_o}^{t} \dot{h}(s,\bar{x}(s))^2 ds]}\right]$$

$$= - \frac{\dot{h}(t,\bar{x}(t))^2}{[P^{-1} + \int_{t_o}^{t} \dot{h}(s,\bar{x}(s))^2 ds]^2} = - [\dot{h}(t,\bar{x}(t))P(t)]^2 . \quad (3.29)$$

Now, (3.29) is the same as (2.14) when $f(t,x) = 0$; i.e., when parameter estimation is considered as a filtering problem. The algorithm specified when the gain function (3.28) is placed in (3.1) defines an approximately-optimal estimator of θ. Theorem 3.1 and Corollary 3.3 show that this estimator asymptotically converges to θ with probability one. Thus, the algorithm provides an estimator that qualifies as a useful algorithm by satisfying in some sense, all three performance criteria in Chapter 1.

3.5 Summary

The continuous-time, parameter estimation problem has been considered in this chapter. The main results, Theorems 3.1 and 3.2, show that the sub-optimal estimators defined by (3.1) converge to the true value of the unknown parameter with probability one and in the mean-square sense.

A secondary, but still significant, result was obtained for the class of gain functions in (3.27). Corollary 3.3 demonstrates that this class of gains satisfies the hypotheses of Theorems 3.1 and 3.2. In an interesting result, this class is shown to contain function (3.28). When algorithm (3.1) employs the gain in (3.28), the algorithm is identical to the approximately-optimal scheme in (2.13)-(2.14). In other words, the approximately-optimal filter (2.13)-(2.14), if used for parameter estimation, qualifies as a useful algorithm satisfying requirements R.1, R.2, and R.3.

CHAPTER 4

CONVERGENCE INVESTIGATIONS FOR THE FILTERING PROBLEM

In Chapter 3, convergence theorems were developed for a class of sub-optimal estimators used for solving the continuous-time parameter estimation problem. The techniques used to prove these theorems can be applied in analyzing the behavior of certain sequential, sub-optimal, continuous-time filters. The main objective of this chapter is to formulate convergence theorems for the class of sub-optimal filters represented mathematically by (2.15).

Section 4.1 contains a list of the assumptions which are sufficient for convergence; basically, they allow the analytical tools and ideas described in Chapter 3 to be applied to the task of demonstrating the asymptotic convergence of time-domain filters.[5] The two convergence theorems are found in Sec. 4.2 and show that the output of the sub-optimal filter, (2.15), converges to the message as time becomes infinite. The convergence is both in the mean-square sense and with probability one.

Section 4.3 shows that, except for the most general assumptions on the message and observation models, the convergence theorems developed for the non-linear filtering problem, can be applied to the

---

[5] During the remainder of the thesis the phrases "asymptotic convergence of filters," "filter is shown to converge", etc. means that the difference between the filter output and message process converges to zero, in some probabilistic sense, as time increases indefinitely.

linear filtering problem and that the Kalman-Bucy filter is asymptotically convergent. This chapter concludes with Sec. 4.4 where properties of the message and observations models which guarantee the convergence of the approximately-optimal filtering solution developed in Chapter 2 are listed. The convergence ideas presented in this thesis have demonstrated that a non-linear filter presently being applied to real-world problems [30], [37], [7] is asymptotically convergent. This filter meets the three requirements enumerated in Chapter 1.

The demonstration of a convergent, approximately-optimal filter illustrates only one type of result which can be realized by using the mathematical techniques considered in this chapter. The convergence theorems exhibit the applicability of Ito calculus and martingale theory to the error analysis of sequential filtering algorithms. The approach, theoretical proofs, and conclusions have originated with this thesis investigation.

4.1 Assumptions and Preliminary Derivations

This section lists the assumptions for the convergence theorems to be proved in Sec. 4.2 and makes preliminary calculations on the error equation.

Section 3.1 dealt with the parameter estimation problem. Assumptions 1 through 5 concerned the behavior of the non-linear function in the observation process, $h(t,x)$ in (2.4), and the behavior of the gain function in the sub-optimal estimator, (3.1). Assumptions of this type are retained in the filtering solutions suggested in this chapter. In addition, conditions are placed on the non-linear function defining the message process, $f(t,x)$ in

(2.1). Assumptions are also made involving a combination of the gain function, (2.15), and both $h(t,x)$ and $f(t,x)$. All assumptions used in Theorems 4.1 and 4.2 are listed below. The first two assumptions are the same as 1, 2 and 3 in Chapter 3.

<u>Assumption B.1</u>: $h(t,x)$ is monotone and differentiable in x for every t.

<u>Assumption B.2</u>: $a(t,\bar{x}_{t_o,t})$ has the same sign as $\dot{h}(t,x)$ for every t and is measurable on $F\{[t_o,t], \bar{x}(s)\}$, the σ-algebra generated by $[\bar{x}(s): t_o \le s \le t]$.

The remainder of the assumptions are given in terms of the following simplified notation.

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

$$b(t) = \inf_x |\dot{h}(t,x)|$$

$$\phi(t,t_o) = \exp[\int_{t_o}^{t} G(s)ds]$$

$$a(t) = a(t,\bar{x}_{t_o,t})$$

$$\inf|a(t)| = \inf_{X \in C_{t_o}^t} |a(t,X)|$$

$$\sup|a(t)| = \sup_{X \in C_{t_o}^t} |a(t,X)|$$

where $C_{t_o}^t$ is the space of continuous functions on $[t_o,t]$.

<u>Assumption B.3</u>: $f(t,x)$ is differentiable in x for every t.

<u>Assumption B.4</u>: a) $\lim_{T\to\infty} \sup_{t \ge T} \phi^2(t,t_o) = K < \infty$

b) $\int_{t_o}^{\infty} b(s)\inf|a(s)|ds = \infty$

c) $\int_{t_o}^{\infty} \phi^2(t_o,t)\sup|a(t)|^2 dt < \infty$

<u>Assumption B.4'</u>: a) $\lim_{t\to\infty} \frac{\phi(t,t_o)^2}{\exp[2\int_{t_o}^{t} b(s)\inf|a(s)|ds]} = 0$

b) $\int_{t_o}^{\infty} \sup|a(t)|^2 dt < \infty$

Before the two convergence theorems are stated, a basic equation involving the filter error is developed from (2.1) and (2.4). The error at time t is denoted by $e(t) = \bar{x}(t) - x(t)$. The following differential equation can be written from (2.1), (2.4), and (2.15).

$$de(t) = [f(t,\bar{x}(t)) - f(t,x(t))]dt + a(t)[h(t,x(t)) - h(t,\bar{x}(t))]dt + a(t)dB(t)$$

If $Z(t,e) = e^2$, then Ito's lemma (Appendix B) gives an equation for $e(t)^2$:

$$de(t)^2 = 2\,e(t)[f(t,\bar{x}(t)) - f(t,x(t))]dt + 2\,e(t)a(t)[h(t,x(t)) - h(t,\bar{x}(t))]dt + a(t)^2 dt + 2\,e(t)a(t)dB(t)$$

The mean-value theorem gives

$$f(t,x(t)) = f(t,\bar{x}(t)) + \dot{f}(t,\sigma(t))[x(t) - \bar{x}(t)]$$

$$h(t,x(t)) = h(t,\bar{x}(t)) + \dot{h}(t,\theta(t))[x(t) - \bar{x}(t)]$$

where $\sigma(t)$ and $\theta(t)$ are both in between $x(t)$ and $\bar{x}(t)$. The equation for the error squared becomes:

$$de(t)^2 = 2\,e(t)^2[\dot{f}(t,\sigma(t)) - a(t)\dot{h}(t,\theta(t))]dt + a(t)^2 + 2e(t)a(t)dB(t) \tag{4.1}$$

Introducing a new variable $y(t) = \phi(t_o,t)^2 e(t)^2$, (4.1) leads to the following differential equation.

$$dy(t) = \phi(t_o,t)^2 de(t)^2 + e(t)^2 \frac{\partial}{\partial t}\phi(t_o,t)^2 dt$$

$$= 2\,e(t)^2\phi(t_o,t)^2[f(t,\sigma(t)) - G(t)]dt - 2\,\phi(t_o,t)^2 a(t)\dot{h}(t,\theta(t))e(t)^2 dt$$

$$+ \phi(t_o,t)^2 a(t)^2 dt + 2\,\phi(t_o,t)^2 e(t)a(t)dB(t)$$

The definition of $G(t)$ and Assumptions B.1 and B.2 lead to the following inequality for $y(t)$.

$$y(t) \le y(r) - \int_r^t 2|a(s)||\dot{h}(s,\theta(s))|y(s)ds + \int_r^t \phi(t_o,s)^2 \sup|a(s)|^2 ds$$

$$+ \int_r^t \phi(t_o,s)^2 e(s)a(s)dB(s) \;; \qquad t,r \ge t_o \qquad (4.2)$$

Equation (4.2) is the basic inequality used in the proofs of Theorems 4.1 and 4.2.

4.2 The Convergence Theorems

In the previous section, (4.2) was derived under Assumptions B.1-B.3. This equation is significant because it displays certain characteristics of the errors produced by the filters defined in (2.15) and because it is in a form which allows application of the convergence techniques developed in Chapter 3. The way in which the ideas of Secs. 3.2 and 3.3 provide two convergence theorems for the filters in (2.15) is shown in this section.

The two convergence theorems indicate the potential of using stochastic approximation ideas for analyzing continuous-time sequential filters. Theorem 4.1 demonstrates the asymptotic convergence of the filters in (2.15) with probability one and in the mean-square sense when (2.1), (2.4), and the gain functions in (2.15) satisfy B.1, B.2, B.3, and B.4. Theorem 4.2 provides mean-square convergence under assumptions B.1, B.2, B.3, and B.4'.

Theorem 4.1 Let $y(t) = \frac{dz(t)}{dt}$ be the observed process defined by (2.4) and let $x(t)$ be the message process defined by (2.1). Then, if Assumptions B.1 through B.3 and B.4 are satisfied, the output $\bar{x}(t)$ of the filter represented by

$$d\bar{x}(t) = f(t,\bar{x}(t))dt + a(t,\bar{x}_{t_o,t})[dz(t) - h(t,\bar{x}(t))dt]; \ \bar{x}(t_o) \text{ arbitrary} \tag{2.15}$$

converges to $x(t)$ with probability one and in the mean-square sense.

