Asymptotic Behavior of the Extended Kalman Filter as a Parameter Estimator for Linear Systems

LENNART LJUNG, MEMBER, IEEE

Abstract—**The extended Kalman filter is an approximate filter for nonlinear systems, based on first-order linearization. Its use for the joint parameter and state estimation problem for linear systems with unknown parameters is well known and widely spread. Here a convergence analysis of this method is given. It is shown that in general, the estimates may be biased or divergent and the causes for this are displayed. Some common special cases where convergence is guaranteed are also given. The analysis gives insight into the convergence mechanisms and it is shown that with a modification of the algorithm, global convergence results can be obtained for a general case. The scheme can then be interpreted as maximization of the likelihood function for the estimation problem, or as a recursive prediction error algorithm.**

I. Introduction

NONLINEAR filtering is an important and well-studied field of estimation and control theory. Many different approaches have been taken, among which perhaps the extended Kalman filter (EKF) is the best-known one. It is based on linearization of the state equations at each time step and on the use of linear estimation theory (the Kalman filter).

A description of the EKF is given in Jazwinski [1, Theorem 8.1], and for future reference we shall immediately give a brief account of the algorithm. Let the nonlinear, discrete-time system be given by

$$\begin{aligned}\xi(t+1)&=f(t,\xi(t))+w(t)\\ y(t)&=h(t,\xi(t))+v(t).\end{aligned}\tag{1.1}$$

The EKF estimate of the state $\xi(t+1)$, based upon observations $y(0),\cdots,y(t)$ is denoted by $\hat{\xi}(t+1)$ and obtained recursively by

$$\hat{\xi}(t+1)=f(t,\hat{\xi}(t))+N(t)\left[y(t)-h(t,\hat{\xi}(t))\right]\tag{1.2}$$

where $N(t)$ is given by

$$\begin{aligned}N(t)=F(t,\hat{\xi}(t))P(t)H^T(t,\hat{\xi}(t))&\\ \times\left[H(t,\hat{\xi}(t))P(t)H^T(t,\hat{\xi}(t))+Q^v(t)\right]^{-1}&\end{aligned}\tag{1.3}$$

$$\begin{aligned}P(t+1)=F(t,\hat{\xi}(t))P(t)F^T(t,\hat{\xi}(t))+Q^w(t)&\\ -N(t)\left[Q^v(t)+H(t,\hat{\xi}(t))\right.&\\ \left.\cdot P(t)H^T(t,\hat{\xi}(t))\right]N^T(t)&\end{aligned}\tag{1.4}$$

and

$$\begin{aligned}F(t,\hat{\xi})&=\frac{\partial}{\partial\xi}f(t,\xi)\Big|_{\xi=\hat{\xi}}\\ H(t,\hat{\xi})&=\frac{\partial}{\partial\xi}h(t,\xi)\Big|_{\xi=\hat{\xi}}\\ Q^v(t)&=Ev(t)v(t)^T;\\ Q^w(t)&=Ew(t)w(t)^T.\end{aligned}\tag{1.5}$$

In [1] this algorithm is given for a continuous-time system with discrete-time measurements. Different variants with relinearizations made iteratively within each recursion are also discussed. A very simple and natural feature is to make the recursion (1.2) in two steps as a measurement update and a time update and to make a relinearization in between. While such features may have a major influence on transient behavior and effective convergence rate of the algorithm, they will not effect the convergence results to be discussed here, and hence we shall not have to go into detail with them.

Recursive identification algorithms for dynamic systems is another topic of considerable current interest, cf. [2]–[4]. Most suggested algorithms of this kind seem to have their origins in some off-line identification procedure or to be based on stochastic approximation considerations. The joint state and parameter estimation problem can of course be understood as a state estimation problem for a nonlinear system. In general there is not much to be gained from such an approach, since then much of the structure of the identification problem is lost. However, if

Manuscript received April 10, 1978; revised September 18, 1978. Paper recommended by A. H. Haddad, Past Chairman of the Estimation Committee.

The author is with the Department of Electrical Engineering, Linköping University, Linköping, Sweden.

0018-9286/79/0200-0036$00.75 ©1979 IEEE

the EKF is applied to this particular problem, interesting algorithms are obtained.

The EKF approach to the estimation of parameters in dynamical systems has a rather long history, and a considerable number of applications of this method has been reported. The approach seems to have been first suggested and discussed in [5] and [6] and among the many papers dealing with various aspects we may mention [7]–[12].

There are several reasons for the popularity of this method. It is a fairly natural thing to include the unknown parameters in the state vector, and once this is done, standard Kalman filter programs can be applied for the estimation. The algorithm is consequently not very complex. A nice and natural approach to the joint state and parameter estimation problem is also obtained.

In view of this, and of the number of applications made, amazingly little analysis of the method has been performed. The author is not aware of any systematic convergence analysis of the EKF for this application. Instead, the collected experiences from simulations and applications to real data seem to have been condensed into "rumors" about the convergence properties. It is thus known that the method may give biased estimates, e.g., [11], and that it does not seldom diverge if the initial estimates are not sufficiently good, e.g., [9].

The objective of the present paper is to give a fairly systematic and comprehensive treatment of the EKF, when applied to parameter estimation for linear stochastic systems. The analysis is based on the methods of [13]. It will lead to a certain understanding of the causes of bias and divergence. The mechanisms for parameter adjustments will be exposed and in that way the relationship between the EKF and other recursive parameter estimation methods is clarified. A fairly general convergence result for deterministic models will be shown. Perhaps the most important feature is that the insight gained into the convergence mechanism directly leads to a slightly modified version for models of innovation representation character, which has excellent global convergence properties.

II. The System

The measured input–output data

$$u(0), y(0), u(1), y(1), \cdots$$

will throughout this paper be assumed to be obtained from the system

$$\begin{aligned} x(t+1) &= A_0 x(t) + B_0 u(t) + v(t) \\ y(t) &= C_0 x(t) + e(t) \end{aligned} \tag{2.1}$$

where $u(t)$, $y(t)$, and $x(t)$ are vectors of dimensions n_u, n_y, and n_x, respectively. The sequences $\{v(t)\}$ and $\{e(t)\}$ consist of independent random vectors with zero-means and covariances

$$\begin{aligned} Ev(t)v(s)^T &= Q_0^v \delta_{ts} \\ Ee(t)e(s)^T &= Q_0^e \delta_{ts} \\ Ev(t)e(s)^T &= Q_0^c \delta_{ts}. \end{aligned} \tag{2.2}$$

Furthermore, it is assumed that the initial state $x(0)$ is a random vector with zero-mean and covariance matrix Π_0. It is independent of future values of $\{v(t)\}$ and $\{e(t)\}$ $t \geqslant 0$. All the matrices A_0, B_0, C_0, Q_0^v, Q_0^e, Q_0^c are assumed to be time invariant.

