JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 96, 209–225 (1983)

Convergence Results for Continuous-Time Adaptive Stochastic Filtering Algorithms

J. H. VAN SCHUPPEN

Mathematical Centre, P. O. Box 4079, 1009 AB Amsterdam, The Netherlands

Submitted by K. J. Astrom

The adaptive stochastic filtering problem for Gaussian processes is considered. The self-tuning synthesis procedure is used to derive two algorithms for this problem. Almost sure convergence for the parameter estimate and the filtering error will be established. The convergence analysis is based on an almost-supermartingale convergence lemma that allows a stochastic Lyapunov-like approach.

1. INTRODUCTION

The goal of this paper is to present two algorithms for a continuous-time adaptive stochastic filtering problem and to establish almost sure convergence results for these algorithms.

What is the adaptive stochastic filtering problem? Problems of prediction and filtering arise in many areas of engineering and economics. For these problems mathematical models in the form of stochastic dynamic systems may be formulated. When the parameter values of these systems are known, the prediction or filtering problem may be solved by applying known filtering techniques such as the Kalman filter. When the parameter values are not known these have to be estimated. The parameter estimation may be done off-line, before the filtering operation starts, or on-line, concurrent with the filtering operation. The adaptive stochastic filtering problem for a stochastic system whose parameter values are not known, is to simultaneously estimate the parameter values and to predict or filter the state of the process. This problem is highly relevant for applications. Algorithms for this problem are especially of interest when the parameter values are slowly changing as is often the case in industrial applications.

In discrete time the adaptive stochastic filtering problem has been investigated by many researchers. Why should one consider the continuous-time version of the problem? Time is generally perceived to be continuous. In practice a continuous-time signal is sampled and the subsequent data processing is done in a discrete time mode. One question then is what happens with the predictions when the sampling time gets smaller and smaller? Does the discrete-time algorithm converge in some sense? To study these and related questions continuous-time algorhtims must be derived and their relationship with discrete-time algorithms investigated.

The questions that one would like to solve for the adaptive stochastic filtering problem are how to synthesize algorithms, and how to evaluate the performance of these algorithms.

Synthesis procedures for the adaptive stochastic filtering problem are summarized below. The self-tuning synthesis procedure prescribes to estimate, separately but concurrently, the parameter values and perform the filtering operation. On the contrary, the second synthesis procedure prescribes to estimate the parameter values and states jointly. In the latter procedure the extended Kalman filter is often used. A criticism of the second procedure is that it treats states and parameters on an equal basis. In this paper attention is restricted to the self-tuning synthesis procedure. This procedure suggests first to solve the associated stochastic filtering problem, and secondly to estimate the values of the parameters of the filter system in a recursive or on-line fashion. A continuous-time recursive parameter estimation algorithm is thus needed.

What is known about continuous-time parameter estimation algorithms? A search of the literature has turned up mainly nonrecursive or off-line algorithms [1–4, 20], for which convergence questions are discussed. However, for adaptive stochastic filtering, recursive algorithms are absolutely necessary. Two such algorithms are presented below.

In the performance evaluation of the algorithms the major question is the convergence of the error in the filtering estimate and the parameter estimate. For these variables one should consider almost sure convergence and the asymptotic distribution. Convergence results for these error processes will be provided below. This result is based on a convergence theorem that is of independent interest.

A brief outline of the paper follows. The problem formulation is given in Section 2. The main results are presented in Section 3, while their proofs may be found in Section 5. Section 4 is devoted to a convergence theorem. A preliminary version of this paper, without proofs, has been presented elsewhere [18].

2. The Problem Formulation

The adaptive stochastic filtering problem is to predict or to filter a stochastic process when the parameters of the distribution of this process are unknown. The object of this section is to make this problem formulation precise. Recall that the self-tuning synthesis procedure for this problem has been adopted which prescribes first to derive the solution of the stochastic filtering problem and then to estimate recursively the parameters of the filter system.

Throughout this paper (Ω, F, P) denotes a complete probability space. Let $T = R$. The terminology of Dellacherie and Meyer [6, 7] will be used.

Assume to be given an R-valued Gaussian process with stationary increments. Under certain additional conditions it follows from weak Gaussian stochastic realization theory [9] that this process has a minimal stochastic realization as the output of what will be called a Gaussian system

$$dx_t = Ax_t\, dt + B\, dv_t, \tag{1}$$

$$dy_t = Cx_t\, dt + D\, dv_t, \tag{2}$$

where $y : \Omega \times T \to R$, $x : \Omega \times T \to R^n$, $v : \Omega \times T \to R^m$ is a standard Brownian motion process, $A \in R^{n \times n}$, $B \in R^{n \times m}$, $C \in R^{1 \times n}$, $D \in R^{1 \times m}$. The precise definition of a realization is that it is a stochastic system such that the distribution of the output y of this system is the same as that of the given process.