Proof. The inequality (4.2) can be weakened by noting that the second term on the Right Hand Side is always negative.

$$y(t) \le y(r) + \int_r^t \phi(t_o,s)^2 \sup|a(s)|^2 ds + \int_r^t \phi(t_o,s)^2 e(s)a(s)dB(s)$$

Letting $w(t) = y(t) + \int_t^\infty \phi(t_o,s)^2 \sup|a(s)|^2 ds$ gives

$$w(t) \le w(r) + \int_r^t 2\, \phi(t_o,s)^2 e(s)a(s)dB(s) \tag{4.3}$$

Equation (4.3) satisfies all of the hypotheses of Lemma 3.1, so $w(t)$ is a positive super-martingale. The previous chapter stated that the expected value of a super-martingale is non-increasing so that

$$\lim_{t\to\infty} Ew(t) = E\zeta = K_1 < \infty$$

and

$$\lim_{t\to\infty} Ey(t) = E\zeta$$

Returning to (4.2) and using the inequality $\inf|a(t)|b(t) \le |a(t)||\dot{h}(t,\theta(t))|$ provides another inequality for $y(t)$.

$$y(t) \le y(r) - 2\int_r^t \inf|a(s)|b(s)y(s)ds + \int_r^t \phi(t_o,s)^2 \sup|a(s)|^2 ds$$
$$+ \int_r^t \phi(t_o,s)^2 e(s)a(s)dB(s) \tag{4.4}$$

The expectation operator is applied to (4.4) and the fact, derived in Lemma 3.1, that $E \int_r^t \phi(t_o,s)^2 e(s)a(s)dB(s) = 0$ is employed to obtain an equation for $Ey(t)$.

$$Ey(t) \le Ey(r) - 2\int_r^t \inf|a(s)|b(s)Ey(s)ds + \int_r^t \phi(t_o,s)^2 \sup|a(s)|^2 ds$$

Combining Assumption B.4(c) and the convergence of $Ey(t)$,

$$\int_r^\infty \inf|a(s)|b(s)Ey(s)ds < \infty \qquad \text{for every } r \ge t_o. \qquad (4.5)$$

If (4.5) is to be consistent with Assumption B.4(b),

$$\lim_{t\to\infty} Ey(t) = \lim_{t\to\infty} \phi(t_o,t)^2 e(t)^2 = 0.$$

Since, from Assumption B.4(a), $\limsup \phi(t,t_o)^2 = K < \infty$, the function $\phi(t_o,t)^2 = 1/\phi(t,t_o)^2$ must be non-zero for large t; thus

$$\lim_{t\to\infty} E\ e(t)^2 = 0.$$

The basic super-martingale inequality (Appendix B) demonstrates probability-one convergence. The process $w(t)$ has been shown to be a positive super-martingale so

$$\text{Prob}\{\sup_{r\ge u\ge t} w(u) > \epsilon\} < \frac{Ew(t)}{\epsilon}$$

The process $w(t)$ depends on $y(t)$ in such a way that

$$\text{Prob}\{\sup_{r\ge u\ge t} y(u) > \epsilon\} < \text{Prob}\{\sup_{r\ge u\ge t} w(u) > \epsilon\}$$

Combining Assumption B.4(c) and the mean-square convergence result,

$$\lim_{t\to\infty} y(t) = 0 \qquad \text{w.P.1.}$$

It has already been pointed out that $\phi(t_o,t)^2 > 0$ for large t so that

$$\lim_{t\to\infty} e(t) = 0 \qquad \text{w.P.1.}$$

The proof of Theorem 4.1 relied on the reasonable assumption that $\limsup \phi(t,t_o)^2 = K < \infty$. In Theorem 4.2, $\phi(t,t_o)^2$ is allowed to grow large as t approaches ∞. This theorem differs from Theorem 4.1 both in its list of assumptions (B.4 is replaced by B.4') and in its conclusions since only mean-square convergence is demonstrated.

<u>Theorem 4.2</u> Under Assumption B.1 through B.3 and B.4', the output $\bar{x}(t)$ of the filter represented by (2.15) converges in the mean-square sense to $x(t)$:

$$\lim_{t\to\infty} E(\bar{x}(t) - x(t))^2 = 0.$$

<u>Proof</u>. The inequality $\inf|a(s)|b(s) \le |a(s)||\dot{h}(s,\theta(s))|$ and the expectation of (4.2) show that

$$Ey(t) \le Ey(r) - \int_r^t 2\inf|a(s)|b(s)Ey(s)ds + \int_r^t \phi(t_o,s)^2 \sup|a(s)|^2 ds;$$

$$r,t \ge t_o.$$

In the proof of Theorem 3.2, a function which dominates $Ey(t)$ was shown to exist. Following that argument, the bound for $Ey(t)$ becomes

$$Ey(t) \le \bar{D}(t,t_o)Ey(t_o) + \int_{t_o}^t \bar{D}(t,s)\phi(t_o,s)^2 \sup|a(s)|^2 ds$$

where

$$\bar{D}(t,t_o) = \exp[-\int_{t_o}^t \inf|a(s)|b(s)ds].$$

Now, from the formulation of $y(t)$,

$$Ee(t)^2 \le \phi(t,t_o)^2\bar{D}(t,t_o)Ee(t_o)^2 + \phi(t,t_o)^2\int_{t_o}^{t} \bar{D}(t,s)\phi(t_o,s)^2 \sup|a(s)|^2 ds. \tag{4.6}$$

Assumption B.4'(a) guarantees that

$$\phi(t,t_o)^2\bar{D}(t,t_o) = \frac{\phi(t,t_o)^2}{\exp[2\int_{t_o}^{t} \inf|a(s)|b(s)ds]} \to 0 \quad \text{as} \quad t \to \infty . \tag{4.7}$$

Letting $D(t,t_o) = \phi(t,t_o)^2\bar{D}(t,t_o)$, (4.6) can be written as

$$Ee(t)^2 \le D(t,t_o)Ee(t_o)^2 + \int_{t_o}^{t} D(t,s)\sup|a(s)|^2 ds . \tag{4.8}$$

Equation (4.8) is the same as (3.25). Equation (4.7) shows that $D(t,s) \to 0$ as $t \to \infty$. Assumption B.4'(b) and the same argument as that used on (3.25) can be applied to (4.8), giving

$$\lim_{t\to\infty} Ee(t)^2 = 0.$$

4.3 A Convergence Theorem for the Linear Kalman-Bucy Filter

The two convergence theorems of Sec. 4.2 are self-contained and developed independently of any other studies into convergence of filtering solutions, except those of discrete-time stochastic approximation. To properly evaluate these two new theorems, other known filtering convergence theorems should be consulted. The only published convergence results in the area of continuous-time, sequential filters are for the optimal solution to the linear filtering problem, the problem defined in Sec. 1.2.1. For this optimal solution, the Kalman-Bucy filter, it is well known that the filter's error converges to zero in the mean square sense when the signal and observation models satisfy certain conditions.

One way to use the established linear convergence theorem for checking the quality of the non-linear theories of Sec. 4.2 is to see if the linear result can be obtained from the non-linear theorems when the non-linear functions are reduced to linear functions. This section proceeds on this basis, checking to see if the assumptions about the message and observation processes along with the Kalman-Bucy filter's gain function satisfy the hypotheses of Theorems 4.1 and 4.2.

The linear filtering problem uses the following definition for the signal and observed processes:

$$\frac{dx(t)}{dt} = F(t)x(t) \;;\quad x(t_o) = b \text{ is an unknown constant.} \quad (4.9)$$

$$dx(t) = H(t)x(t)dt + dB(t) \,. \quad (4.10)$$

Equations (4.9) and (4.10) are the linear cases of (2.1) and (2.4), respectively. When b is assumed to be Gaussian with a mean of zero and a variance of P, the optimal, minimum-variance, filtering solution, which is the Kalman-Bucy filter, is represented by

$$d\hat{x}(t) = F(t)\hat{x}(t) + P(t)H(t)[dx(t) - H(t)\hat{x}(t)] \;;\; \hat{x}(t_o) = 0 \quad (4.11)$$

$$\frac{dP(t)}{dt} = 2\,F(t)P(t) - [H(t)P(t)]^2 \qquad ;\; P(t_o) = P \,. \quad (4.12)$$

In (4.11) and (4.12), $\hat{x}(t)$ is the filter output at time t and $P(t) = E(\hat{x}(t) - x(t))^2$.

Theorem 4.3 describes the asymptotic behavior of the Kalman-Bucy filter.

<u>Theorem 4.3</u> Let $\frac{\partial}{\partial t}\psi(t,t_o) = F(t)\psi(t,t_o)$ with $\psi(t_o,t_o) = 1$, that is,

$$\psi(t,t_o) = \exp[\int_{t_o}^{t} F(s)ds].$$

If

$$\lim_{t\to\infty} \frac{\psi(t,t_o)^2}{\int_{t_o}^{t} \psi(s,t_o)^2 H(s)^2 ds} = 0$$

then $$\lim_{t\to\infty} E(\hat{x}(t) - x(t))^2 = 0.$$

Proof. The solution to (4.12) is

$$P(t) = \frac{\psi(t,t_o)^2}{P^{-1} + \int_{t_o}^{t} \psi(s,t_o)^2 H(s)^2 ds} .$$

Since

$$\frac{1}{P^{-1} + \int_{t_o}^{t} \psi(s,t_o)^2 H(s)^2 ds} \le \frac{1}{\int_{t_o}^{t} \psi(s,t_o)^2 H(s)^2 ds} \quad \text{for} \quad t > t_o$$

it follows from the hypotheses that

$$0 \le P(t) \le \frac{\psi(t,t_o)^2}{\int_{t_o}^{t} \psi(s,t_o)^2 H(s)^2 ds} \to 0 \quad \text{as} \quad t \to \infty .$$

This theorem is analogous to the multi-dimensional discrete-time linear filtering result by Aoki [16] and Sorenson [43].

The object of this section is to see if the non-linear convergence Theorems 4.1 and 4.2 imply the convergence of the Kalman-Bucy filter from the hypotheses of Theorem 4.3. The function $\phi(t,t_o)^2 = \psi(t,t_o)^2$ (since $G(t) = F(t)$) may become unbounded so Theorem 4.2 is checked to see if the gain of the Kalman-Bucy filter satisfies its hypotheses.

Assumption B.1 through B.3 are satisfied by inspection, so only Assumption B.4' need be examined. The gain function for the Kalman-Bucy filter is

$$a(t) = P(t)H(t) = \frac{\phi(t,t_o)^2 H(t)}{[P^{-1} + \int_{t_o}^{t} \phi(s,t_o)^2 H(s)^2 ds]} .$$

Define the function

$$m(t) = \frac{\phi(t,t_o)^2}{\exp[2 \int_{t_o}^{t} b(s)\inf|a(s)|ds]}$$

where

$$b(t)\inf a(t) = \frac{H(t)^2 \phi(t,t_o)^2}{[P^{-1} + \int_{t_o}^{t} \phi(s,t_o)^2 H(s)^2 ds]}.$$

Letting the denominator of $m(t)$ be

$$u(t) = \frac{1}{P} \exp\left[2 \int_{t_o}^{t} \frac{H(s)^2 \phi(s,t_o)^2}{[P^{-1} + \int_{t_o}^{s} \phi(u,t_o)^2 H(u)^2 du]} ds \right]$$

produces the differential equation

$$\frac{du(t)}{dt} = \frac{2\, \phi(t,t_o)^2 H(t)^2 u(t)}{[P^{-1} + \int_{t_o}^{t} \phi(s,t_o)^2 H(s)^2 ds]} \quad ; \quad u(t_o) = \frac{1}{P}.$$

The function $w(t) = [P^{-1} + \int_{t_o}^{t} \phi(s,t_o)^2 H(s)^2 ds]^2$ satisfies the differential equation,

$$\frac{dw(t)}{dt} = \frac{2\, \phi(t,t_o)^2 H(t)^2 w(t)}{[P^{-1} + \int_{t_o}^{t} \phi(s,t_o)^2 H(s)^2 ds]} \quad ; \quad w(t_o) = \frac{1}{P}.$$

Since $w(t)$ and $u(t)$ satisfy the same linear differential equation with identical initial conditions, they are equal; i.e., $w(t) = u(t)$. Now, to examine Assumption B.4', this equality is used in $m(t)$.

$$m(t) = \frac{\phi(t,t_o)^2}{\exp[2 \int_{t_o}^{t} b(s)\inf|a(s)|ds]} = \frac{\phi(t,t_o)^2}{P[P^{-1} + \int_{t_o}^{t} \phi(s,t_o)^2 H(s)^2 ds]^2}$$

The hypothesis in Theorem 4.3 is sufficient for

$$\lim_{t\to\infty} \frac{\phi(t,t_o)^2}{\exp[2\int_{t_o}^{t} b(s)\inf|a(s)|ds]} = 0 .$$

The assumption made for (4.9) and (4.10) in Theorem 4.3 suffices for making the gain in the Kalman-Bucy filter satisfy Assumption B.4' (a). Since

$$\int_{t_o}^{\infty} \sup|a(s)|^2 ds = \int_{t_o}^{\infty} \frac{\phi(s,t_o)^4 H(s)^2 ds}{[P^{-1} + \int_{t_o}^{s} \phi(u,t_o)^2 H(u)^2 du]^2}$$

the hypotheses of Theorem 4.3 do not guarantee that $a(t)$ satisfy Assumption B.4'(b). Thus, Theorem 4.2 cannot be used to show that the Kalman-Bucy filter converges since the gain function may not be square-integrable.