In some cases we shall consider a time series, i.e., the input signal is absent, corresponding to $B_0 = 0$. If an input is present we shall, for technical reasons, assume that it is a weakly stationary stochastic process with rational spectral density, and hence can be understood as obtained from white noise by linear, exponentially stable filtering. Furthermore, we shall, again for technical reasons, assume that all absolute moments of the stochastic processes $\{u(t)\}$, $\{v(t)\}$, and $\{e(t)\}$ exist and are bounded.

Remark: These assumptions on the stochastic processes are introduced in order to be able to apply the "B-conditions" of [13]. With a similar technique, using instead the "C-conditions" of [13], less restrictive assumptions about these processes may be introduced. This is treated in [21].

The system (2.1) with (2.2) is assumed to be (partly) unknown to the user. The problem he is faced with is to determine the matrices A_0, B_0, C_0 and possibly also Q_0^v, Q_0^e, and Q_0^c together with the state estimates, based on measurements of input–output data.

If these matrices are all known, then the linear least-squares state estimate for (2.1) is obtained from the familiar Kalman filter:

$$\hat{x}_0(t+1) = A_0 \hat{x}_0(t) + B_0 u(t) + K_0(t)\left[y(t) - C_0 \hat{x}_0(t)\right] \tag{2.3}$$

$$\hat{x}_0(0) = 0$$

where

$$K_0(t) = \left[A_0 P_0(t) C_0^T + Q_0^c\right]\left[C_0 P_0(t) C_0^T + Q_0^e\right]^{-1} \tag{2.4}$$

$$\begin{aligned} P_0(t+1) = {} & A_0 P_0(t) A_0^T + Q_0^v \\ & - K_0(t)\left[C_0 P(t) C_0^T + Q_0^e\right] K_0^T(t) \end{aligned} \tag{2.5}$$

$$P_0(0) = \Pi_0.$$

Let

$$\bar{K}_0 = \lim_{t \to \infty} K_0(t). \tag{2.6}$$

III. The algorithm for a General Linear State Model

A. The Model

In order to determine the system that has generated the observed input–output data, we assume the following model for it:

$$
\begin{aligned}
x(t+1)&=A(\theta)x(t)+B(\theta)u(t)+v_\theta(t)\\
y(t)&=C(\theta)x(t)+e_\theta(t)
\end{aligned}
\tag{3.1}
$$

where

$$
\begin{aligned}
Ev_\theta(t)v_\theta^T(s)&=Q^v(\theta)\delta_{ts},\\
Ee_\theta(t)e_\theta^T(s)&=Q^e(\theta)\delta_{ts},\\
Ev_\theta(t)e_\theta^T(s)&=Q^c(\theta)\delta_{ts},\\
Ex(0)&=0,\\
Ex(0)x^T(0)&=\Pi(\theta).
\end{aligned}
\tag{3.2}
$$

The matrices $A(\theta)$, $B(\theta)$, $C(\theta)$, $Q^v(\theta)$, $Q^e(\theta)$, and $Q^c(\theta)$ depend on a finite-dimensional parameter vector θ in an arbitrary way. It is assumed, though, that the matrix elements are differentiable with respect to θ. Usually, the noise characteristics matrices Q^v, Q^e, and Q^c do not depend on θ, but are chosen fixed in some ad hoc way, most often with $Q^c=0$. This corresponds to the fact that in (1.1), the noise characteristics are independent of the state.

We shall in the remainder of this section assume that $\{e_\theta(t)\}$ and $\{v_\theta(t)\}$ are independent of θ (and drop this index). This is of no importance as such; the reader may easily append this dependence. It, however, touches a fundamental issue, that of how the linearization should be done, and we shall return to this question in Sections VII and VIII.

It could be remarked that the converse situation is also widely discussed in the literature; that of known dynamics and unknown noise statistics in (3.1), (3.2). This case may also cover the approach to describe modeling errors in the dynamics as additive disturbances. It is usually referred to as "adaptive filtering," and among many papers dealing with different aspects of this problem, [14]–[17] could be mentioned.

B. The Algorithm

The EKF approach to determine the unknown parameter vector θ now is obtained by extending the state vector x with the parameter vector $\theta=\theta(t)$.

$$
z(t)=\begin{pmatrix} x(t)\\ \theta(t)\end{pmatrix}.
\tag{3.3}
$$

We then have the following state equation

$$
\begin{aligned}
z(t+1)&=f(z(t),u(t))+\begin{pmatrix} v(t)\\ 0\end{pmatrix}\\
y(t)&=h(z(t))+e(t)
\end{aligned}
\tag{3.4}
$$

where

$$
f(z(t),u(t))=\begin{bmatrix} A(\theta)x(t)+B(\theta)u(t)\\ \theta\end{bmatrix}
\tag{3.5}
$$

$$
h(z(t))=C(\theta)x(t).
\tag{3.6}
$$

We are consequently faced with a nonlinear filtering problem and if this is attacked by the EKF (1.2)–(1.5) we obtain

$$
\hat z(t+1)=f(\hat z(t),u(t))+N(t)[y(t)-h(\hat z(t))]
\tag{3.7}
$$

$$
\hat z(0)=\hat z_0
$$

$$
\begin{aligned}
N(t)=&\left[F(\hat z(t),u(t))\bar P(t)H^T(\hat z(t))+\bar Q^c\right]\\
&\times\left[H(\hat z(t))\bar P(t)H^T(\hat z(t))+Q^e\right]^{-1}
\end{aligned}
\tag{3.8}
$$

$$
\begin{aligned}
\bar P(t+1)=&F(\hat z(t),u(t))\bar P(t)F^T(\hat z(t),u(t))+\bar Q^v\\
&-N(t)\left[H(\hat z(t))\bar P(t)H^T(\hat z(t))+Q^e\right]N^T(t)
\end{aligned}
\tag{3.9}
$$

$$
\bar P(0)=\bar P_0
$$

where

$$
\begin{aligned}
F(\hat z(t),u(t))&=\frac{\partial}{\partial z}f(z,u)\bigg|_{z=\hat z(t)}\\
&=\left[\begin{array}{c:c} A(\hat\theta(t)) & M(\hat\theta(t),\hat x(t),u(t))\\ \hdashline 0 & I\end{array}\right]\\
H(\hat z(t))&=\frac{\partial}{\partial z}h(z)\bigg|_{z=\hat z(t)}\\
&=\left[\begin{array}{c:c} C(\hat\theta(t)) & D(\hat\theta(t),\hat x(t))\end{array}\right]
\end{aligned}
\tag{3.11}
$$

$$
\begin{aligned}
\bar Q^v&=\begin{bmatrix} Q^v & 0\\ 0 & 0\end{bmatrix}\\
\bar Q^c&=\begin{bmatrix} Q^c\\ 0\end{bmatrix}\\
\hat z_0&=\begin{bmatrix} 0\\ \hat\theta_0\end{bmatrix}\\
\bar P_0&=\begin{pmatrix} \Pi(\hat\theta_0) & 0\\ 0 & \Sigma_0\end{pmatrix}.
\end{aligned}
$$