One may construct the asymptotic Kalman–Bucy filter for the above Gaussian system, which is

$$d\hat{x}_t = A\hat{x}_t\, dt + K(dy_t - C\hat{x}_t\, dt),$$

where

$$\hat{x}_t = E[x_t | F_t^y], \qquad F_t^y = \sigma(\{ y_s, \forall s \leqslant t\}),$$

is constructed such that it satisfies the "usual conditions" [6]. This filter may be rewritten as a Gaussian system

$$d\hat{x}_t = A\hat{x}_t\, dt + K\, d\bar{v}_t, \tag{3}$$

$$dy_t = C\hat{x}_t\, dt + d\bar{v}_t, \tag{4}$$

where $\bar{v} : \Omega \times T \to R$ is the innovations process, a Brownian motion process, say with variance $\sigma^2 t$. It is a result of stochastic realization theory that the two realizations (1), (2) and (3), (4) are indistinguishable on the basis of information about the distribution of y only. For adaptive stochastic filtering one may therefore limit attention to the realization (3), (4). That realization has the additional advantage that it is suitable for prediction purposes.

The minimality of (1), (2), and hence the minimality of (3), (4), implies that (A, C) is an observable pair and that the spectrum of A is in $C^- := \{c \in C | \text{Re}(c) < 0\}$.

2.1. PROBLEM. Assume given an R-valued Gaussian process with stationary increments having a minimal past-output based stochastic realization given by

$$d\hat{x}_t = A\hat{x}_t\,dt + K\,d\bar{v}_t, \tag{5}$$

$$dy_t = C\hat{x}_t\,dt + d\bar{v}_t, \tag{6}$$

$$\hat{z}_t = C\hat{x}_t, \tag{7}$$

with the properties given above. Assume further that the values of the dimension n and of σ^2, occurring in the variance of $\bar{v}$, are known, but that the calues of A, K, C are unknown. The *adaptive stochastic filtering problem* for the above defined Gaussian system is to recursively estimate $\hat{z}$ given y.

The second step of the self-tuning synthesis procedure prescribes to recursively estimate the parameters of the filter system (3), (4). To solve this parameter estimation problem another representation of this dynamic system is required. This representation is derived below. For notational convenience the time set is taken to be $T = R_+$ in the following.

2.2. PROPOSITION. *Given the Gaussian system as defined in* (1), (2) *and* (3), (4), *the two following representations describe the same relation between* $\bar{v}$ *and* $\hat{z}$:

(a)

$$d\hat{x}_t = A\hat{x}_t\,dt + K\,d\bar{v}_t, \qquad \hat{x}_0 = 0,$$
$$\hat{z}_t = C\hat{x}_t$$
$$dy_t = \hat{z}_t\,dt + d\bar{v}_t, \qquad y_0 = 0.$$

(b)

$$dh_t = Fh_t\,dt + G_1\,dy_t + G_2\,d\bar{v}_t, \qquad h_0 = 0, \tag{8}$$

$$\hat{z}_t = h_t^{\mathrm{T}}p, \tag{9}$$

$$dy_t^{\mathrm{T}} = h_t^{\mathrm{T}}p\,dt + d\bar{v}_t, \qquad y_0 = 0, \tag{10}$$

where $h : \Omega \times T \to R^{2n}$,

$$h_t^{\mathrm{T}} = (y_t^{(1)},\ldots, y_t^{(n)}, \bar{v}_t^{(1)},\ldots, \bar{v}_t^{(n)}),$$
$$y_t^{(1)} = y_t, \qquad \bar{v}_t^{(1)} = \bar{v}_t,$$
$$y_t^{(i)} = \int_0^t y_s^{(i-1)}\,ds, \qquad \textit{for} \quad i = 2, 3,\ldots, n,$$

$p \in R^{2n}$ is related to A, K, C, as indicated in the proof,

$$F_1 = \begin{pmatrix} 0 & \cdots & 0 \\ & & \vdots \\ I_{n-1} & & 0 \end{pmatrix} \in R^{n \times n}, \qquad F = \begin{pmatrix} F_1 & 0 \\ 0 & F_1 \end{pmatrix} \in R^{2n \times 2n},$$

$$G_1 = e_1 \in R^{2n}, \qquad G_2 = e_{n+1} \in R^{2n},$$

where e_i is the ith unit vector.

Proof. (a) → (b). By the remark below (1), (2), (A, C) is an observable pair. Then there exists a basis transformation, say $T \in R^{n \times n}$ nonsingular, such that with $\hat{w}_t = T\hat{x}_t$

$$d\hat{w}_t = \begin{pmatrix} a_1 & I_{n-1} & & \\ \vdots & & & \\ a_n & 0 & \cdots & 0 \end{pmatrix} \hat{w}_t \, dt + \begin{pmatrix} k_1 \\ \vdots \\ k_n \end{pmatrix} d\bar{v}_t, \qquad \hat{w}_0 = 0,$$

$$\hat{z}_t = (10 \cdots 0)\, \hat{w}_t.$$

By successive substitution it is then shown that

$$\hat{z}_t = h_t^{\mathrm{T}} p,$$

where h is as given before, and

$$p^{\mathrm{T}} = (a_1, a_2, \ldots, a_n, k_1 - a_1, \ldots, k_n - a_n) \in R^{2n}.$$

The representation (b) then follows.