Conclusions about convergence cannot be drawn from the non-linear theorems if $\phi(t,t_o)^2$ is unbounded for large t, so the hypotheses of Theorem 4.3 will be broken into three separate cases with one case containing the unbounded $\phi(t,t_o)^2$ condition.

<u>Case 1</u>: a) $\limsup \phi(t,t_o)^2 = K < \infty$

b) $\lim_{t\to\infty} \int_{t_o}^{t} \phi(s,t_o)^2 H(s)^2 ds = \infty$

<u>Case 2</u>: a) $\lim_{t\to\infty} \phi(t,t_o)^2 = 0$

b) $\lim_{t\to\infty} \int_{t_o}^{t} \phi(s,t_o)^2 H(s)^2 ds = K_1 < \infty$

<u>Case 3</u>: a) $\limsup \phi(t,t_o)^2 = \infty$

b) $\dfrac{\phi(t,t_o)^2}{\int_{t_o}^{t} \phi(s,t_o)^2 H(s)^2 ds} \to 0$ as $t \to \infty$

If f(t,x) and h(t,x) satisfy case 1, case 2, or case 3, then the functions satisfy the hypotheses of Theorem 4.3. Conversely, if the non-linear functions agree with the hypotheses of Theorem 4.3, they fit into one of the above cases. Thus, the three cases are equivalent to the hypotheses.

Since Theorem 4.1 and 4.2 cannot handle case 3, only the other two cases will be considered. Case 2 does satisfy the conditions of Theorem 4.2. Assumption B.4'(a) was examined above and now Assumption B.4'(b) is obeyed since

$$\int_{t_o}^{\infty} \sup|a(t)|^2 dt = \int_{t_o}^{\infty} \frac{\phi(t,t_o)^4 H(t)^2}{[P^{-1} + \int_{t_o}^{t} \phi(s,t_o)^2 H(s)^2 ds]^2} dt$$

$$\leq \frac{1}{P^{-2}} \int_{t_o}^{\infty} \phi(t,t_o)^4 H(t)^2 dt$$

and with $\phi(t,t_o)^2 < 1$ for $t >$ some $T > t_o$

$$\int_{t_o}^{\infty} \sup|a(t)|^2 dt \leq \frac{1}{P^{-2}} \int_{T}^{\infty} \phi(t,t_o)^2 H(t)^2 dt + \frac{1}{P^{-2}} \int_{t_o}^{T} \phi(t,t_o)^4 H(t)^2 dt$$

$$\leq K_1 + \frac{1}{P^{-2}} \int_{t_o}^{T} \phi(t,t_o)^4 H(t)^2 dt.$$

Theorem 4.1 handles case 1. First, Assumption B.4(c) requires $\int_{t_o}^{\infty} \phi(t_o,s)^2 \sup|a(s)|^2 ds$ be integrable. This follows directly when part (b) of case 1 and Lemma 3.2 are applied to the integral,

$$\int_{t_o}^{\infty} \phi(t_o,t)^2 \sup|a(t)|^2 dt = \int_{t_o}^{\infty} \frac{\phi(t,t_o)^2 H(t)^2}{[P^{-1} + \int_{t_o}^{t} \phi(s,t_o)^2 H(s)^2 ds]^2} dt.$$

Furthermore, for Assumption B.4(b), part b of case 1 and Lemma 3.2 show that

$$\int_{t_o}^{\infty} b(s) \inf|a(s)| ds = \int_{t_o}^{\infty} \frac{\phi(s,t_o)^2 H(s)^2}{[P^{-1} + \int_{t_o}^{s} \phi(u,t_o)^2 H(u)^2 ds]} ds = \infty.$$

Consequently, Theorem 4.1 and 4.2 show that the optimal, Kalman-Bucy solution to the linear filtering problem is asymptotically convergent in the mean-square sense if (4.9) and (4.10) satisfy the hypotheses of Theorem 4.3 and $\lim\sup \phi(t,t_o)^2 < \infty$. This is the first time that restrictions have been placed on (4.9) and (4.10) which guarantee the convergence of the optimal linear filter with probability one.

4.4 A Class of Gain Functions Guaranteeing Convergence

This section returns to the filters represented by (2.15) which were proposed as solutions to the non-linear filtering problem. In Sec. 4.2, these sub-optimal filters were shown to converge when the models of the message and observation processes, (2.1) and (2.4), and the gain function, $a(t,\bar{x}_{t_o,t})$, satisfied Assumptions B.1 - B.3 and B.4 or B.4'. The gain function does not have any fixed structure and only has to fulfill conditions such as: $\int_{t_o}^{\infty} b(t)\inf|a(t)|dt = \infty$ and $\int_{t_o}^{\infty} \phi(t_o,t)^2 \sup|a(t)|^2 dt < \infty$. In this section, a class of gain functions is formulated and more assumptions are made about (2.1) and (2.4). The assumptions and Theorem 4.1 imply that the sub-optimal filters converge with probability one and in the mean-square sense.

The following assumptions are made about (2.1) and (2.4):

Assumption B.1, B.2, B.3, and

Assumption C.1: $\sup_t \frac{g(t)}{b(t)} < \infty$ where $g(t) = \sup_x |\dot{h}(t,x)|$

Assumption C.2: Let $G(t) \geq \dot{f}(t,x) \geq D(t)$ for $t \geq t_o$ and for every x and define

$$\phi(t,t_o) = \exp\left[\int_{t_o}^{t} G(s)ds\right]$$

$$\psi(t,t_o) = \exp\left[\int_{t_o}^{t} D(s)ds\right];$$

then $\sup_t \frac{\phi(t,t_o)^2}{\psi(t,t_o)^2} < \infty$ and $\limsup \phi(t,t_o)^2 = K < \infty$.

<u>Assumption C.3</u>: $\int_{t_o}^{\infty} \phi(s,t_o)^2 b(s)^2 ds = \infty$

<u>Corollary 4.4</u> If (2.1) and (2.4) satisfy Assumptions B.1 - B.3 and C.1 - C.3 and if the gain function $a(t,\bar{x}_{t_o},t)$ has the form

$$|a(t,\bar{x}_{t_o},t)| = \frac{g_1(t,\bar{x}_{t_o},t)c_1(t,\bar{x}(t))}{[a + \int_{t_o}^{t} g_2(s,\bar{x}_{t_o},s)c_2(s,\bar{x}(s))ds]} \quad (4.13)$$

where

$$\psi(t,t_o)^2 \le g_1(t,\bar{x}_{t_o},t),\ g_2(t,\bar{x}_{t_o},t) \le \phi(t,t_o)^2$$

$$b(t) \le c_1(t,\bar{x}(t)),\ c_2(t,\bar{x}(t)) \le g(t)$$

then $a(t,\bar{x}_{t_o},t)$, (2.1) and (2.4) satisfy Assumption B.4.

<u>Proof</u>. Assumption B.4(a) is satisfied by inspection. Examining the integrand of the integral in Assumption B.4(b),

$$b(t)\inf|a(t)| \ge \frac{\psi(t,t_o)^2 b(t)^2}{[a + \int_{t_o}^{t} \phi(s,t_o)^2 g(s)^2 ds]} .$$

By Assumption C.2, there exists a K_1 and K_2 such that for every t

$$1 \le \frac{\phi(t,t_o)^2}{\psi(t,t_o)^2} \le K_1 \qquad \text{and} \qquad 1 \le \frac{g(t)}{b(t)} \le K_2 .$$

These inequalities weaken the lower bound on the integrand to

$$b(t)\inf|a(t)| \ge \frac{\phi(t,t_o)^2 b(t)^2}{K_1 K_2^2 [a' + \int_{t_o}^{t} \phi(s,t_o)^2 b(s)^2 ds]} .$$

Using Assumption C.3 and Lemma 3.2,

$$\int_{t_o}^{\infty} b(t)\inf|a(t)|dt = \infty .$$

Assumption B.4(c) is now investigated.

$$\phi(t,t_o)^2 \sup|a(t)|^2 \le \frac{\phi(t_o,t)^2 \phi(t,t_o)^4 g(t)^2}{[a + \int_{t_o}^{t} \psi(s,t_o)^2 b(s)^2 ds]^2}$$

$$\le \frac{K_1^4 K_2^2 \phi(t,t_o)^2 b(t)^2}{[a'' + \int_{t_o}^{t} \phi(s,t_o)^2 b(s)^2 ds]^2}$$

Again, Assumption C.3 and Lemma 3.2 show that

$$\int_{t_o}^{\infty} \phi(t_o,t)^2 \sup|a(t)|^2 dt < \infty,$$

completing the proof.

Except for the discussions of the linear Kalman-Bucy filter in Sec. 4.3, this chapter has concentrated on demonstrating the asymptotic convergence of general sub-optimal, sequential filters. One reason this investigation analyzed sub-optimal filters such as (2.15) was to obtain theorems which provide analytical properties for optimal, non-linear filters. With this motivation, the ideas of stochastic approximation exhibited by Theorems 4.1 and 4.2 were specialized to produce the corollary in this section, Corollary 4.4. This corollary is now used to show that the approximately-optimal filter represented by (2.13)-(2.14) is asymptotically convergent.

Define the gain term in (2.15) as

$$a(t,\bar{x}_{t_o,t}) = P(t)\dot{h}(t,\bar{x}(t)) \qquad (4.14)$$

with

$$P(t) = \frac{\Phi(\bar{x}_{t_o,t})}{[\frac{1}{var(b)} + \int_{t_o}^{t} \Phi(\bar{x}_{t_o,s})\dot{h}(s,\bar{x}(s))^2 ds]}$$

and $$\Phi(\bar{x}_{t_o},t) = \exp[2 \int_{t_o}^{t} \dot{f}(s,\bar{x}(s))ds] .$$

The function $P(t)$ satisfies the differential equation (2.14) since

$$\frac{dP(t)}{dt} = 2\ \dot{f}(t,\bar{x}(t))P(t) - [\dot{h}(t,\bar{x}(t))P(t)]^2 \ ; \quad P(t_o) = \text{var}(b).$$

The gain term $a(t,\bar{x}_{t_o},t)$ is the gain term in (2.13) and also satisfies (4.13) in Corollary 4.4. Therefore, Theorem 4.1 combined with Corollary 4.4 show that the approximately-optimal filter represented by (2.13) and (2.14) converges asymptotically with probability one and in the mean-square sense if (2.1) and (2.4) satisfy Assumptions B.1, B.3, C.1, C.2, and C.3 and if the initial condition on (2.1), $x(t_o) = b$, has a prior distribution. The convergence theorems of this chapter have provided the means for displaying, under certain conditions, a filtering solution which meets the three requirements of a useful filter; optimality, convergence, and sequential structure.