Here

$$
M(\hat\theta,\hat x,u)=\frac{\partial}{\partial\theta}(A(\theta)\hat x+B(\theta)u)\bigg|_{\theta=\hat\theta}(a\ n_x|n_\theta\ \text{matrix})
\tag{3.12}
$$

$$
D(\hat\theta,\hat x)=\frac{\partial}{\partial\theta}(C(\theta)\hat x)\bigg|_{\theta=\hat\theta}(a\ n_y|n_\theta\ \text{matrix}).
\tag{3.13}
$$

These functions are of course linear in $\hat x$ and u, and depend in an essential way on the parametrization.

$\hat\theta_0$ and Σ_0 represent some *a priori* information about the parameter vector θ. Common choices are $\hat\theta_0=0$ and $\Sigma_0=100\cdot$(variance of y) if no *a priori* information is available.

Introduce for short

$$
\begin{aligned}
M_t&=M(\hat\theta(t),\hat x(t),u(t))\\
D_t&=D(\hat\theta(t),\hat x(t))\\
A_t&=A(\hat\theta(t))\\
B_t&=B(\hat\theta(t))\\
C_t&=C(\hat\theta(t))\\
S_t&=\left[(C_t \,\vdots\, D_t)\bar P(t)(C_t \,\vdots\, D_t)^T+Q^e\right].
\end{aligned}
$$

Introduce also the natural block structure

$$N(t)=\left[\begin{array}{c} K(t) \\ \hdashline L(t) \end{array}\right]$$

$$\bar{P}(t)=\left[\begin{array}{c:c} P_1(t) & P_2(t) \\ \hdashline P_2^T(t) & P_3(t) \end{array}\right].$$

Equations (3.7)–(3.9) can now be rewritten explicitly as

$$\hat{x}(t+1)=A_t\hat{x}(t)+B_tu(t)+K(t)[y(t)-C_t\hat{x}(t)] \tag{3.14}$$

$$\hat{x}(0)=0$$

$$\hat{\theta}(t+1)=\hat{\theta}(t)+L(t)[y(t)-C_t\hat{x}(t)] \tag{3.15}$$

$$\hat{\theta}(0)=\hat{\theta}_0$$

$$K(t)=[A_tP_1(t)C_t^T+M_tP_2^T(t)C_t^T+A_tP_2(t)D_t^T + M_tP_2(t)D_t^T+Q^c]S_t^{-1} \tag{3.16a}$$

$$S_t=C_tP_1(t)C_t^T+C_tP_2(t)D_t^T + D_tP_2^T(t)C_t^T+D_tP_3(t)D_t^T+Q^e \tag{3.16b}$$

$$L(t)=[P_2^T(t)C_t^T+P_3(t)D_t^T]S_t^{-1} \tag{3.17}$$

$$P_1(t+1)=A_tP_1(t)A_t^T+A_tP_2(t)M_t^T + M_tP_2^T(t)A_t^T+M_tP_3(t)M_t^T - K(t)S_tK^T(t)+Q^v \tag{3.18}$$

$$P_1(0)=\Pi_0(\hat{\theta}_0)$$

$$P_2(t+1)=A_tP_2(t)+M_tP_3(t)-K(t)S_tL^T(t) \tag{3.19}$$

$$P_2(0)=0$$

$$P_3(t+1)=P_3(t)-L(t)S_tL^T(t) \tag{3.20}$$

$$P_3(0)=\Sigma_0.$$

Note that (3.19) with the aid of (3.17) also can be written as

$$P_2(t+1)=(A_t-K(t)C_t)P_2(t) + (M_t-K(t)D_t)P_3(t). \tag{3.19$'$}$$

It could be mentioned that certain numerical problems will arise in the algorithm (3.14)–(3.20) if $M_t-K(t)D_t$ is not a full rank stochastic process (i.e., its covariance matrix is singular). This is a question associated with the parameterization of the model (3.1). In such a case, there are also certain technical problems in the analytic treatment. Therefore, we shall in the sequel assume that some measures are taken to come around these numerical problems. A simple way is to replace (3.20) by

$$P_3(t+1)=\left[\{P_3(t)-L(t)S_tL^T(t)\}^{-1}+\delta I\right]^{-1}, \tag{3.20$'$}$$

for some small positive δ. Notice that for small δ (3.20′) is approximately given by

$$P_3(t+1)=P_3(t)-L(t)S_tL^T(t)-\delta P_3(t)P_3(t).$$

IV. The Method of Analysis—An Associated Differential Equation

Convergence of the algorithm (3.14)–(3.20′) will here be analyzed using the theory of [13]. In that reference it is shown how convergence of recursive, stochastic algorithms can be analyzed in terms of the stability properties of an associated differential equation.

We shall in this section determine the differential equation that is associated with the algorithm (3.14)–(3.20′) and show that the regularity conditions of [13] are satisfied.