(b) → (a). Set p as above

$$d\bar{v}_t = dy_t - h_t^{\mathrm{T}} p \, dt,$$

$$\hat{w}_t^1 = h_t^{\mathrm{T}} p,$$

$$d\hat{w}_t^{n-1} = a_n \hat{w}_t^1 \, dt + k_n \, d\bar{v}_t,$$

$$d\hat{w}_t^{n-1} = a_{n-1} \hat{w}_t^1 \, dt + \hat{w}_t^n \, dt + k_{n-1} \, d\bar{v}_t,$$

$$\vdots$$

$$d\hat{w}_t^2 = a_2 \hat{w}_t^1 \, dt + \hat{w}_t^3 \, dt + k_2 \, d\bar{v}_t.$$

It is then shown by induction that

$$d\hat{w}_t^1 = a_1 \hat{w}_t^1 \, dt + \hat{w}_t^2 \, dt + k_1 \, d\bar{v}_t. \quad \blacksquare$$

3. The Main Results

In this section two algorithms are presented for the continuous-time adaptive stochastic filtering problem, and convergence results are provided. The proofs of the convergence results may be found in Section 5.

In the following attention is restricted from the Gaussian system defined by (3), (4), or by (5), (6), to the autoregressive case described by

$$y_t = \sum_{i=1}^{n} a_i y_t^{(i+1)} + \bar{v}_t,$$

or

$$dy_t = h_t^{\mathrm{T}} p \, dt + d\bar{v}_t, \qquad y_0 = 0, \tag{11}$$

where now $h : \Omega \times T \to R^n$, $p \in R^n$,

$$\begin{aligned} h_t^{\mathrm{T}} &= (y_t^{(1)}, \dots, y_t^{(n)}), \\ p^{\mathrm{T}} &= (a_1, \dots, a_n). \end{aligned} \tag{12}$$

Then

$$dh_t = \begin{pmatrix} a_1 & \cdots & a_n \\ & & 0 \\ & & \vdots \\ I_{n-1} & & 0 \end{pmatrix} h_t \, dt + \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} d\bar{v}_t, \qquad h_0 = 0. \tag{13}$$

One concludes that asymptotically h is a stationary Gauss–Markov process. Since the interest here is in the stationary situation, it will henceforth be assumed that h is a stationary Gauss–Markov process. Because of the stability of the Gaussian system, the covariance function of h is integrable, hence h is an ergodic process [19, p. 69].

3.1. Definition. The adaptive stochastic filtering algorithm RLS for the autoregressive representation (11), (13) based on the least-squares parameter estimation algorithm is defined by

$$d\hat{p}_t = Q_t h_t \sigma^{-2} [dy_t - h_t^{\mathrm{T}} \hat{p}_t \, dt], \qquad \hat{p}_0 = 0, \tag{14}$$

$$dQ_t = -Q_t h_t h_t^{\mathrm{T}} Q_t \sigma^{-2}, \qquad Q_0, \tag{15}$$

$$\hat{\hat{z}} = h_t^{\mathrm{T}} \hat{p}_t, \tag{16}$$

where $\hat{p} : \Omega \times T \to R^n$, $Q : \Omega \times T \to R^{n \times n}$, $Q_0 \in R^{n \times n}$ such that $Q_0 = Q_0^{\mathrm{T}} > 0$, $\hat{\hat{z}} : \Omega \times T \to R$. Here $\hat{\hat{z}}$ is the desired estimate of $\hat{z}$ and $\hat{p}$ is an estimate of the parameter p.

It follows from [8] that the stochastic differential equation for $\hat{p}$ (14) has a unique solution. Here y is assumed to be generated by (11), the underlying σ-algebra family generated by the Brownian motion process $\bar{v}$, and $p \in R^n$.

In the following digression a derivation of the algorithm 3.1 via the Bayesian method is given. Consider the representation

$$dp_t = 0, \qquad p_0 = 0,$$
$$dy_t = h_t^{\mathrm{T}} p_t \, dt + d\bar{v}_t, \qquad y_0 = 0,$$

where it is now assumed that $\bar{v}$ is a Brownian motion process, $p : \Omega \times T \to R^n$, p is a Gaussian random variable with mean 0 and variance Q_0, and that p and $\bar{v}$ are independent objects. From (12) one concludes that $(h_t, F_t^y, t \in T)$ is adapted. The conditional Kalman–Bucy filter [13, 12.1] applied to the above representation then yields the algorithm given in 3.1. Actually the conditions of [13, 12.1] are stronger than necessary; a similar result holds under weaker conditions. This is the end of the digression and in the following the assumptions above 3.1 will be in force.

To evaluate adaptive stochastic filtering algorithms two questions are relevant:

(1) is $\lim_{t\to\infty} \hat{z}_t - \hat{\hat{z}}_t = 0$ in some sense, and if so what is the asymptotic distribution of this difference;

(2) is $\lim_{t\to\infty} \hat{p}_t - p = 0$ in some sense, and if so what is the asymptotic distribution of this difference.

The first question concerns the difference of the filter estimate $\hat{z}$ obtained with knowledge of the parameters, and the adaptive filter estimate $\hat{\hat{z}}$. The second question deals with the error in the parameter estimate.

In the literature the second question is often emphasized. In the opinion of the author the first question is much more relevant, because the adaptive filter estimate is available to an outside observer and is what one is ultimately interested in; the parameters are inaccessible to an outside observer anyway.