4.5 Summary

The first major contribution of this chapter is the statement and proof, in Sec. 4.2, of two theorems demonstrating the asymptotic convergence of the non-linear filters defined by (2.15). Theorem 4.1 shows that the filtering error goes to zero both with probability one and in the mean-square sense, with limitations placed on (2.1) and (2.4) and on the gains $a(t,\bar{x}_{t_o},t)$. One prime restriction in the hypotheses of Theorem 4.1 is that the message process is bounded; i.e., $\lim\sup \phi(t,t_o)^2 < \infty$. Theorem 4.2 places other conditions on the gain functions and allows the message process to become unbounded yet shows that the output of the sub-optimal filter converges in the mean-square sense to the message.

To compare Theorems 4.1 and 4.2 to known convergence results, a convergence theorem for the optimal, Kalman-Bucy solution to the linear filtering problem is given in Sec. 4.3 as a special case of these theorems. Theorem 4.1 and 4.2 prove the convergence of the Kalman-Bucy filter when the message process is bounded and when the functions used in the mathematical models of the message and observation processes, (2.1) and (2.4), guarantee the divergence of a certain integral. Thus, Theorems 4.1 and 4.2 imply convergence of the optimal linear solution but under restricted hypotheses. Section 4.3 produces a minor original result in the linear theory; Theorem 4.1 demonstrates convergence of the Kalman-Bucy filter with probability one, a type of convergence that has not been proved in the literature.

In Sec. 4.4, additional assumptions were made about (2.1) and (2.4). Corollary 4.4 showed that Assumptions B.1 - B.3 and C.1 - C.3, when combined with any of the gain functions specified by (4.13), satisfy the hypotheses of Theorem 4.1. If any of the gain functions given in (4.13) is used in (2.15) and if (2.1) and (2.4) fulfill the corollary's assumptions, the filter defined by (2.15) converges, as time increases, to the message process.

The second major contribution of this chapter concerns the convergence of an approximately-optimal non-linear filter. The gain function given in (4.14) satisfies the conditions in Corollary 4.4 and, when placed in (2.15), defines a filter identical to the approximately-optimal filter derived in Chapter 2. The combination of Assumptions B.1 - B.3 and C.1 - C.3, Theorem 4.1, and Corollary 4.4 shows that an approximately-optimal filter is asymptotically convergent. Under the specified restrictions, this filter, given by

(2.13)-(2.14), provides nearly-minimum-variance estimates of the message throughout the period of observation and, furthermore, is guaranteed to converge to the message after lengthy processing of the observations. The filter satisfies all three requirements of a useful filtering solution as discussed in Chapter 1.

This thesis provides the first study into the asymptotic convergence of filters for the continuous-time, non-linear filtering problem. The style of analysis and its associated theoretical tools used to prove Theorems 4.1 and 4.2 appear to provide logical means for probing deeper into the behavior of non-linear, sequential filters.

CHAPTER 5

AN EXAMPLE OF A FILTERING PROBLEM

This chapter considers an example of the filtering problem defined in Sec. 2.1. The filter uses the approximately-optimal algorithm discussed in Sec. 4.4 and is shown to satisfy the hypotheses of Corollary 4.4 so that the filter is asymptotically convergent. This convergence is verified when the message process, observation process, and filter are simulated on a digital computer.

5.1 The Example and Computer Simulation

The filtering problem discussed in this chapter is defined as:

$$\dot{x}(t) = f(t,x) \ ; \qquad x(0) = b; \ t_o = 0 \tag{2.1}$$

$$y(t) = h(t,x) + v(t) \tag{2.2}$$

where

$$f(t,x) = \frac{\sin(1000\ t)}{100\ \pi\ t}\ x + \frac{3.0}{\left[1 + \left\{\frac{x}{(t+1)^2}\right\}^2\right]} \tag{5.1}$$

$$h(t,x) = 0.5\ x + 5\ \tan^{-1} x \tag{5.2}$$

and b is a Gaussian random variable with a mean = -1.5 and a variance = 1.0.

The approximately-optimal filter derived in Sec. 2.2 and represented by (2.13)-(2.14) is used to process $y(t)$. This filter is asymptotically convergent if (5.1) and (5.2) satisfy Assumptions B.1, B.3, C.1, C.2, and C.3. These assumptions are now examined. With

$h(t,x)$ being monotone in x, Assumptions B.1 and B.3 are verified directly. Assumption C.1 is satisfied as shown below.

$$\dot{h}(t,x) = 0.5 + \frac{5}{1+x^2}$$

$$g(t) = \sup_x |\dot{h}(t,x)| = 5.5 \quad \text{and}$$

$$b(t) = \inf_x |\dot{h}(t,x)| = 0.5$$

Also

$$\dot{f}(t,x) = \frac{\sin(1000\ t)}{100\ \pi\ t} - \frac{6.0\ x}{(t+1)^4[1 + \{\frac{x}{(t+1)^2}\}^2]^2}$$

which implies that

$$G(t) = \frac{\sin(1000\ t)}{100\ \pi\ t} + \frac{3.0}{(t+1)^2}$$

$$D(t) = \frac{\sin(1000\ t)}{100\ \pi\ t} - \frac{3.0}{(t+1)^2}$$

and $$\phi(t,t_o) = \exp[\int_0^t \frac{\sin(1000\ s)}{100\ \pi\ s}\, ds + \int_0^t \frac{3.0}{(s+1)^2}\, ds]$$

$$\psi(t,t_o) = \exp[\int_0^t \frac{\sin(1000\ s)}{100\ \pi\ s}\, ds - \int_0^t \frac{3.0}{(s+1)^2}\, ds]$$

Therefore

$$\lim_{t\to\infty} \phi(t,t_o)^2 = K < \infty ,$$

$$\sup_t \frac{\phi(t,t_o)^2}{\psi(t,t_o)^2} = \sup_t \exp[2 \int_0^t \frac{3.0}{(s+1)^2}] < \infty ,$$

and $$\int_0^\infty \phi(s,t_o)^2 b(s)^2 ds = (0.5)^2 \int_0^\infty \phi(s,t_o)^2 ds = \infty .$$

Thus the non-linearities (5.1) and (5.2) also satisfy Assumption C.2 and C.3, demonstrating the asymptotic convergence of the filter.

A digital computer has been used to simulate the performance of this filter. Equations (2.1), (2.13), and (2.14) were approximated on the computer by the following difference equations.

$$x(t_{i+1}) = x(t_i) + f(t_i, x(t_i))\Delta t \;;\quad x(0) = b$$

$$\hat{x}(t_{i+1}) = \hat{x}(t_i) + f(t_i, \hat{x}(t_i))\Delta t + P(t_i)\dot{h}(t_i, \hat{x}(t_i))[h(t_i, x(t_i)) - h(t_i \hat{x}(t_i))]\Delta t$$

$$+ P(t_i)\dot{h}(t_i, \hat{x}(t_i))\Delta B_i \;;\quad \hat{x}(0) = -1.5$$

$$P(t_{i+1}) = P(t_i) + 2\, P(t_i)\dot{f}(t_i, \hat{x}(t_i))\Delta t$$

$$- [P(t_i)\dot{h}(t_i, \hat{x}(t_i))]^2 \Delta t \;;\quad P(0) = 1.0$$

Here, $\Delta t = t_i - t_{i-1}$ for $i \geq 1$ and $\{\Delta B_i\}$ are Gaussian random variables with mean zero and variance Δt for each $i \geq 0$.

The solution to these difference equations has been shown to converge to the actual solutions of (2.1), (2.13), and (2.14) in the mean-square sense as $\Delta t \to 0$ [44].

Using $\Delta t = 10^{-3}$ sec., 200 runs were performed on the computer. For each run, the initial condition for (2.1), b, was sampled from a Gaussian distribution and a new sequence of observation noises $\{\Delta B_i\}$ was selected. Using these 200 runs, the sample mean and variance of the error $e(t) = \hat{x}(t) - x(t)$ were calculated. In this simulation the filter does converge as time increases; Figures 2 and 3 display the results.