The differential equation is defined in terms of the processes that (3.14)–(3.20′) would produce if the model parameter were kept constant $=\theta$. This means that (3.15) would be replaced by $\hat{\theta}(t)=\theta$. Consequently, in the θ-dependent matrices $[A_t, B_t, C_t, M_t,$ and $D_t]$ the estimate $\hat{\theta}(t)$ should be replaced by θ. It is easy to see (and it will be shown later) that then P_2 and P_3 would tend to zero and that $P_1(t)$, S_t, and $K(t)$ would tend to $\bar{P}_1(\theta)$, $\bar{S}(\theta)$, and $\bar{K}(\theta)$, respectively, which are given as the solutions of

$$\bar{P}_1(\theta)=A(\theta)\bar{P}_1(\theta)A^T(\theta)+Q^v-\bar{K}(\theta)\bar{S}(\theta)\bar{K}^T(\theta) \tag{4.1a}$$

$$\bar{S}(\theta)=C(\theta)\bar{P}_1(\theta)C^T(\theta)+Q^e \tag{4.1b}$$

$$\bar{K}(\theta)=[A(\theta)\bar{P}_1(\theta)C^T(\theta)+Q^c]\bar{S}^{-1}(\theta). \tag{4.1c}$$

Existence of these solutions follows, provided the model $\{A(\theta),B(\theta),C(\theta),Q^v,Q^c\}$ satisfies certain detectability and stabilizability conditions.

Then define the process $\bar{x}(t;\theta)$ as the estimates that would be obtained with this constant model; corresponding to the parameter value θ:

$$\bar{x}(t+1;\theta)=A(\theta)\bar{x}(t;\theta)+B(\theta)u(t)+\bar{K}(\theta)\bar{\epsilon}(t;\theta) \tag{4.2}$$

where

$$\bar{\epsilon}(t;\theta)=y(t)-C(\theta)\bar{x}(t;\theta). \tag{4.3}$$

Furthermore, let $\bar{w}(t;\theta)$ be defined by

$$\bar{w}(t+1;\theta)=[A(\theta)-\bar{K}(\theta)C(\theta)]\bar{w}(t;\theta) + \left[M(\theta,\bar{x}(t;\theta),u(t))-\bar{K}(\theta)D(\theta,\bar{x}(t;\theta))\right]. \tag{4.4}$$

Recall that M and D were defined in (3.12) and (3.13), respectively. We see by comparing (3.19′) and (4.4) that $\bar{w}(t,\theta)\tilde{P}_3$ is the P_2-process (3.19′) would produce for given constant θ and $\tilde{P}_3$. Let the $n_\theta|n_y$ matrix $\bar{\psi}(t,\theta)$ be given by

$$\bar{\psi}(t;\theta)=\left[C(\theta)\bar{w}(t;\theta)+D(\theta,\bar{x}(t;\theta))\right]^T. \tag{4.5}$$

Again, by comparing (4.5) with (3.17), having in mind that $P_2\sim\bar{w}(t,\theta)\tilde{P}_3$, we see that the L-process that (3.17) would produce for constant θ and $\tilde{P}_3$ would be $\tilde{P}_3\bar{\psi}(t,\theta)\bar{S}^{-1}(\theta)$.

Equations (4.1)–(4.5) now define the processes $\bar{\psi}(t;\theta)$ and $\bar{\epsilon}(t;\theta)$ uniquely, for the given θ, from $y(s),u(s)$; $s\leqslant t$.

(Take the initial conditions $\bar{x}(0;\theta)$, $\bar{w}(0,\theta)$ such that the processes become stationary.) Since, according to (3.15), $L(t)[y(t)-C_t\hat{x}(t)]$ updates $\hat{\theta}$ and since $L_2 \sim \bar{P}_3\bar{\psi}(t,\theta)\bar{S}^{-1}(\theta)$, it seems reasonable that this update, asymptotically should be related to the vector

$$f(\theta)=E\bar{\psi}(t;\theta)\bar{S}^{-1}(\theta)\bar{\epsilon}(t;\theta) \quad (n_\theta|1 \text{ matrix}) \tag{4.6}$$

where "E" denotes expectation with respect to the stochastic processes $y(s)$ and $u(s)$. Define also

$$G(\theta)=E\bar{\psi}(t;\theta)\bar{S}^{-1}(\theta)\bar{\psi}^T(t;\theta) \quad (n_\theta|n_\theta \text{ matrix}). \tag{4.7}$$

We may thus interpret $f(\theta)$ as the direction (modified by $R^{-1}=\bar{P}_3$; see below) in which the estimates asymptotically are adjusted.

Thus far, what we have done is given a formal definition of the functions $f(\theta)$ and $G(\theta)$ via (4.1)–(4.7). We have also given intuitive arguments as to why the function $f(\theta)$ should be related to the asymptotic properties of the $\hat{\theta}(t)$-sequence. The latter result indeed holds, which is proved in the following lemma, which is the basic result for the convergence analysis.

Lemma 4.1: Consider the differential equation given by

$$\frac{d}{d\tau}\theta(\tau)=R^{-1}(\tau)f(\theta(\tau)) \tag{4.8a}$$

$$\frac{d}{d\tau}R(\tau)=G(\theta(\tau))+\delta I-R(\tau). \tag{4.8b}$$

Let $D_s=\{\theta|(A(\theta),Q^v)$ stabilizable and $(A(\theta),C(\theta))$ detectable$\}$. Let $\{\hat{\theta}(t),\hat{x}(t)\}$ be given by algorithm (3.14)–(3.20′).

1) Suppose that the differential equation (4.8) has an invariant set D_c with domain of attraction $\theta \in D_A$ (which will be a subset of D_s). Suppose further that $\hat{\theta}(t)$ belongs to a compact subset of D_A *and* $\hat{x}(t)$ is bounded infinitely often with probability one (with probability 1). Then

$$\hat{\theta}(t)\rightarrow D_c \quad \text{with probability 1 as } t\rightarrow\infty. \tag{4.9}$$

2) Suppose that $\hat{\theta}(t)\rightarrow\theta^*$ with probability greater than zero. Then θ^* must be a stable stationary point of the differential equation (4.8).

3) Let $\bar{D}$ be a compact subset of D_s such that the trajectories of (4.8) that start in $\bar{D}$ do not leave $\bar{D}$. Suppose that the estimates $\hat{\theta}(t)$ are projected into $\bar{D}$ and that (4.8) has an invariant set D_c with a domain of attraction $D_A\supset\bar{D}$. Then $\hat{\theta}(t)\rightarrow D_c$ with probability 1 as $t\rightarrow\infty$.

Remark: An invariant set of a differential equation is a set, such that the trajectories remain in there for $-\infty<\tau<\infty$. The domain of attraction of an invariant set D_c consists of those points from which the trajectories converge into D_c as τ tends to infinity. For further comments on the formulation of the lemma, see [13].

Proof: The lemma follows directly from [13, Theorems 1, 2, 4], once it is verified that the algorithm (3.14)–(3.20′) satisfies the regularity conditions of these theorems. This is verified in Appendix I.

With the aid of this lemma the convergence analysis is effectively reduced to stability analysis of the differential equation (4.8). The next section is devoted to such analysis.