3.2. THEOREM. *Consider the adaptive stochastic filtering problem* 2.1 *for the system* (5), (6) *restricted to the autoregressive case as indicated above. Assume that the conditions of* 2.1 *hold, in particular that* n, σ^2 *are known. If the algorithm* RLS *is applied to this stochastic system, then*

(a) as-$\lim_{t\to\infty} t^{-1} \int_0^t (\hat{z}_s - \hat{\hat{z}})^2 \, ds = 0$;

(b) as-$\lim_{t\to\infty} \hat{p}_t = p$.

The above result means that under the conditions given the error in the filter estimate goes to zero in the above defined sense. Why convergence can only be proven in the sense of 3.2(a) is not clear. It is related to the fact that in adaptive stochastic control only results for the average cost function can be proven.

One might conjecture that a result like 3.2 holds if the restriction to the autoregressive case is relaxed and an extended least-squares algorithm is applied. An investigation has indicated that such a conjecture may not be true. The reason for this may be explained as follows. Consider the representation (11). The recursive least-squares algorithm RELS applied to this representation is given by

$$\begin{aligned} d\hat{p}_t &= Q_t \hat{h}_t \sigma^{-2}[dy_t - \hat{h}_t^{\mathrm{T}} p_t\, dt], & \hat{p}_0 &= 0, \\ dQ_t &= -Q_t \hat{h}_t \hat{h}_t^{\mathrm{T}} Q_t \sigma^{-2}\, dt, & Q_0&, \\ d\hat{h}_t &= F\hat{h}_t dt + G_1\, dy_t + G_2(dy_t - \hat{h}_t^{\mathrm{T}} \hat{p}_t\, dt), & \hat{h}_0 &= 0, \\ \hat{\hat{z}}_t &= \hat{h}_t^{\mathrm{T}} \hat{p}_t. \end{aligned}$$

A detailed derivation of this algorithm, as given below 3.1 for the RLS algorithm, runs into serious trouble, but let us not consider that question here. The process $\hat{h}$ contains, besides y, the second innovation process

$$d\bar{\bar{v}}_t = dy_t - \hat{h}_t^{\mathrm{T}} \hat{p}_t\, dt,$$

and its integrals. Furthermore, $\hat{h}$ is not a stationary process, while in the proof of 3.2 the stationarity of h plays a key role. Convergence of the estimates produced by the RELS algorithm has not been established, and is unlikely in the author's opinion. Prefiltering of the observations and the innovations seems necessary. A consequence of these remarks is that the value of the estimates produced by a discrete-time RELS algorithm may be doubtful when the sampling time goes to zero.

The second algorithm for the autoregressive case is related to that of Goodwin, Ramadge, and Caines [10], and that of Chen [5]. The latter also provides a continuous-time algorithm not only for the autoregressive case but also for the general case of 2.1.

3.3. Definition. The adaptive stochastic filtering algorithm for the autoregressive representation (11) based on the parameter estimation algorithm AML2 [10] is defined to be

$$d\hat{p}_t = h_t r_t^{-1} \sigma^{-2}[dy_t - h_t^{\mathrm{T}} \hat{p}_t dt], \qquad \hat{p}_0 = 0, \tag{17}$$

$$dr_t = \sigma^{-2} h_t^{\mathrm{T}} h_t\, dt, \qquad r_0 = 1, \tag{18}$$

$$\hat{\hat{z}} = h_t^{\mathrm{T}} \hat{p}_t, \tag{19}$$

where $\hat{p} : \Omega \times T \to R^n$, $r : \Omega \times T \to R$, $\hat{\hat{z}} : \Omega \times T \to R$, and h is as given in (12). Here $\hat{\hat{z}}$ is the desired adaptive filter estimate of $\hat{z}$ and $\hat{p}$ is an estimate of p.

3.4. THEOREM. *Consider the adaptive filtering problem* 2.1 *for the system* (5), (6) *restricted to the autoregressive case as indicated above. If the algorithm* AML2 *is applied to this system, then*

$$\text{as-}\lim_{t \to \infty} t^{-1} \int_0^t (\hat{z}_s - \hat{\hat{z}}_s)^2 \, ds = 0.$$

The comments given below 3.2. also apply here. The method of proof does not provide information on the question whether as-lim $\hat{p}_t = p$. One may pose the question how the asymptotic variances of $(\hat{z}_s - \hat{\hat{z}}_s)$ of the estimates produced by the algorithm RLS and AML2 are related. Chen [5] considers also the algorithm AML2 but applies it to the representation (10). Almost sure convergence for such an algorithm is established under an unnatural assumption [5, (54)].

4. A CONVERGENCE RESULT

The convergence results of Section 3 are based on an almost sure convergence theorem that is of independent interest. In this section this result is stated and proven.

As some of the other concepts and results of system identification, the convergence theorem is also inspired by the statistics literature, in particular by the area of stochastic approximation. Robbins and Siegmund [15] established a discrete-time convergence result for use in stochastic approximation theory. A simplified version of that result is given as an exercise in [14, II–4]. Solo [16, 17] has been the first to use this result in the system identification literature, and since then it has become rather popular [10, 12]. This popularity is due not only to the ease with which convergence results are proven but also to the formulation in terms of martingales which show up naturally in stochastic filtering and stochastic control problems. Below the continuous time analog of [15, Theorem 1] is given.

A few words about notation follow. $(F_t, t \in T)$ denotes a σ-algebra family satisfying the usual conditions. A^+ is the set of increasing processes, $M_{1\text{uloc}}$ the set of locally uniformly integrable martingales, and $\Delta x_t = x_t - x_{t-}$ the jump of the process x at time $t \in T$.