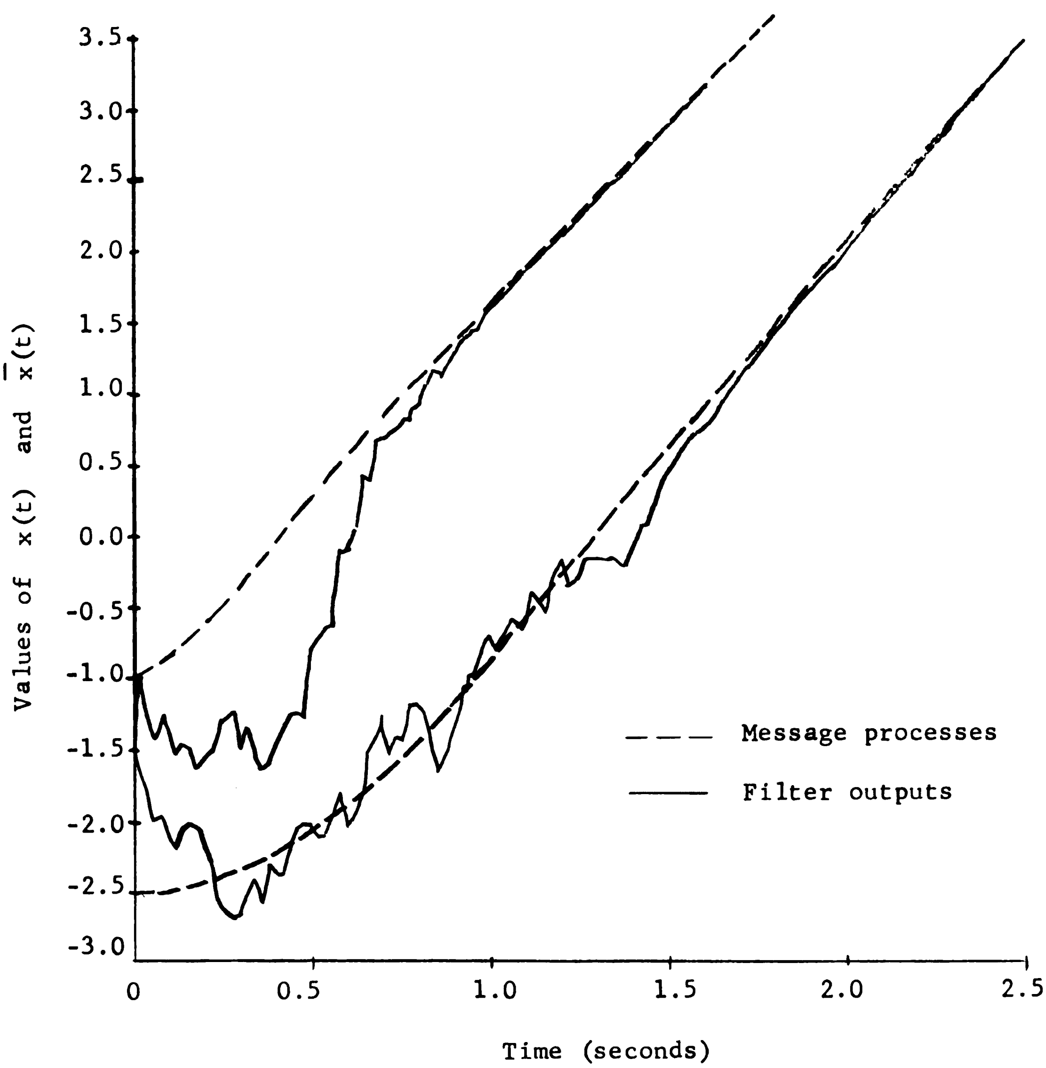

Figure 2 Filter Behavior for Two Runs

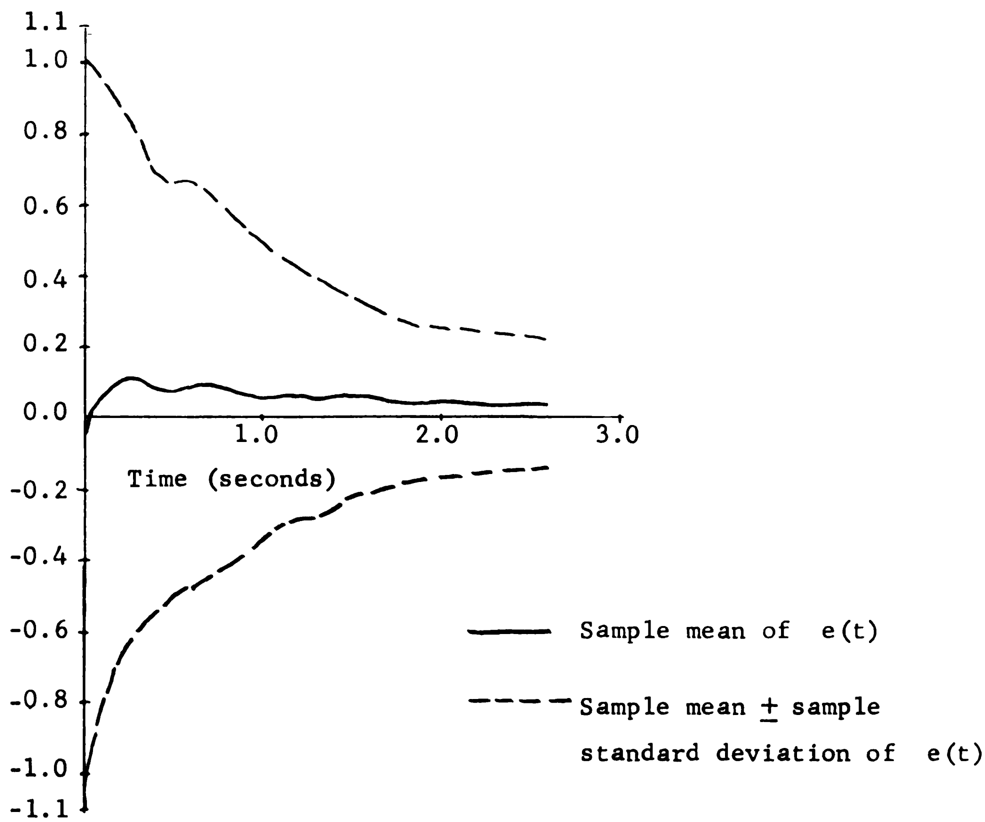

Figure 3 Statistical Analysis of Filter Performance

5.2 Summary

An approximately-optimal filter has been applied to a non-linear filtering problem which satisfies the conditions for asymptotic convergence given in Sec. 4.4. A computer simulation has verified the convergence.

CHAPTER 6

CONCLUSIONS

Section 6.1 outlines the major results of the thesis. Possible extensions of the investigations in this thesis are outlined in Sec. 6.2.

6.1 Conclusions and Results

This thesis has proved the existence, under certain conditions, of an approximately-optimal, asymptotically convergent, sequential, non-linear filter. This culminates an effort devoted to developing useful and practical filtering algorithms. The primary results are theorems which demonstrate the asymptotic convergence of sequential, non-linear algorithms both for filtering and parameter estimation.

After briefly describing the stochastic filtering problem and the characteristics of useful solutions, Chapter 1 discusses the need for analyzing the asymptotic behavior of non-linear, continuous-time filters. A literature review suggests that the method of stochastic approximation, a well-developed technique of classical, discrete-time statistics, could be applied in an investigation of the asymptotic properties of sequential filters.

In Chapter 2, a scalar version of the filtering problem is formulated in terms of non-linear stochastic differential equations. Modeling the problem in this manner allows a calculus developed by Ito to be applied to the performance analysis of filters. Chapter 2 also proposed the algorithms in (2.15) as sub-optimal solutions to the filtering problem. Besides possessing the general features of presently-documented time-domain filters, these algorithms have unstructured gain terms. This thesis marks the first time that continuous-time filters with arbitrary gain terms have been investigated.

Before pursuing an analysis of the asymptotic behavior of these filters, the method of stochastic approximation is extended to the continuous-time, parameter estimation problem in Chapter 3. The large-time behavior of a special version of the algorithms in (2.15) is investigated. A parameter of a signal is estimated where the signal is observed in the presence of additive, white noise. Theorems 3.1 and 3.2 are the most important results in this thesis. Conditions are imposed on the signal and the gains of estimators which guarantee convergence to the true parameter as time increases. The sequence of estimates converges with probability one and in the mean-square sense. These theorems are original and illustrate the use of Ito calculus and martingale theory in asymptotic convergence studies. Section 3.4 illustrates a way of effectively exploiting the arbitrary gain structure permitted in the theorems. A particular gain function is selected that leads to a nearly-optimal (in a Bayesian sense), convergent algorithm.

In Chapter 4, the filters (2.15) are considered. Two key theorems, 4.1 and 4.2, were developed from the ideas expounded in Chapter 3. These theorems show that when the message process, the observed process, and gain function in (2.15) satisfy certain hypotheses, the outputs of the filters in (2.15) converge to the message process as time grows large. Both probability-one and mean-square convergence are demonstrated. Section 4.3 examines the conclusions of Theorems 4.1 and 4.2 for the special case when the message and observation processes are generated by appropriate linear stochastic differential equations. Conditions are specified which guarantee that the Kalman-Bucy filter is asymptotically convergent with probability one.

In Sec. 4.4, a basic goal of this study is accomplished when Theorem 4.1 is used to demonstrate that an approximately-optimal, non-linear filter converges asymptotically, given that the message and observation processes satisfy certain conditions. That is, a filter is specified which provides nearly-minimum-variance estimates of the message process throughout the observation time interval and which converges surely to this message as time increases. This thesis contains the first investigation of the asymptotic convergence of non-linear, continuous-time, sequential filters and indicates a way in which stochastic approximation can be applied to the filtering problem.

An example of a filtering problem is given in Chapter 5. The approximately-optimal filter of Sec. 4.4 is applied as a filter and convergence is demonstrated by use of a digital computer simulation.

6.2 Extensions

There are two different directions in which the results of this thesis can be extended. Either the hypotheses of the convergence theorems can be weakened or these theorems can be extended to include the multi-dimensional filtering problem.

The convergence theorems in both Chapters 3 and 4 are for one-dimensional problems. These theorems may be extended to the multi-dimensional case, thus enabling them to be applicable to large scale engineering problems. Insight into this type of extension may be gained by consulting Albert and Gardner [20] for an application of the method of stochastic approximation to estimating, via discrete-time observations, a vector of parameters.

The analyses of Chapters 3 and 4 may be closely examined to see if less restrictive hypotheses can be placed on the convergence theorems. Albert and Gardner's analysis of constrained, discrete-time estimators offers a way of reducing the restrictions on the observed processes. Another possibility is to interconnect the concepts of Liapunov stability and the method of stochastic approximation in continuous-time. Introducing a Liapunov function for the differential equation representing the message process into the convergence ideas of this thesis may lead to asymptotic convergence results more effective than those in Chapter 4.

The approach exhibited in this thesis may also be applied to the general time-domain filtering problem defined in Sec. 1.1, i.e. where $u(t) \neq 0$ in (1.1). An investigation into the asymptotic behavior of filters for this type of problem should begin with an examination of the discrete-time results of Albert and Gardner discussed earlier in Sec. 1.2.7.

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Wiener, N., Extrapolation, Interpolation, and Smoothing of Stationary Time Series, John Wiley & Sons, Inc., New York, 1949.

2. Kushner, H.J., "Dynamical Equations for Optimal Nonlinear Filtering," Journal of Differential Equations, Vol. 3, pp. 179-190, 1967.

3. Bucy, R.S., "Nonlinear Filtering Theory," IEEE Trans Automatic Control, Vol. 10, pp. 198, 1965.

4. Kalman, R. and R.S. Bucy, "New Results in Linear Filtering and Prediction Theory," Trans. ASME ser. D, J. Basic Eng. 83, pp. 95-108, March 1961.

5. Deutsch, R., Estimation Theory, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1965.

6. Synder, D.L., "The State-Variable Approach to Continuous Estimation," Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, Mass., 1966.

7. Richman, J. and F. Thau, "Elements of the Kalman Filtering Technique," Technical News Bulletin, General Precision, Inc., Vol. 9, No. 4, 1966.

8. Fisher, J.R., "Conditional Probability Density Functions and Optimal Non-Linear Estimation," Ph.D. Thesis, University of California, Los Angeles, 1966.

9. Ho, Y.C. and R.C.K. Lee, "Identification of Linear Dynamic Systems," Proc. N.E.C., 1964.

10. Kalman, R.E., "Fundamental Study of Adaptive Control Systems," Tech. Rept. ACD-TR-61, Vol. 1, App. I, 1961.

11. Irwin, J.D. and J.B. O'Neal, Jr., "The Design of Optimum DPCM (Differential Pulse Code Modulation) Encoding Systems Via the Kalman Predictor," Proc. 1968 JACC.

12. Stratonovich, R.L., Conditional Markov Processes and Their Application to the Theory of Optimal Control, American Elsevier Publishing Co., Inc., New York, 1968.

13. Kallianpur, G. and C. Striebel, Stochastic Differential Equations Occurring in the Estimation of Continuous Parameter Stochastic Processes, Technical Report No. 103, University of Minnesota, Minneapolis, Minn., September 1967.

14. Kushner, H.J., "Approximations to Optimal Nonlinear Filters," IEEE Trans. Automatic Control, Vol. AC-12, pp. 546-556, October 1967.

15. Bucy, R.S., "Recent Results in Linear and Non-Linear Filtering," Stochastic Problems in Control, 1968 Joint Automatic Control Conference, University of Michigan, Ann Arbor, Mich., ASME, pp. 87-106, June 1968.

16. Aoki, M., Optimization of Stochastic Systems, Academic Press, New York, 1967.

17. Smith, G.L., "Multivariable Linear Filter Theory Applied to Space Vehicle Guidance," J. SIAM Control, Ser. A, Vol. 2, pp. 19-32, 1964.

18. Friedland, B. and I. Bernstein, "Estimation of the State of a Non-Linear Process in the Presence of Non-Gaussian Noise and Distrubances," J. of the Franklin Institute, Vol. 281, No. 6, pp. 455-480, June 1966.

19. McLean, J.D., S.F. Schmidt and L.A. McGree, Optimal Filtering and Linear Prediction Applied to a Space Navigation System for the Circumlunar Mission, NASA TND-1208, March 1962.

20. Albert, A.E. and L.A. Gardner, Jr., Stochastic Approximation and Nonlinear Regression, Research Monograph No. 42, the M.I.T. Press, Cambridge, Mass., 1967.

21. Pearson, J.B., "A Note on Nonlinear Filtering," IEEE Trans. Automatic Control, AC-13, No. 1, pp. 101-105, February 1968.