V. Convergence Analysis

A. Stationary Points of the Differential Equation and the Question of Bias

In the differential equation (4.8)

$$\frac{d}{d\tau}\theta(\tau)=R^{-1}(\tau)f(\theta(\tau)) \tag{5.1a}$$

$$\frac{d}{d\tau}R(\tau)=G(\theta(\tau))+\delta I-R(\tau) \tag{5.1b}$$

the function $f(\cdot)$ represents the "corrective force." The positive definite matrix $R^{-1}(\tau)$ only modifies the direction of correction. We shall later interpret (5.1) as a descent algorithm, and then R^{-1} yields a modification from a negative gradient direction $f(\cdot)$ to an approximate Newton direction.

It can also be seen that if $\theta(\tau)$ converges to θ^* then $R(\tau)$ tends to $G(\theta^*)+\delta I$.

Now, as seen from the definition of $f(\theta)$ in (4.7), this function is the correlation between the residuals (innovations) $\bar{\epsilon}(t;\theta)$ obtained from model θ and the variable $\bar{\psi}(t;\theta)$. This random variable is according to (4.4) and (4.5) obtained by linear filtering of the state estimates corresponding to this same model θ. Therefore, $f(\theta)$ is a measure of the correlation between the current innovation and previous state estimates. These in turn can be obtained from the previous residuals; see (4.2). Therefore, it is clear that $f(\theta)$ measures the correlation of the sequence $\{\bar{\epsilon}(t;\theta)\}$.

If this sequence is uncorrelated for some $\theta=\theta^*$ (which means that θ^* gives a "good" model), then $f(\theta^*)=0$ and $\theta=\theta^*$, $R=G(\theta^*)+\delta I$ is a stationary point of (5.1). The EKF therefore has certain relations to adaptive filtering techniques, for which monitoring the correlation of the innovations often is the chief tool, see, e.g., [15]–[17]. Notice, however, that the converse is not necessarily true, i.e., $f(\theta^*)=0$ does not in general imply that $\{\bar{\epsilon}(t;\theta^*)\}$ is a sequence of uncorrelated random vectors. It merely implies that $\bar{\epsilon}(t;\theta^*)$ is orthogonal to certain sets of linear combinations of $\bar{\epsilon}(s;\theta^*)$, $s<t$. If model orders and structures are chosen suitably, this in turn may imply orthogonality of $\{\bar{\epsilon}(t;\theta^*)\}$, and some such choices will be discussed below.

Now, suppose that for some θ_0,

$$\begin{aligned} A_0&=A(\theta_0)\\ B_0&=B(\theta_0)\\ C_0&=C(\theta_0) \end{aligned} \tag{5.2}$$

that is the true system matrices (2.1) are obtained for a certain parameter vector. Then $\{\bar{\epsilon}(t;\theta_0)\}$ will in general be an orthogonal sequence only if in addition

$$\overline{K}_0 = \overline{K}(\theta_0) \tag{5.3}$$

where $\overline{K}_0$ is defined by (2.6), and $\overline{K}(\theta_0)$ by (4.1). The condition (5.3) holds only if the assumptions (3.2) about the noise structure of the model (3.1) are in accordance with those of the true system (2.2). Consequently, a value θ_0 corresponding to the true system, i.e., satisfying (5.2) will in general be a stationary point of (5.1) and hence a possible convergence point only if the assumed noise structure coincides with the true one. Otherwise the estimates will be biased.

It is of course somewhat unrealistic to assume that the noise structure of the system is known, while the dynamics are unknown. Therefore, if the noise characteristics of the model (3.1) is chosen ad hoc (as, apparently, is usually done) then the system parameter estimates will in general be biased. We have consequently arrived at the perhaps trivial observation that the cause of the bias does not lie in the EKF-method itself, but comes from incorrect noise assumptions associated with the model.

B. Local Stability Analysis

The local convergence properties of the algorithm are according to Lemma 4.1:2) associated with stability of the linearized differential equation (4.8) around the stationary point in question. In [18] the linearization is carried out. There does not seem to be any guarantee that the linearized equation should be stable for all systems. Therefore, the possible lack of convergence of the EKF method does not necessarily have to be caused by large parameter deviations from the true values, as is sometimes claimed.

C. Global Convergence Analysis—Simple Example

Let the true system be given by

$$\begin{aligned} x(t+1) &= a_0 x(t) + v_0(t) \\ y(t) &= x(t) + e_0(t) \end{aligned} \tag{5.4}$$

where x and y are scalars and

$$\begin{aligned} Ev_0(t)v_0(s) &= \delta_{ts}\lambda \\ Ee_0(t)e_0(s) &= \delta_{ts} \\ Ev_0(t)e_0(s) &= 0. \end{aligned} \tag{5.5}$$

Let the model be given by

$$\begin{aligned} x(t+1) &= ax(t) + v(t) \\ y(t) &= x(t) + e(t) \end{aligned} \tag{5.6}$$

where we assume that

$$\begin{aligned} Ev(t)v(s) &= \delta_{ts} \\ Ee(t)e(s) &= \delta_{ts} \\ Ev(t)e(s) &= 0. \end{aligned}$$

To determine the differential equation associated with this problem we first note that, in obvious notation, cf. Section IV:

$$\overline{K}(a) = \left(\frac{a^3}{2} + a\sqrt{\frac{a^4}{4}+1}\right) \Big/ \left(1 + \frac{a^2}{2} + \sqrt{\frac{a^4}{4}+1}\right) \tag{5.7}$$

$$\begin{aligned} M(a,\hat{x},u) &= \hat{x} \\ D(a,\hat{x}) &= 0 \\ \bar{\hat{x}}(t+1;a) &= (a-\overline{K}(a))\bar{\hat{x}}(t,a) + \overline{K}(a)y(t) \\ \overline{w}(t+1;a) &= (a-\overline{K}(a))\overline{w}(t;a) + \bar{\hat{x}}(t;a) \\ \overline{S}(a) &= 1 + \frac{a^2}{2} + \sqrt{\frac{a^4}{4}+1}\ . \end{aligned}$$