4.1. THEOREM. *Let* $x : \Omega \times T \to R_+$, $a : \Omega \times T \to R_+$, $b : \Omega \times T \to R_+$, $e : \Omega \times T \to R_+$, *and* $m : \Omega \times T \to R$ *be stochastic processes. Assume that*

(i) $x_0 : \Omega \to \quad_+$ *is* F_0 *measurable;*

(ii) $(a_t, F_t, t \in T) \in A^+$, $a_0 = 0$, $a_\infty < \infty$ *a.s., and there exists a* $c_1 \in R_+$ *such that for all* $t \in T$, $\Delta a_t \leqslant c_1$; $(b_t, F_t, t \in T) \in A^+$ *and* $b_0 = 0$;

(iii) $(e_t, F_t, t \in T)$ *is adapted and* $\int_0^\infty e_s\, ds < \infty$ *a.s.;*

(iv) $(m_t, F_t, t \in T) \in M_{1\text{uloc}}$, $m_0 = 0$;

(v) x *is the unique solution of*

$$dx_t = e_t x_t\, dt + da_t - db_t + dm_t, x_0.$$

Then

(a) $x_\infty := \text{as-lim}_{t\to\infty} x_t$ *exists in* R_+, *thus* $x_\infty < \infty$ *a.s.;*

(b) $b_\infty := \text{as-lim}_{t\to\infty} b_t$ *exists or* $b_\infty < \infty$ *a.s.*

Proof. (1) Define $\phi : \Omega \times T \times T \to R$, $\phi(t, s) = \exp(\int_s^t e_r\, dr)$ which is well defined by e positive and assumption (iii). Then

$$\phi(t, 0) \leqslant \phi(\infty, 0) < \infty \text{ a.s.}, \qquad \phi(0, t) \leqslant 1,$$

and

$$\partial\phi(0, t)/\partial t = -e_t \phi(0, t).$$

By [8] the stochastic differential equation

$$dx_t = e_t x_t\, dt + da_t - db_t + dm_t, \qquad x_0,$$

has an unique solution, and x is a semimartingale. Define $y : \Omega \times T \to R_+$, $y_t = \phi(0, t)\, x_t$. Application of the stochastic calculus rule yields

$$dy_t = \phi(0, t)\, da_t - \phi(0, t)\, db_t + \phi(0, t)\, dm_t, \qquad y_0 = x_0.$$

(2) For $c \in R_+$ define

$$\tau = \inf \left\{ t \in T \,\middle|\, \int_0^t \phi(0, s)\, da_s > c \right\},$$

$$= +\infty, \qquad \text{otherwise.}$$

Then

$$\int_0^\tau \phi(0, s)\, da_s \leqslant c + \Delta a_\tau \leqslant c + c_1$$

by (1) above and assumption (ii). Furthermore,

$$\begin{aligned} I_{\{x_0<c\}} &\int_0^{t\wedge\tau} \phi(0,s)\,dm_s \\ &= \left[y_{t\wedge\tau} - x_0 - \int_0^{t\wedge\tau} \phi(0,s)\,da_s + \int_0^{t\wedge\tau} \phi(0,s)\,db_s\right] I_{\{x_0<c\}} \\ &\geqslant -2c - c_1. \end{aligned}$$

Let

$$r : \Omega \times T \to R,$$

$$r_t = \int_0^t \phi(0,s)\,dm_s.$$

Then

$$(r_t, F_t, t \in T) \in M_{1\mathrm{uloc}},$$

and if $\{\tau_n, n \in Z_+\}$ is a fundamental sequence [7], then so is $\{\tau_n \wedge \tau, n \in Z_+\}$ for r^τ. By the above

$$\{I_{\{x_0<c\}} r_{t\wedge\tau}, F_t, t \in T\}$$

is bounded from below. For $s, t \in T$, $s \leqslant t$, then

$$\begin{aligned} &E[r_{t\wedge\tau} I_{\{x_0<c\}} \mid F_s] \\ &\qquad \leqslant \text{as-}\lim_n E[r_{t\wedge\tau\wedge\tau_n} \mid F_s]\, I_{\{x_0<c\}}, \end{aligned}$$

by Fatou's lemma,

$$= r_{s\wedge\tau} I_{\{x_0<c\}},$$

by $\{\tau \wedge \tau_n, \; n \in Z_+\}$ a fundamental sequence for r^τ. Thus $(r_{t\wedge\tau} I_{\{x_0<c\}}, F_t, t \in T) \in \mathrm{Sup}M$ is bounded from below. By [7]

$$\text{as-}\lim_{t\to\infty} \int_0^{t\wedge\tau} \phi(0,s)\,dm_s I_{\{x_0<c\}}$$

exists and is finite almost surely,

(3) Consider

$$\begin{aligned} y_{t\wedge\tau} I_{\{x_0<c\}} &+ I_{\{x_0<c\}} \int_0^{t\wedge\tau} \phi(0,s)\,db_s \\ &= x_0 I_{\{x_0<c\}} + I_{\{x_0<c\}} \int_0^{t\wedge\tau} \phi(0,s)\,da_s \\ &\quad + I_{\{x_0<c\}} \int_0^{t\wedge\tau} \phi(0,s)\,dm_s. \end{aligned}$$