22. Wolverton, C.T., "A Note on the Application of Dvoretzky's Theorem of Non-linear Filtering," IEEE Trans. Automatic Control, AC-14, No. 1, pp. 117, February 1969.

23. Blaydon, C.C., "Recursive Algorithms for Pattern Classification," Ph.D. Thesis, Harvard University, Cambridge, Mass., 1967.

24. Donoghue, P.J., "System Identification by Bayesian Learning," Ph.D. Thesis, Michigan State University, East Lansing, Mich., 1968.

25. Driml, M. and J. Nedoma, "Stochastic Approximations for Continuous Random Processes," Trans of 2nd Prague Conf. on Inform. Theory, Statistical Decision Functions, and Random Processes, Prague, Czechoslovakia, pp. 145-158, June 1959.

26. Driml, M. and O. Hans, "Continuous Stochastic Approximations," Trans. of 2nd Prague Conf. on Inform. Th., Statist. Decision Functions, and Random Processes, Prague, Czechoslovakia, pp. 113-122, June 1959.

27. Sakrison, D.J., "A Continuous Kiefer-Worlfowitz Procedure for Random Processes," Ann. Math. Statist., Vol. 35, pp. 590-599, 1964.

28. Sakrison, D.J., "Stochastic Approximation: A Recursive Method for Solving Regression Problems," Advances in Communication Systems, Vol. 2, pp. 51-106, 1966.

29. Bass, R.W., V.D. Norum, and L. Schwartz, "Optimal Multichannel Nonlinear Filtering," J. of Math. Analysis and Application, Vol. 16, pp. 152-164, 1966.

30. Bellman, R.E., H.H. Kagiwada, R.E. Kalaba, and R. Sridhar, Invariant Imbedding and Nonlinear Filtering Theory, Rand Corporation Memo., RM-4374-PR, December 1964.

31. Kushner, H.J., "On the Differential Equations Satisfied by Conditional Probability Densities of Markov Processes with Applications," J. SIAM Control, Ser. A, Vol. 2, No. 1, pp. 106-119, 1964.

32. Stratonvich, R.L., "A New Representation for Stochastic Integrals and Equations," J. SIAM Control, Vol. 4, No. 2, 1966.

33. Coddington, E.A. and N. Levinson, Theory of Ordinary Differential Equations, McGraw-Hill, New York, 1955.

34. Kailath, T. and P. Frost, "Mathematical Modeling of Stochastic Processes," Stochastic Problems in Control, 1968 Joint Automatic Control Conference, University of Michigan, Ann Arbor, Mich., ASME, pp. 1-38, June 1968.

35. Wonham, W.M., "Some Applications of Stochastic Differential Equations to Optimal Nonlinear Filtering," J. SIAM Control, Ser. A, Vol. 2, No. 3, pp. 347-369, 1965.

36. Perfitt, T.E., Adaptive and Non-Adaptive Digital Phase Lock Loops, Lear Siegler, Inc. Engineering Memorandum Report No. 204, Grand Rapids, Mich., 1968.

37. Stear, E.B. and L. Schwartz, "A Computational Comparison of Several Nonlinear Filters," IEEE Trans. Automatic Control, AC-13, No. 1, pp. 83-86, February 1968.

38. Morozan, T., "Sur L'Approximation Stochastic," C.R. Acad. Sci., Vol. 264, Paris, pp. 633-635, 1967.

39. Kushner, H.J., "A Note on the Maximum Sample Excursions of Stochastic Approximation Processes," Ann. Math. Statist., Vol. 37, pp. 513-516, 1966.

40. Gladyshev, E.G., "On Stochastic Approximation," Theory Prob. Applications, Vol. 10, No. 2, pp. 275-278, 1965.

41. Doob, J.L., Stochastic Processes, John Wiley and Sons, Inc., New York, 1953.

42. Fulks, W., Advanced Calculus, John Wiley and Sons, Inc., New York, 1961.

43. Sorenson, H.W., "On the Error Behavior in Linear Minimum Variance Estimation Problems," IEEE Trans. Automatic Control, AC-12, No. 5, pp. 557-562, October 1967.

44. Maruyama, G., "Continuous Markov Processes and Stochastic Equations," Circolo Matematico Di Palermo, Series 2, Vol. 4, pp. 48-90, 1955.

45. Skorokhod, A.V., Studies in the Theory of Random Processes, Addison-Wesley Publishing Co., Inc., Reading, Mass., 1965.

46. Gel'fand, I.M. and N. Ya. Vilenkin, Generalized Functions, Vol. 4, Academic Press, New York, 1964.

47. Loeve, M., Probability Theory, D. Van Nostrand Co., Inc., Princeton, New Jersey, 1963.

APPENDICES

APPENDIX A

STOCHASTIC DIFFERENTIAL EQUATIONS AND THE ITO CALCULUS

Stochastic differential equations have only recently been used in engineering applications. The equations were initially misunderstood, creating a confusion about the proper analytical procedures for handling them. The purpose of this appendix is to paraphrase recent publications and briefly explain the available facts about stochastic differential equations.

The following equation defines the n vector $\underline{x}(t)$:

$$\frac{dx(t)}{dt} = \underline{c}(t) + \underline{D}(t)\underline{u}(t) \qquad (A.1)$$

where $\underline{u}(t)$ is a zero-mean, white noise vector process, i.e., $E\, u_i(t)u_j(s) = \delta(t-s)\delta_{ij}$, and $\underline{c}(t)$ and $\underline{D}(t)$ are, respectively, random vector and matrix functions of t, which may depend on the present and past values of $u(t)$. Equation (A.1) is a general engineering definition of a stochastic differential equation. The filtering error equations in this thesis are of this type.

To display the need for the strict attention necessary when dealing with stochastic differential equations, an example from Kailath [34] will be discussed. Consider the following system of equations involving Gaussian white noise $u(t)$.

$$\dot{B}(t) = u(t) \;; \quad B(0) = 0 \qquad (A.2)$$

$$\dot{x}(t) = B(t)u(t) \;; \quad x(0) = 0 \qquad (A.3)$$

The solution for $x(t)$ is:

$$x(t) = \int_0^t B(s)u(s)ds \tag{A.4}$$

where
$$B(t) = \int_0^t u(s)ds \tag{A.5}$$

Equation (A.4) can be written as:

$$x(t) = \int_0^t B(s)dB(s) \tag{A.6}$$

and formal integration gives:

$$x(t) = \frac{B(s)^2}{2}\Big|_0^t = B(t)^2/2. \tag{A.7}$$

As indicated, the integration of (A.6) was formal, following normal calculus rules. Equation (A.6) will now be examined by applying the fundamental formula for evaluating integrals:

$$x(t) = \lim_{|t_{i+1}-t_i|\to 0} \sum_{t_i} B(\rho_i)[B(t_{i+1}) - B(t_i)] \tag{A.8}$$

where $\{0, t_1, t_2, \ldots, t_{n-1}, t\}$ is a partition of $[0,t]$ and ρ_i is any point in the interval $[t_i, t_{i+1}]$.

Before continuing with this discussion, certain properties of the random process $B(t)$, which has the following characteristics, are discussed.

$$E\ B(t) = 0$$

$$E\ B(t)B(s) = \min(t,s)$$

The Gaussian process with these properties, called a Brownian motion process, has been studied extensively and has the following features: independent increments, Markov and martingale behavior, and continuity but non-differentiability for almost all sample functions. Related to this last feature, and very important in the analysis of the integral (A.6), is the Levy oscillation property:[6] If $\{0 = t_o, t_1, \ldots, t_n = t\}$ is a partition of $[0,t]$ then

$$\lim_{|t_{i+1}-t_i|\to 0} \sum_{t_i} [B(t_{i+1}) - B(t_i)]^2 = t \qquad (A.9)$$

where the limit exists w.P.1 and in the mean square sense.

To analyze (A.6), let $\rho_i = t_i$. Then (A.8) becomes, from (A.9):

$$\begin{aligned} x(t) &= \lim \Sigma\, B(t_i)[B(t_{i+1}) - B(t_i)] \\ &= \lim\{\Sigma \frac{B(t_{i+1})^2 - B(t_i)^2}{2} - \Sigma \frac{[B(t_{i+1}) - B(t_i)]^2}{2}\} \\ &= \frac{B(t)^2}{2} - \frac{t}{2} \end{aligned} \qquad (A.10)$$

The solution (A.10) disagrees with (A.7). If $\rho_i = t_{i+1}$ in (A.8), the result also disagrees with (A.7).

$$\begin{aligned} x(t) &= \lim \Sigma\, B(t_{i+1})[B(t_{i+1}) - B(t_i)] \\ &= \lim\{\Sigma \frac{B(t_{i+1})^2 - B(t_i)^2}{2} + \Sigma \frac{[B(t_{i+1}) - B(t_i)]^2}{2}\} \\ &= \frac{B(t)^2}{2} + \frac{t}{2} \end{aligned}$$

These conflicts indicate the need for carefully formulating stochastic differential equations such as (A.1) and for strictly defining the integration procedure to be utilized in the solution of such equations.

The formulation presently being applied in the literature and being relied upon in this thesis has a number of important theoretical advantages (Kailath). This formulation, which applies

---

[6] The Levy oscillation property is shown in Doob [41], p. 395.

when $\underline{u}(t)$ is Gaussian, formally makes $\underline{u}(t) = \frac{d\underline{B}(t)}{dt}$, where $\underline{B}(t)$ is a vector Brownian motion process, and states that (A.1) is equivalent to (A.11).

$$\underline{x}(t) = \underline{x}(t_o) + \int_{t_o}^{t} \underline{c}(s)ds + \int_{t_o}^{t} \underline{D}(s)d\underline{B}(s) \tag{A.11}$$

The solution to (A.11) is symbolically represented as:

$$d\underline{x}(t) = \underline{c}(t)dt + \underline{D}(t)d\underline{B}(t). \tag{A.12}$$

The first integral in (A.11) is treated as a Reimann integral, while the second is defined as the Ito integral, first defined by Ito [45]. Since the Ito calculus is used throughout the thesis, its properties are listed in Appendix B. One important property of this calculus is the chain-rule for differentiation (Theorem B.1, Ito's Lemma), which is used when determining the differential equation for a function of $\underline{x}(t)$.

To illustrate Ito's lemma and to show the difference between Ito and non-Ito calculus, the stochastic differential equation in (A.13), is considered.

$$dz(t) = \tfrac{1}{2} z(t)dt + z(t)dB(t) \;\; ; \;\; z(0) = 1 \tag{A.13}$$

Theorem B.1 shows that the solution in the Ito sense is:

$$z(t) = \exp[B(t)]. \tag{A.14}$$

Using the notation of Theorem B.1, let $a(t) = 0$ and $b(t) = 1$. Then, $x(t) = B(t)$. Now, let $Z(t,x) = \exp[x]$. Theorem B.1 says that

$$\begin{aligned} dz(t) &= \tfrac{1}{2} \frac{\partial^2}{\partial x^2} \exp[x(t)]dt + \frac{\partial}{\partial x} \exp[x(t)]dB(t) \\ &= \tfrac{1}{2} z(t)dt + z(t)dB(t). \end{aligned}$$

Thus, when Ito calculus is used to analyze (A.13), (A.14) is the correct solution. The solution to (A.13) when applying the rules of ordinary calculus to (A.13) divided by dt is:

$$z(t) = \exp[B(t) + t/2].$$

Other ways of treating stochastic differential equations have been documented. All assume (A.1) is equivalent to (A.11) and specify different definitions for the Stieltjes-type integral in (A.11). Stratonovich [32] has provided the most popular alternative to Ito's definition; his method selects $\rho_i = \frac{t_{i+1} + t_i}{2}$ in (A.8) and provides solutions to (A.11) which would be obtained by formal integration, at least in the scalar case. Unfortunately, his integral doesn't possess analytical properties as convenient as those listed in Appendix B for the Ito integral.