With the backward shift operator $q^{-1}(q^{-1}x(t) = x(t-1))$, we obtain

$$\begin{aligned} \overline{w}(t;a) &= \left(1-q^{-1}(a-\overline{K}(a))\right)^{-2}\overline{K}(a)q^{-2}y(t) \\ \bar{\epsilon}(t;a) &= 1-\left(1-q^{-1}(a-\overline{K}(a))\right)^{-1}\overline{K}(a)q^{-1}y(t) \\ &= \frac{1-q^{-1}a}{1-q^{-1}(a-\overline{K}(a))}y(t) \end{aligned}$$

and, evaluating the expected value using complex integrals

$$\begin{aligned} E\overline{w}(t;a)\bar{\epsilon}(t;a) &\triangleq \tilde{f}(a) \\ &= \frac{1}{2\Pi i}\oint \frac{\overline{K}(a)z^2}{\left(1-z(a-\overline{K}(a))\right)^2} \\ &\quad \cdot \frac{1-a/z}{1-(a-\overline{K}(a))/z} \\ &\quad \cdot \Phi_{yy}(z)\frac{dz}{z}. \end{aligned} \tag{5.8}$$

Here $\Phi_{yy}(z)$ is the spectrum of the signal y:

$$\Phi_{yy}(z) = 1 + \frac{\lambda}{(1-za_0)\left(1-\frac{1}{z}a_0\right)}. \tag{5.9}$$

The integral (5.8) with (5.9) is easy to evaluate using residue calculus which gives

$$\tilde{f}(a) = \frac{(a_0-a)(1-a_0f_0)a_0\overline{K}_0\overline{K}(a)}{(1-a_0f)^2(1-a_0^2)(a_0-f)} + \frac{(f-f_0)(1-ff_0)f\overline{K}(a)^2}{(1-f^2)^2(1-fa_0)(a_0-f)} \tag{5.10}$$

where

$$\begin{aligned} \overline{K}_0 &= \frac{a_0\left[(\lambda+a_0^2-1)/2+\sqrt{(\lambda+a_0^2-1)^2/4+\lambda}\ \right]}{1+(\lambda+a_0^2-1)/2+\sqrt{(\lambda+a_0^2-1)^2/4+\lambda}} \\ f &= a-\overline{K}(a) \\ f_0 &= a_0-\overline{K}_0. \end{aligned}$$

The differential equation (5.1a) is now

$$\frac{d}{d\tau}a(\tau)=R^{-1}(\tau)S^{-1}(a(\tau))\tilde{f}(a(\tau)). \tag{5.11}$$

However, since $a(\tau)$ is scalar and $R^{-1}(\tau)S^{-1}(a(\tau))$ is a positive scalar, the stability properties of (5.11) are entirely determined by the sign changes of $\tilde{f}(a)$. Sketches of this function are shown in Fig. 1 for some choices of a_0 and λ.

Several conclusions can be drawn from this figure. Remember that the convergence properties of the EKF-algorithm are determined by the stability properties of (5.11). A sign change of $\tilde{f}(a)$ from plus to minus for increasing a corresponds to a stable stationary point of (5.11) and the domain of attraction extends to the nearest neighboring sign changes. We see that for $a_0=0$ we have global convergence to this value even when $\lambda=10$ (wrong noise assumptions in the model). For a_0 larger than zero and $\lambda=1$ (correct noise assumption) the true value is always a stable stationary point with domain of attraction equal to all positive a. There exist, however, other attraction points for negative values of a, to which we may converge with a nonzero probability. Also, for $a_0=0.4$, 0.6, and 0.8 (in fact also for $a_0=0.2$) the estimate may with a nonzero probability tend to minus infinity. If the system is known to have been obtained from a continuous-time system by sampling, it is natural to exclude negative a, e.g., using a projection facility. Then we have guaranteed convergence, with probability 1 to a_0 according to Lemma 4.1.

Notice also that for the case $\lambda=10$ we obtain biased estimates of a_0, even if we converge to the "best" stationary point.

VI. Deterministic Models (Output Error Methods)

The results of the previous section showed that in general the convergence properties of the EKF are not satisfactory. It turns out that in the special case where $\bar{K}(\theta)$ defined by (4.1) happens to be independent of θ, the convergence properties are much better. The reason for this will become quite clear in the next two sections. Let us first, however, give a result for an interesting special case of this sort.

If, in the model (3.1), $v_\theta(t)$ is assumed to be absent, i.e., $Q^v(\theta)=0$, and $A(\theta)$ is stable for $\theta\in D_s$ then, as seen from (4.1), the stationary solution $\bar{K}(\theta)=0$ for $\theta\in D_s$, and in particular it will not depend on θ. The case $Q^v(\theta)=0$ (all θ) will here be called the case of a *deterministic model*, but it should be noted that we assume a nonzero measurement noise $e_\theta(t)$. In fact, for systems operating in open loop this measurement noise is allowed to be colored, so the model is quite a general one. However, only the input-output dynamics will be modeled, and no information about the noise structure is gained. It is clear that such a model makes sense only if an input $u(t)$ indeed is present. In the case $\bar{K}(\theta)=0$ the state estimate $\bar{x}(t;\theta)$ is given by (see (4.2))

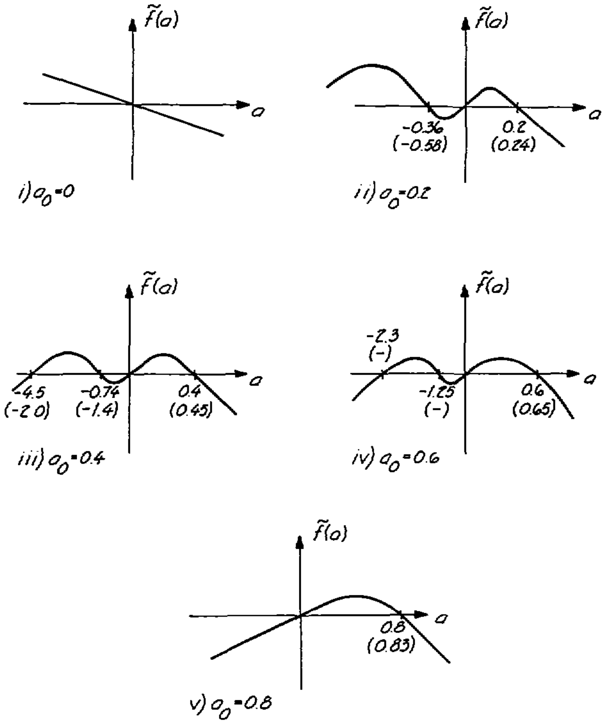

Fig. 1. Sketches of the function $f(a)$ for different a_0. Only sign changes are shown. Numbers in parenthesis correspond to $\lambda=10$, others to $\lambda=1$.