By (2) above, the third term on the right-hand side converges, while by the definition of τ and assumption (ii)

$$\text{as-}\lim_{t\to\infty} I_{\{x_0<c\}} \int_0^{t\wedge\tau} \phi(0,s)\,da_s \leqslant c + c_1$$

exists and is finite almost surely. Because y is positive and b increasing both terms on the left-hand side of the above equality must converge to finite limits. Then as-$\lim_{t\to\infty} y_t$ exists and is finite on $\{x_0 < c\} \cap \{\tau = \infty\}$. Furthermore,

$$\{a_\infty \leqslant c\} \subset \left\{ \int_0^\infty \phi(0,s)\,da_s \leqslant c \right\} \subset \{\tau = \infty\},$$

thus as-lim y_t exists and is finite on $\{x_0 < c\} \cap \{a_\infty \leqslant c\}$. Since this holds for all $c \in R_+$, $x_0 < \infty$, and $a_\infty < \infty$ a.s., as-lim y_t exists and is finite almost surely. Similarly,

$$\text{as-}\lim_{t\to\infty} \int_0^t \phi(0,s)\,db_s$$

exists and is finite almost surely.

(4) Finally, by assumption (iii),

$$\text{as-}\lim_{t\to\infty} \phi(t,0) = \phi(\infty,0) < \infty \text{ a.s.},$$

hence

$$\text{as-}\lim_{t\to\infty} x_t = \text{as-}\lim_{t\to\infty} y_t \phi(t,0)$$

exists and is finite almost surely, while also

$$\text{as-}\lim_{t\to\infty} b_t = \text{as-lim} \int_0^t \phi(s,0)\,\phi(0,s)\,db_s$$

$$\leqslant \phi(\infty,0)\ \text{as-lim} \int_0^t \phi(0,s)\,db_s$$

exists and is finite almost surely. ∎

5. Proofs

In this section the proofs of Theorems 3.2. and 3.4. are given. The convergence result of Section 4 is used. The method of the proofs is analogous to the Lyapunov method for proving stability of deterministic differential systems.

5.1. *Proof of* 3.2. (1) Let $\tilde{p} : \Omega \times T \to R^n$, $\tilde{p}_t = \hat{p}_t - p$, $\tilde{z} : \Omega \times T \to R$, $\tilde{z}_t = \hat{z}_t - \hat{\hat{z}}_t$, $u : \Omega \times T \to R$,

$$u_t = \tilde{p}_t^{\mathrm{T}} Q_t^{-1} \tilde{p}_t + \int_0^t \sigma^{-2} \tilde{z}_s^2 \, ds.$$

Elementary calculations then show that

$$\begin{aligned} \tilde{z}_t &= \hat{z}_t - \hat{\hat{z}}_t = -h_t^{\mathrm{T}} \tilde{p}_t, \\ d\tilde{p}_t &= Q_t h_t \sigma^{-2} [\tilde{z}_t \, dt + d\bar{v}_t], \\ dQ_t^{-1} &= h_t h_t^{\mathrm{T}} \sigma^{-2} \, dt, \\ du_t &= h_t^{\mathrm{T}} Q_t h_t \sigma^{-2} \, dt + 2(h_t^{\mathrm{T}} \tilde{p}_t) \, \sigma^{-2} \, d\bar{v}_t. \end{aligned}$$

(2) Define $r : \Omega \times T \to R$,

$$dr_t = h_t^{\mathrm{T}} h_t \sigma^{-2} \, dt, \qquad r_0 = \mathrm{tr}(Q_0^{-1}).$$

Then

$$\mathrm{tr}(Q_t^{-1}) = \mathrm{tr}\left(Q_0^{-1} + \int_0^t \sigma^{-2} h_s h_s^T \, ds \right) = r_t.$$

Define $w : \Omega \times T \to R$, $w_t = u_t / r_t$. Then

$$\begin{aligned} dw_t = {} & h_t^{\mathrm{T}} Q_t h_t r_t^{-1} \sigma^{-2} \, dt - w_t (h_t^{\mathrm{T}} h_t r_t^{-1} \sigma^{-2}) \, dt \\ & + 2 r_t^{-1} (h_t^{\mathrm{T}} \tilde{p}_t) \, \sigma^{-2} \, d\bar{v}_t. \end{aligned}$$

(3) To be able to apply 4.1, its conditions are checked. Because Q^{-1} is positive definite, so is Q, and hence u. Thus r and w are positive, and

$$\begin{aligned} \int_0^t r_s^{-1} h_s^T Q_s h_s \sigma^{-2} \, ds &\leqslant \int_0^t r_s^{-1} \, \mathrm{tr}(Q_s^{-1}) \, h_s^{\mathrm{T}} Q_s^2 h_s \sigma^{-2} \, ds \\ &= \mathrm{tr}\left(\int_0^t Q_s h_s h_s^{\mathrm{T}} Q_s \sigma^{-2} \, ds \right) \\ &= \mathrm{tr}(-Q_t + Q_0) \leqslant \mathrm{tr}(Q_0), \end{aligned}$$

$$\text{as-}\lim_{t \to \infty} \int_0^t r_s^{-1} h_s^{\mathrm{T}} Q_s h_s \sigma^{-2} \, ds \leqslant \mathrm{tr}(Q_0) < \infty.$$