This discussion has concentrated on stochastic differential equations containing Gaussian white noise. The remainder of this appendix describes the available literature concerning definitions for stochastic integrals, such as the one in (A.8), when the noise in (A.1) is non-Gaussian. Portions of the following comments are also found in Kailath [34] and Fisher [8].

The correct way of representing white noise is to define it as a generalized random process, such as discussed in Gel'fand and Vilenkin [46]. Part of Gel'fand's development shows that the generalized derivative of Brownian motion has a delta function for its correlation function. More generally, the random variables comprising the formal derivative of any process with independent increments are shown to be independent. In other words, the formal derivative of an independent increment process is white. This fact has been used in formulating stochastic differential equations since both the Ito and the Stratonovich integral have been defined when $B(t)$ is replaced by any independent increment process [13], [45]. However, this more general representation of white noise cannot be applied when analyzing observation processes such as (1.2) and (2.2). Independent increment processes can be shown to be either impulsive (Poisson) processes or Brownian motion processes or a combination of the two. Kailath and Fisher indicate that it is not possible to define a filtering problem which has an observation process containing impulsive noise.

In summary, if the white noise in the models of the observation processes used for the filtering problem is considered equivalent to the formal derivative of an independent increment process, the white noise must be Gaussian.

There is at least one other way to interpret white noise. Assume it is equivalent to the formal derivative of any process with orthogonal increments. Results on the derivative of independent increment processes indicate that the random variables comprising these processes are independent when all that is needed is orthogonality. Thus, the derivatives of orthogonal-increment processes do seem logical as representations for white noise. There does not seem to be any published work on the formulation of stochastic differential equations using this interpretation of white noise. Because of these limitations in the theory of stochastic differential equations, this thesis treats only Gaussian, white noise in the observed process (2.2).

APPENDIX B

σ-ALGEBRAS, ITO CALCULUS, AND MARTINGALES

This Appendix defines and discusses three important concepts in stochastic processes.

B.1 σ-Algebras Generated by Stochastic Processes

Let $m(t,\omega)$ be a stochastic process defined on the probability space $\{\Omega,H,P\}$ where $\omega \in \Omega$ and $t \in I$, the index set; Ω is a sample space, H is a σ-algebra of measurable sets and P is a measure assigned to the sets in H. For the remainder of this appendix, and for the entire thesis, $m(t,\omega)$ is denoted as $m(t)$.

The σ-algebra generated by the random process $[m(s),\ t_o \le s \le t]$ is defined as the minimal σ-algebra of events that contains all events of the form: $\{\omega : \omega \in \Omega$ and $m(u,\omega) \in A\}$ where $u \in [t_o,t]$ and $A \in R_1$ is any Borel set. This σ-algebra is denoted by $F\{[t_o,t],m(u)\}$. Theorems from measure theory [41], [47] show that $F\{[t_o,t],m(u)\} \subset H$, that $F\{[t_o,s],m(u)\} \subset F\{[t_o,t],m(u)\}$ for $t \ge s$, and that $m(t)$ is measurable on $F\{[t_o,t],m(u)\}$.[7]

b.2 Properties of the Ito Integral

Let $B(t)$ be a Brownian motion process defined on $I = [t_o,T]$. This section lists the properties of the Ito calculus that follow from Ito's definition of the integral $\int_{t_o}^{t} g(s)dB(s)$. This integral cannot be defined in the ordinary Riemann-Stietjes sense since $B(t)$ is of unbounded variation with probability one.

---

[7] For a discussion of σ-algebras (σ-fields) consult Loeve [47] (Chap. 1). Doob [41] (pp. 599-602) has a discussion of σ-algebras generated by random processes.

Ito calculus is thoroughly discussed in Skorokhod [45]. The σ-algebra generated by $B(s)$, where $t_o \le s \le t$, is $F\{[t_o,t],B(s)\}$. Let $M_2(F_t)$ be the class of functions $g(t)$ which are measurable on $F\{[t_o,t],B(s)\}$ and satisfy $\text{Prob}\{\int_{t_o}^{T} |g(s)|^2 dt < \infty\} = 1$. The Ito integral $\int_{t_o}^{t} g(s)dB(s)$ is defined for all $g(t) \in M_2(F_t)$ and has the following properties.

<u>Property B.1</u>: $E \int_{t_o}^{t} g(s)dB(s) = 0 \qquad t \in I$

<u>Property B.2</u>: $E[\int_s^t g(u)dB(u) | F\{[t_o,s],B(r)\}] = 0$ for $t \ge s$ and $t,s \in I$[8]

<u>Property B.3</u>: $E|\int_{t_o}^{t} g(s)dB(s)|^2 = \int_{t_o}^{t} E|g(s)|^2 ds$ for $t \in I$[9]

<u>Property B.4</u>: The process $w(t) = \int_{t_o}^{t} g(s)dB(s)$ is a martingale.

<u>Property B.5</u>: The process $w(t)$ is a continuous process with probability one.

Appendix A indicated that the Ito integral has been extremely useful in the formulation and analysis of stochastic differential equations. Special care must be taken when applying the chain-rule for differentiation to these equations. Theorem B.1 provides the proper differentiation formula.

<u>Theorem B.1</u> (Ito's Lemma) Let $x(t)$ be a process satisfying

$$dx(t) = a(t)dt + b(t)dB(t) \qquad \text{w.P.1} \quad \text{for} \quad t \in [t_o,T],$$

where $B(t)$ is Brownian motion and $a(t)$, $b(t)$ and $b(t)^2$ belong to $M_2(F_t)$. If $Z(t,x)$ is continuous and has continuous derivatives $\frac{\partial}{\partial t} Z(t,x)$, $\frac{\partial}{\partial x} Z(t,x)$, and $\frac{\partial^2}{\partial x^2} Z(t,x)$ for $t \in [t_o,T]$, then the process $z(t) = Z(t,x(t))$ satisfies the relation

---

8 For a discussion of conditional expectations refer to Doob [41], p. 37.

9 All integrals not of the Ito type are treated as Riemann integrals.

$$dz(t) = [\frac{\partial}{\partial t} Z(t,x(t)) + a(t) \frac{\partial}{\partial x} Z(t,x(t)) + \tfrac{1}{2}b(t)^2 \frac{\partial^2}{\partial x^2} Z(t,x(t))]dt$$
$$+ b(t) \frac{\partial}{\partial x} Z(t,x(t))dB(t).$$

The notation $\frac{\partial}{\partial t} Z(t,x(t))$ has been substituted for $\frac{\partial}{\partial s} Z(s,x(t))|_{s=t}$.

B.3 Martingale Theory

A random process $w(t)$ defined on a time interval I is called a martingale if for every $t \in I$ there corresponds a σ-algebra A_t relative to which the random variables $w(s)$ are measurable for $s \leq t$, $s \in I$, and which possesses the property that for $t_1 \leq t_2$ where $t_1, t_2 \in I$

$$E[w(t_2)|A_{t_1}] = w(t_1) \qquad \text{w.P.1.} \qquad \text{(B.1)}$$

If the equality in (B.1) is replaced by $\leq$ ($\geq$), a super-martingale (sub-martingale) is defined.

Some of the martingale theorems given later require that $w(t)$ be a separable process. A process $w(t)$ is separable on I if there is a denumerable, everywhere dense subset D of I such that for any $a < b$ in I

$$\sup_{t\in(a,b)} w(t) = \sup_{t\in(a,b)\cap D} w(t) \quad \text{and} \quad \inf_{t\in(a,b)} w(t) = \inf_{t\in(a,b)\cap D} w(t) \quad \text{w.P.1.}$$

The separability condition is satisfied if the process $w(t)$ is right-continuous in t (w.P.1) on the interval I. In this thesis, all processes are continuous (Appendix D), so separability is automatically satisfied and need not be mentioned.

The following three theorems on martingales are used in this thesis.

Theorem B.2: If $w(t)$ is a non-negative, super-martingale,

$$\lim_{t\to\infty} w(t) = w_\infty$$

exists with probability one and is finite.

Theorem B.3: If $w(t)$ is a (separable) non-negative super martingale on any interval I, then for any $t \in I$ and any constant c,

$$\text{Prob}\{\sup_{u \geq t, u \in I} w(u) \geq c\} \leq \frac{Ew(t)}{c} .$$

Theorem B.4: If $w(t)$ is a (separable) sub-martingale on some interval I, then for any $t \in I$ and any constant c,

$$\text{Prob}\{\sup_{u \leq t, u \in I} w(u) \geq c\} \leq \frac{E|w(t)|}{c} .$$

The proofs of these properties can be found in Loeve [47] (sub-sections 29.3 and 36.1) and Doob [41] (chapter 7).

APPENDIX C

THE INFINITE DIMENSIONAL REPRESENTATION FOR THE OPTIMAL FILTER

This short exposition develops the system of random differential equations which must be solved to provide the minimum-variance estimate of the message process $x(t)$ in (2.1). This system represents the filter which is an optimal solution to the filtering problem discussed in Sec. 2.2.

In Sec. 2.2, the minimum-variance (conditional mean) estimator is represented by the following stochastic differential equation.

$$d\hat{x}(t) = \hat{f}(t)dt + [\widehat{xh}(t) - \hat{x}(t)\hat{h}(t)][dz(t) - \hat{h}(t)dt] \qquad (2.7)$$

In (2.7), $\hat{x}(t)$ is the conditional mean at time t. The other functions are defined in Sec. 2.2. To solve (2.7), where the observed process $dz(t)$ is an input driver, the functions $\hat{f}(t)$, $\widehat{xh}(t)$, $\hat{h}(t)$ must be found.

The function $\hat{h}(t)$ is examined first. Equation (2.6) represents $\hat{h}(t)$, the conditional expectation of $h(t,x(t))$, when $g(t,x) = h(t,x)$. Then

$$d\hat{h}(t) = \widehat{fh_1}(t)dt + [\widehat{h^2}(t) - \hat{h}(t)^2][dz(t) - \hat{h}(t)dt]$$

where $h_1 = \frac{\partial}{\partial x} h(t,x)$. Now, $\widehat{fh_1}(t)$ and $\widehat{h^2}(t)$ need to be determined so (2.6) must be consulted.

$$d\widehat{fh_1}(t) = \widehat{ff_1h_1}(t)dt + [\widehat{f^2h_2}(t)dt + \widehat{fhh_1}(t) - \widehat{fh_1}(t)\hat{h}(t)][dz(t) - \hat{h}(t)dt]$$

$$d\widehat{h^2}(t) = 2\,\widehat{fhh_1}(t)dt + [\widehat{h^3}(t) - \widehat{h^2}(t)\hat{h}(t)][dz(t) - \hat{h}(t)dt]$$

To continue this procedure, $\widehat{ff_1h_1}(t)$, $\widehat{f^2h_2}(t)$, $\widehat{fhh_1}(t)$, $\widehat{h^3}(t)$ must be determined. In fact, an endless string of differential equations need to be solved to determine all needed conditional expectations. The same situation occurs when determining $\hat{f}(t)$ and $\widehat{xh}(t)$.

This endless string of differential equations, which must be solved simultaneously with (2.7), provides the infinite dimensional representation for the optimal filter.