$$\bar{x}(t+1;\theta)=A(\theta)\bar{x}(t;\theta)+B(\theta)u(t). \tag{6.1}$$

Therefore $C(\theta)\bar{x}(t;\theta)$ is the output of model corresponding to the parameter value θ and the difference $y(t)-C(\theta)\bar{x}(t;\theta)$ on which the correction of θ is based, is then the discrepancy between measured output and model output. Such methods are usually called *output error methods* or *model reference identification techniques*. Consequently, the EKF-approach with $Q^v(\theta)=0$ yields a particular output error method that apparently has not been discussed previously. In order to make the "boundedness" condition [the condition preceding (4.9)] satisfied, we assume that the estimates are projected into the set $D_s=\{\theta|A(\theta)$ stable$\}$ as described, e.g., in [13, Section VI.] We now have the following result:

Theorem 6.1: Consider input-output data generated by the system (2.1), (2.2). Let the "deterministic" model be given by

$$\begin{aligned}x(t+1)&=A(\theta)x(t)+B(\theta)u(t)\\ y(t)&=C(\theta)x(t)+e(t)\end{aligned} \tag{6.2}$$

where $\{e(t)\}$ is supposed to be white noise with covariance matrix Q^e. Suppose that the parameter vector θ is estimated by the extended Kalman filter scheme described in Section III [(3.14)–(3.20′)]. Assume further that the algorithm is complemented with a projection facility to keep $\hat{\theta}(t)$ in a compact subset of $\{\theta|A(\theta)$ stable$\}$. Then the estimate $\hat{\theta}(t)$ converges with probability 1 to a stationary point of the function

$$V(\theta)=E\bar{\epsilon}^T(t;\theta)(Q^e)^{-1}\bar{\epsilon}(t;\theta)$$

and among isolated stationary points only local minima are possible convergence points. If there exists a value $\theta_0 \in D_s$ such that

$$C(\theta_0)(zI-A(\theta_0))^{-1}B(\theta_0)=C_0(zI-A_0)^{-1}B_0 \quad \text{a.e.z} \tag{6.3}$$

holds, and the system operates in open loop, then θ_0 will yield a global minimum of $V(\theta)$. If the system operates in closed loop, this last conclusion holds only if in addition $Q_0^v=0$ (Q_0^v defined in (2.2).) □

Proof: The proof is given in Appendix II.

Remark: Since we have not specified how the projection into the compact subset of D_s is to be implemented, there may be a possibility that the estimates are "caught" at a boundary point of this compact subset. With a sensible projection mechanism this can, however, always be avoided. This remark will apply also to Theorems 7.1 and 8.1 below. □

The objective of the modeling procedure is to reduce "the unexplained" and $V(\theta)$ clearly is a measure of how much is left unexplained by the model θ. Therefore, Theorem 6.1 must be considered as a good result for this particular case of the EKF-approach.

VII. A Modified Algorithm

Guided by the results of the previous section, it is natural to interpret the EKF as an attempt to minimize the expected value of the squared residuals associated with model θ. Let us pursue this idea further. Let $\bar{K}(\theta)$, $\bar{S}(\theta)$, $\bar{\hat{x}}(t;\theta)$, and $\bar{\epsilon}(t;\theta)$ be given by (4.1)–(4.3). A suitable criterion to seek to minimize would be

$$V(\theta)=E|\bar{\epsilon}(t,\theta)|^2. \tag{7.1}$$

[Expectation is here over $\{e(t),v(t),u(t)\}$].

A reasonable adjustment scheme to achieve minimization of (7.1) should be related to the gradient of $V(\theta)$. We have, allowing ourselves to interchange differentiation and mathematical expectation,

$$\frac{d}{d\theta}V(\theta)=2E\left[\left(\frac{d}{d\theta}\bar{\epsilon}^T(t;\theta)\right)\bar{\epsilon}(t;\theta)\right].$$

Denote the $n_\theta|n_y$ matrix

$$-\frac{d}{d\theta}\bar{\epsilon}^T(t;\theta)=\bar{\eta}(t;\theta). \tag{7.2}$$

Then the negative gradient of $V(\theta)$ (a column vector) can be written

$$-\frac{d}{d\theta}V(\theta)=2E\bar{\eta}(t;\theta)\bar{\epsilon}(t;\theta). \tag{7.3}$$

One would prefer that, asymptotically, the parameter values θ be corrected in this direction, or in a modified (e.g., "Newton") negative gradient direction. Compare with (4.6)! The asymptotic direction of modification for the EKF is given by $f(\theta)$. The similarity between (4.6) and (7.3) is striking. The relationship between $\bar{\psi}(t;\theta)$ (given by (4.5)) and $\bar{\eta}(t;\theta)$ (given by (7.2)) clearly should be investigated.

Differentiating (4.2) and (4.3) gives

$$\frac{d}{d\theta_i}\bar{\epsilon}(t;\theta)=-\left[\frac{d}{d\theta_i}C(\theta)\right]\bar{\hat{x}}(t;\theta) - C(\theta)\left[\frac{d}{d\theta_i}\bar{\hat{x}}(t;\theta)\right] \tag{7.4}$$

and

$$\begin{aligned}\frac{d}{d\theta_i}\bar{\hat{x}}(t;\theta)=&\left[A(\theta)-\bar{K}(\theta)C(\theta)\right]\frac{d}{d\theta_i}\bar{\hat{x}}(t;\theta)\\&+\left[\frac{d}{d\theta_i}B(\theta)\right]u(t)+\left[\frac{d}{d\theta_i}A(\theta)\right]\bar{\hat{x}}(t;\theta)\\&+\left[\frac{d}{d\theta_i}\bar{K}(\theta)\right]\bar{\epsilon}(t;\theta)-\bar{K}(\theta)\left[\frac{d}{d\theta_i}C(\theta)\right]\bar{\hat{x}}(t;\theta).\end{aligned} \tag{7.5}$$

Let $\bar{W}(t;\theta)$ be the $n_x|n_\theta$ matrix $d/d\theta\bar{\hat{x}}(t;\theta)$. Then, with M and D defined by (3.12) and (3.13), respectively, (7.4) and (7.5) can be rewritten as

$$\begin{aligned}\bar{W}(t+1;\theta)=&\left[A(\theta)-\bar{K}(\theta)C(\theta)\right]\bar{W}(t;\theta)\\&+M\big(\theta,\bar{\hat{x}}(t;\theta),u(t)\big)+\left[\frac{d}{d\theta}\bar{K}(\theta)\right]\bar{\epsilon}(t;\theta)\\&-\bar{K}(\theta)D\big(\theta,\bar{\hat{x}}(t;\theta)\big);\end{aligned} \tag{7.6}$$

$$\bar{\eta}(t;\theta)=\left[C(\theta)\bar{W}(t;\theta)+D\big(\theta,\bar{\hat{x}}(t;\theta)\big)\right]^T. \tag{7.7}$$

Compare with (4.4), (4.5)! We see that $\bar{\eta}(t;\theta)$ is "almost" equal to $\bar{\psi}(t;\theta)$. It is just a term $[d/d\theta\bar{K}(\theta)]\bar{\epsilon}(t;\theta)$ that should be included in $M(\theta,\bar{\hat{x}}(t;\theta),u(t))$ in order to make the EKF-algorithm a (stochastic) descent-algorithm. Furthermore the $\bar{S}^{-1}(\theta)$-term in (4.6) should be deleted.