(4) From 4.1 then follows that as-lim w_t exists and that

$$\text{as-lim} \int w_s h_s^{\mathrm{T}} h_s r_s^{-1} \sigma^{-2} \, ds < \infty.$$

(5) As argued below 3.1, h is an ergodic process. Hence

$$\begin{aligned} \text{as-lim } t^{-1} Q_t^{-1} &= \text{as-lim } t^{-1} \int_0^t h_s h_s^{\mathrm{T}} \sigma^{-2} \, ds \\ &= \sigma^{-2} E[h_t h_t^{\mathrm{T}}] > 0, \end{aligned}$$

$$\begin{aligned} \text{as-lim } r_t/t &= \text{as-lim } t^{-1} \int_0^t h_s^{\mathrm{T}} h_s \sigma^{-2} \, ds \\ &= \sigma^{-2} E[h_t^{\mathrm{T}} h_t] > 0. \end{aligned}$$

Then

$$\text{as-lim } r_t = +\infty,$$

$$\begin{aligned} &\text{as-lim} \int_0^t r_s^{-1} h_s^{\mathrm{T}} h_s \sigma^{-2} \, ds \\ &\quad = \text{as-lim} \int_0^t r_s^{-1} \, dr_s = \text{as-lim } \ln(r_t) - \ln(r_0) = +\infty. \end{aligned}$$

(6) One now claims that as-lim $w_t = 0$. For if not, then there exists a set of positive measure and an $\varepsilon \in (0, \infty)$, such that on this set

$$\text{as-lim } w_t \geqslant \varepsilon > 0,$$

$$\begin{aligned} &\text{as-lim} \int_0^t w_s h_s^{\mathrm{T}} h_s \sigma^{-2} \, ds \\ &\quad \geqslant (\text{as-lim } w_t) \left(\text{as-lim} \int_0^t h_s^{\mathrm{T}} h_s \sigma^{-2} \, ds \right) = +\infty, \end{aligned}$$

by using (5), which is a contradiction of the conclusion obtained in (4). Hence as-lim $w_t = 0$, and by definition of u and positivity of the terms in u

$$\text{as-lim } r_t^{-1} \int_0^t \tilde{z}_s^2 \sigma^{-2} \, ds = 0,$$

$$\text{as-lim } r_t^{-1} \tilde{p}_t^{\mathrm{T}} Q_t^{-1} \tilde{p}_t = 0.$$

(7) By using a result of (5) above, one obtains

$$\text{as-lim}\, t^{-1}\int_0^t \tilde{z}_s^2\, ds = (\text{as-lim}\, r_t/t)\left(\text{as-lim}\, r_t^{-1}\int_0^t \tilde{z}_s^2\, ds\right) = 0,$$

$$\text{as-lim}\, \tilde{p}_t^{\mathrm{T}}(Q_t^{-1}/t)\,\tilde{p}_t = (\text{as-lim}\, r_t/t)(\text{as-lim}\, \tilde{p}_t^{\mathrm{T}} Q_t^{-1}\tilde{p}_t r_t^{-1}) = 0.$$

By (5) above as-lim $Q_t^{-1}/t > 0$, hence as-lim $\tilde{p}_t = 0$. ∎

5.2. *Proof of* 3.4. (1) Let $\tilde{p} : \Omega \times T \to R^n$, $\tilde{p}_t = \hat{p}_t - p$, $\tilde{z} : \Omega \times T \to R$,

$$\tilde{z}_t = \hat{z}_t - \hat{\hat{z}}_t,$$

$u : \Omega \times T \to R$,

$$u_t = \tfrac{1}{2}\,\tilde{p}_t^{\mathrm{T}}\tilde{p}_t + r_t^{-1}\int_0^t \tilde{z}_s^2\sigma^{-2}\, ds.$$

Elementary calculations then show that

$$d\tilde{p}_t = h_t r_t^{-1}\sigma^{-2}[\tilde{z}_t\, dt + d\bar{v}_t],$$

$$\tilde{z}_t = \hat{z}_t - \hat{\hat{z}}_t = -h_t^{\mathrm{T}}\tilde{p}_t,$$

$$du_t = \tfrac{1}{2}h_t^{\mathrm{T}}h_t r_t^{-2}\sigma^{-2}\, dt - \left(\int_0^t \tilde{z}_s^2\sigma^{-2}\, ds\right) r_t^{-2}h_t^{\mathrm{T}}h_t\sigma^{-2}\, dt + dm_t,$$

where $(m_t, F_t, t \in T) \in M_{1\text{uloc}}$.