APPENDIX D

IMPORTANT PROPERTIES OF STOCHASTIC DIFFERENTIAL EQUATIONS SIMILIAR TO EQUATION 2.15

The purpose of this appendix is to indicate conditions on $h(t,x)$ and $a(t,\bar{x}_{t_o,t})$ which are sufficient for providing $\bar{x}(t)$ in (2.15) with two important properties: (1) Continuity of $\bar{x}(t)$ w.P.1 and (2) $F\{[t_o,t],\bar{x}(s)\} \subset F\{[t_o,t],B(s)\}$. The following derivations are based on Doob [41] (chapter 6) and Skorokhod [45] (chapters 2 and 3). This appendix does not deal directly with (2.15) but with an equation containing the same features as (2.15) and fewer terms.

Let $w(t)$ be the solution to

$$dw(t) = g(t,w(t))dt + a(t,w_{t_o,t})dB(t) \quad \text{for} \quad t \in [t_o,T] \qquad (C.1)$$

where $B(t)$ is a Brownian motion process and $w_{t_o,t} = \{w(s) : t_o \le s \le t\}$. The following conditions on $g(t,w(t))$ and $a(t,w_{t_o,t})$ are postulated for the remainder of this appendix.

H_1: The function $g(t,w)$ is a continuous function in the pair (t,w) and $a(t,w_{t_o,t})$ is measurable on $F\{[t_o,t],w(s)\}$.

H_2: The function $g(t,w)$ satisfies the uniform Lipschitz condition

$$|g(t,x) - g(t,y)| \le K|x - y|$$

for $x,y \le R_1$ and $t \in [t_o,T]$; K is a fixed constant.

The function $a(\cdot,\cdot)$ satisfies the following function-space condition

$$|a(t,x_{t_o,t}) - a(t,y_{t_o,t})| \le K[\int_{t_o}^{t} |x(s) - y(s)|^2 ds]^{\frac{1}{2}}$$

where $x_{t_o,t}$ and $y_{t_o,t} \in C_{t_o}^{t}$, the space of continuous functions on the time interval $[t_o,t]$. The constant K may be selected equal to the Lipschitz constant for $g(\cdot,\cdot)$ with no loss of generality.

The derivations in the remainder of this appendix are concentrated on demonstrating the following properties of the solution to (C.1).

P_1: The function $w(t)$ is continuous in t with probability one.

P_2: $F\{[t_o,t],w(s)\} \subset F\{[t_o,t],B(s)\}$ for every $t \in [t_o,T]$.

These properties are demonstrated by a proof constructed similar to Doob [41] (pp. 277-281). A by-product of this proof is the existence of the solution to (C.1). The derivation begins with the following lemma.

__Lemma__: If a process $w(s)$ has properties P_1 and P_2 and if $g(\cdot,\cdot)$ and $a(\cdot,\cdot)$ have properties H_1 and H_2 then any process $y(t)$ defined by

$$y(t) = \int_{t_o}^{t} g(s,w(s))ds + \int_{t_o}^{t} a(s,w_{t_o,s})dB(s) \qquad (C.2)$$

has property P_2 and the Ito interpretation of the second integral in (C.2) provides $y(t)$ with property P_1.

__Proof__. According to H_1, H_2, and P_1, the first integrand in (C.2) is a bounded, continuous function of s for almost all sample functions. The first integral is a continuous function of t with probability one. Condition H_1, combined with P_2, indicates that $a(t,w_{t_o,t})$ is measurable on $F\{[t_o,t],B(s)\}$ so the second integral may be defined as an Ito integral (Appendix B). This integral is a martingale and, in addition, is continuous in t with probability one. Continuity for almost all sample functions $y(t)$ follows. It is obvious that $y(t)$ is measurable on $F\{[t_o,t],B(s)\}$ and, since for all $s \le t$, $y(s)$ is measurable on $F\{[t_o,s],B(u)\} \subset F\{[t_o,t],B(u)\}$, the process $[y(s) : t_o \le s \le t]$ is measurable on $F\{[t_o,t],B(u)\}$. The definition of $F\{[t_o,t],y(u)\}$ shows that $F\{[t_o,t],y(u)\} \subset F\{[t_o,t],B(u)\}$. This finishes the proof of the lemma.

A solution can be found for (C.1) by successive approximations that have the properties P_1 and P_2 shown above. Let $w_o(t)$ be any process having properties P_1 and P_2. According to the lemma, it is now possible to define $w_n(t)$ in such a way that every $w_n(t)$ has properties P_1 and P_2.

$$w_n(t) = \int_{t_o}^{t} g(s,w_{n-1}(s))ds + \int_{t_o}^{t} a(s,w_{t_o,s,n-1})dB(s) \qquad (C.3)$$

with $w_{t_o,s,n-1} = \{w_{n-1}(u) : t_o \le u \le s\}$. The following derivation will show that

$$\lim_{n\to\infty} w_n(t) = w(t) \qquad \text{w.P.1} \qquad (C.4)$$

uniformly in t, thus defining a process $w(t)$ with properties P_1 and P_2. Also, it will be shown that

$$\lim_{n\to\infty} \int_{t_o}^{t} g(s,w_n(s))ds = \int_{t_o}^{t} g(s,w(s))ds \qquad \text{w.P.1}$$

$$(C.5)$$

$$\lim_{n\to\infty} \int_{t_o}^{t} a(s,w_{t_o,s,n})dB(s) = \int_{t_o}^{t} a(s,w_{t_o,s})dB(s) \qquad \text{w.P.1}$$

uniformly in t. The process $w(t)$ will be the solution to (C.1).

In proving these facts, the following notation will be convenient.

$$\Delta_n w(t) = w_n(t) - w_{n-1}(t)$$

$$\Delta_n g(t) = g(t,w_n(t)) - g(t,w_{n-1}(t))$$

$$\Delta_n a(t) = a(t,w_{t_o},t,n) - a(t,w_{t_o},t,n-1)$$

From H_2,

$$|\Delta_n g(t)| \le K|\Delta_n w(t)|$$

$$|\Delta_n s(t)|^2 \le K \int_{t_o}^{t} |\Delta_n w(s)|^2 ds .$$

Then, from (C.3) and a property of the Ito integral (Appendix B),

$$\begin{aligned}
E\{[\Delta_n w(t)]^2\} &\le 2\, E\{|\int_{t_o}^{t} \Delta_{n-1} g(s)ds|^2\} + 2\, E\{|\int_{t_o}^{t} \Delta_{n-1} a(s)dB(s)|^2\} \\
&\le 2\, K^2(T-t_o) \int_{t_o}^{t} E[\Delta_{n-1}w(s)]^2 ds + 2 \int_{t_o}^{t} E|\Delta_{n-1}a(s)|^2 ds \\
&\le 2\, K^2(T - t_o) \int_{t_o}^{t} E[\Delta_{n-1}w(s)]^2 ds + \\
&\qquad + 2\, K^2 \int_{t_o}^{t}\int_{t_o}^{s} E[\Delta_{n-1}w(u)]^2 du ds \\
&\le 2\, K^2(T - t_o) \int_{t_o}^{t} E[\Delta_{n-1}w(s)]^2 ds + \\
&\qquad + 2\, K^2 \int_{t_o}^{t} (t-s)E[\Delta_{n-1}w(s)]^2 ds \\
&\le 4\, K^2(T - t_o) \int_{t_o}^{t} E[\Delta_{n-1}w(s)]^2 ds .
\end{aligned}$$

Hence, for some constant c,

$$\begin{aligned}
E[\Delta_n w(t)^2] &\le [4\, K^2(T - t_o)]^{n-1} \int_{t_o}^{t} \frac{(t-s)}{(n-2)!} E[\Delta_1 w(s)]^2 ds \\
&\le \frac{c^n}{n!} \qquad \text{for } t_o \le t \le T.
\end{aligned}$$

Using this inequality and Chebyshev's inequality gives

$$\begin{aligned}
\text{Prob}\{\sup_{t_o \le t \le T} |\int_{t_o}^{t} \Delta_n g(s)ds| \ge 2^{-n}\} &\le \text{Prob}\{K \int_{t_o}^{T} |\Delta_n w(s)|ds \ge 2^{-n}\} \\
&\le 4^n\, E\{[K \int_{t_o}^{T} |\Delta_n w(s)|ds]^2\} \\
&\le 4^n\, K^2(T-t_o)\, \frac{c^n}{n!} .
\end{aligned}$$

Since the last term is the general term of a convergent series, (C.6) holds for sufficiently large n with probability one from the Borel-Cantelli lemma (Doob [41], p. 104).

$$\sup_{t_o \le t \le T} \left| \int_{t_o}^{t} \Delta_n g(s)ds \right| \le 2^{-n} \tag{C.6}$$

The process $\int_{t_o}^{t} \Delta_n a(s)dB(s)$ is a martingale for the reasons given in the lemma. The square of this process is a sub-martingale and the basic martingale inequality for sub-martingales (Theorem B.4) shows that

$$\begin{aligned}
\text{Prob}\{ \sup_{t_o \le t \le T} \left| \int_{t_o}^{t} \Delta_n a(s)dB(s) \right| \ge 2^{-n} \} &\le 4^n \, E\{ \left| \int_{t_o}^{T} \Delta_n a(s)dB(s) \right|^2 \} \\
&\le 4^n \int_{t_o}^{T} E\left|\Delta_n a(s)\right|^2 ds \\
&\le 4^n \, K \int_{t_o}^{T} \int_{t_o}^{s} E[\Delta_n w(u)]^2 du ds \\
&\le 4^n \, K \int_{t_o}^{T} (T-s)E[\Delta_n w(s)]^2 ds \\
&\le 4^n \, K(T - s)^2 \frac{c^n}{n!} .
\end{aligned}$$

Again, the last term is the general term of a convergent series so (C.7) holds for sufficiently large n, w.P.1.

$$\sup_{t_o \le t \le T} \left| \int_{t_o}^{T} \Delta_n a(s)dB(s) \right| < 2^{-n} \tag{C.7}$$

According to (C.6) and (C.7), the integrals on the right hand side of (C.3) converge uniformly in t when $n \to \infty$, w.P.1. Hence, the limit in (C.4) exists uniformly in t, w.P.1 and $w(t)$ has properties P_1 and P_2. The validity of (C.5) follows from two facts: (1) the integrands converge uniformly w.P.1, in the first limit equation; (2) the sub-martingale inequality applied as above, shows that

$$\text{Prob}\{ \sup_{t_o \le t \le T} \left| \int_{t_o}^{t} [a(s, w_{t_o,s,n}) - a(s, w_{t_o,s})] dB(s) \right| \ge \frac{1}{n} \}$$

$$\le n^2 K^2 \int_{t_o}^{T} (T - s) E[w_n(s) - w(s)]^2 ds$$

$$n^2 K^2 \; (T - t_o)^2 \; 2^{-n} \sum_{j=1}^{\infty} \frac{(2c)^j}{j!}$$

where the fact that $E[w_n(t) - w(t)]^2 \le 2^{-n} \sum_{j=1}^{\infty} \frac{(2c)^j}{j!}$ for $t_o \le t \le T$ was used. This last inequality is proved in Doob [41] (p. 281). According to the Borel-Cantelli lemma

$$\sup_{t_o \le t \le T} \left| \int_{t_o}^{t} [a(s, w_{t_o,s,n}) - a(s, w_{t_o,s})] dB(s) \right| < \frac{1}{n}$$

thus proving the second limit in (C.5).

MICHIGAN STATE UNIVERSITY LIBRARIES
3 1293 03058 0264