(2) Let $k : \Omega \times T \to R$,

$$dk_t = h_t^{\mathrm{T}}h_t r_t^{-2}\sigma^{-2}\, dt = r_t^{-2}\, drt = -dr_t^{-1}, \qquad k_0 = 0.$$

Then

$$k_t = 1 - r_t^{-1} \leqslant 1,$$

$$\text{as-}\lim_{t\to\infty} k_t \leqslant 1.$$

From 4.1 then follows that

$$\text{as-lim}\, u_t \quad \text{exists in} \quad R_+,$$

$$\text{as-lim} \int_0^t \left(\int_0^\tau \tilde{z}_s^2\sigma^{-2}\, ds\right) r_\tau^{-2}h_\tau^{\mathrm{T}}h_\tau\sigma^{-2}\, d\tau < \infty.$$

(3) As in the proof of 3.2 one shows that

$$\text{as-lim } r_t/t = \text{as-lim } t^{-1} \int_0^t h_s^{\mathrm{T}} h_s \sigma^{-2}\, ds$$

$$= \sigma^{-2} E[h_t^{\mathrm{T}} h_t > 0,$$

$$\text{as-lim} \int_0^t h_s^{\mathrm{T}} h_s r_s^{-1} \sigma^{-2}\, ds = \infty, \qquad \text{as-lim } r_t^{-1} \int_0^t \tilde{z}_s^2 \sigma^{-2}\, ds = 0.$$

Then

$$\text{as-lim } t^{-1} \int_0^t \tilde{z}_s^2\, ds$$

$$= (\text{as-lim } r_t/t) \left(\text{as-lim } r_t^{-1} \int_0^t \tilde{z}_s^2\, ds\right) = 0. \quad \blacksquare$$

6. Conclusion

The adaptive stochastic filtering problem for Gaussian systems has been considered. For the autoregressive case two algorithms have been presented for which almost sure convergence results have been derived.

In addition a rather general convergence theorem has been stated and proved. This result may be used to establish almost sure convergence for adaptive stochastic filtering problems and adaptive stochastic control problems. This result is also applicable when point-process systems are considered, rather than Gaussian systems.

Future research efforts will be concentrated on synthesizing and establishing convergence for other classes of stochastic systems. The recursive maximum likelihood method is currently under investigation.

References

1. A. Bagchi, Continuous time systems identification with unknown noise covariance, *Automatica–J. IFAC* **11** (1975), 533–536.
2. A. Bagchi, Consistent estimates of parameters in continuous time systems, *in* "Anaysis and Optimization of Stochastic Systems" (O. L. R. Jacobs, M. H. A. Davis, M. A. H. Dempster, C. J. Harris, P. C. Parks, Eds.), pp. 437–450, Academic Press, New York, 1980.
3. A. V. Balakrishnan, "Stochastic Differential Systems I, Filtering and Control, A Function Space Approach," Lecture Notes in Economics and Mathematical Systems, Vol. 84, Springer-Verlag, Berlin, 1973.

4. I. V. BASAWA AND B. L. S. PRAKASA RAO, Asymptotic inference for stochastic processes, *Stochastic Process. Appl.* **10** (1980), 221–254.
5. CHEN HAN FU, Quasi-least-squares identification and its strong consistency, *Internat. J. Control* **34** (1981), 921–936.
6. C. DELLACHERIE AND P. A. MEYER, "Probabilités et Potentiel," Chapitres I–IV, Hermann, Paris, 1975; English translation: "Probabilities and Potential," North-Holland, Amsterdam, 1978.
7. C. DELLACHERIE AND P. A. MEYER, "Probabilités et Potentiel," Chap. V–VIII, Hermann, Paris, 1980.
8. C. DOLÉANS-DADE AND P. A. MEYER, Equations differentielles stochastiques, *in* "Séminaire de Probabilités XI," Lecture Notes in Mathematics (C. Dellacherie, P. A. Meyer, M. Weil, Eds.), Vol. 581, pp. 376–382, Springer-Verlag, Berlin, 1977.
9. P. FAURRE, M. CLERGET, AND F. GERMAIN, "Operateurs Rationnels Positifs," Dunod, Paris, 1979.
10. G. C. GOODWIN, P. J. RAMADGE, AND P. E. CAINES, A globally convergent adaptive predictor, *Automatica–J. IFAC* **17** (1981), 135–140.
11. Y. D. LANDAU, „Adaptive Control," Dekker, New York, 1979.
12. Y. D. LANDAU, Near supermartingales for convergence analysis of recursive identification and adaptive control schemes, *Internat. J. Control* **35** (1982), 197–226.
13. R. S. LIPTSER AND A. N. SHIRYAYEV, "Statistics of Random Processes, I General Theory, II Applications," Springer-Verlag, Berlin, 1977, 1978.
14. J. NEVEU, "Martingales a Temp Discrets," Masson, Paris, 1972; English translation: "Discrete Parameter Martingales," North-Holland, Amsterdam, 1975.
15. H. ROBBINS AND D. SIEGMUND, A convergence theorem for non negative almost supermartingales and some applications, *in* "Optimizing Methods in Statistics," (J. S. Rustagi, Ed.), pp. 233–256, Academic Press, New York, 1971.
16. V. SOLO, The convergence of AML, *IEEE Trans. Automat. Control* **24** (1979), 958–962.
17. V. SOLO, The convergence of an instrumental-variable-like recursion, *Automatica–J. IFAC* **17** (1981), 545–547.
18. J. H. VAN SCHUPPEN, Adaptive stochastic filtering problems—the continuous-time case, *in* "Proceedings 2nd Bad Honnef Workshop on Stochastic Differential Systems," June 1982, to appear.
19. E. WONG, "Stochastic Processes in Information and Dynamical Systems," McGraw–Hill, New York, 1977.
20. P. C. YOUNG, Parameter estimation for continuous-time models-a survey, *Automatica–J. IFAC* **17** (1981), 23–29.