As a result of this heuristic discussion we might expect improved asymptotic convergence properties of the EKF if (an approximation of) $[d/d\theta\bar{K}(\theta)]_{\theta=\hat{\theta}(t)}\epsilon(t)$, where $\epsilon(t)$ is the current residual, is added to the M_t matrix in the scheme (3.14)–(3.20).

There are several ways to achieve this. One that is quite well-suited for practical computations is treated in the next section. Another would be to calculate the derivative $d/d\theta\bar{K}(\theta)$ from (4.1) and evaluate it at $\hat{\theta}(t)$. Since the gain $K(t)$ is calculated by means of the Riccati equation (3.16), (3.18), it is natural to find an expression for the derivative from the sensitivity equations. Let $\kappa_t^{(i)}$ be defined by

$$\begin{aligned}\kappa_t^{(i)}=&\left[\frac{\partial}{\partial\theta_i}A(\theta)P_1(t)C_t^T+A_t\Pi^{(i)}(t)C_t^T\right.\\&\left.+A_tP_1(t)\frac{\partial}{\partial\theta_i}C^T(\theta)+\frac{\partial}{\partial\theta_i}Q^c(\theta)\right]_{\theta=\hat{\theta}(t)}\\&\cdot S_t^{-1}-K(t)S_t^{-1}\sigma_t^{(i)}S_t^{-1};\end{aligned} \tag{7.8a}$$

$$\sigma_t^{(i)}=\Bigg[\frac{\partial}{\partial\theta_i}C(\theta)P_1(t)C_t^T+C_t\Pi^{(i)}(t)C_t^T$$
$$+C_tP_1(t)\frac{\partial}{\partial\theta_i}C(\theta)+\frac{\partial}{\partial\theta_i}Q^e(\theta)\Bigg]_{\theta=\hat\theta(t)};\quad(7.8\text{b})$$

$$\Pi^{(i)}(t+1)=\Bigg[\frac{\partial}{\partial\theta_i}A(\theta)P_1(t)A_t^T+A_t\Pi^{(i)}(t)A_t^T$$
$$+A_tP_1(t)\frac{\partial}{\partial\theta_i}A^T(\theta)$$
$$+\frac{\partial}{\partial\theta_i}Q^v(\theta)-\kappa_t^{(i)}S_tK^T(t)$$
$$-K(t)\sigma_t^{(i)}K^T(t)-K(t)S_t\kappa_t^{(i)}\Bigg]_{\theta=\hat\theta(t)}.\quad(7.8\text{c})$$

Here, $P_1(t)$, $K(t)$, S_t, A_t, and C_t are defined as in Section III.

If $\hat\theta(t)$ is fixed to a given value $=\theta$, then the algorithm (7.8) will obviously yield a sequence $\kappa_t^{(i)}$ that tends to $\partial/\partial\theta_i\bar K(\theta)$ as t tends to infinity.

We may now state the following result.

Theorem 7.1: Consider the algorithm (3.14)–(3.20′) with M_t replaced by M_t^*, whose ith column is given by

$$M_t^{*(i)}=M_t^{(i)}+\kappa_t^{(i)}(y(t)-C_t\hat x(t))\quad(7.9)$$

where $\kappa_t^{(i)}$ is given by (7.8). Let also S_t^{-1} in (3.17) (but not elsewhere) be replaced by I. Assume that the algorithm is complemented with a projection facility to keep $\hat\theta(t)$ in a compact subset of

$$D_s=\{\theta|(A(\theta),C(\theta))\text{ detectable and }(A(\theta),Q^v)\text{ stabilizable}\}.\quad(7.10)$$

Let the input–output data be generated by the system (2.1), (2.2). Then the estimate $\hat\theta(t)$ converges with probability 1 to a stationary point of the function $V(\theta)$ given by (7.1) (where $\bar\epsilon$ is given by (4.1)–(4.3)), and among isolated stationary points only local minima are possible convergence points.

Proof: The proof is analogous to that of Theorem 6.1. The essential point is that the modification (7.9) has the effect that the associated differential equation (4.8) is replaced by

$$\frac{d}{d\tau}\theta(\tau)=R^{-1}(\tau)f^*(\theta(\tau))\quad(7.11\text{a})$$

$$\frac{d}{d\tau}R(\tau)=G^*(\theta(\tau))+\delta I-R(\tau)\quad(7.11\text{b})$$

where

$$f^*(\theta)=E\bar\eta(t;\theta)\bar\epsilon(t;\theta)\left(=-\frac{1}{2}\cdot\frac{d}{d\tau}V(\theta)\right)\quad(7.12\text{a})$$

$$G^*(\theta)=E\bar\eta(t;\theta)\bar\eta^T(t;\theta).\quad(7.12\text{b})$$

□

As a final remark on the descent character of the algorithm, we may note that

$$\frac{1}{2}\cdot\frac{d^2}{d\theta^2}V(\theta)=E\bar\eta(t;\theta)\bar\eta^T(t;\theta)$$
$$+E\left[\frac{d}{d\theta}\bar\eta(t;\theta)\right]\bar\epsilon(t;\theta).\quad(7.13)$$

Now, at the true value θ_0 (if this exists), $\bar\epsilon(t;\theta_0)$ are the true innovations, which are independent of what has happened before time t. Therefore, the last term in (7.13) is zero at θ_0. We thus find that close to θ_0

$$\frac{1}{2}\cdot\frac{d^2}{d\theta^2}V(\theta)\approx G^*(\theta)$$

and the interpretation of the $R^{-1}(\tau)$-term in (7.11a) is that it changes the gradient step to a "Newton" step.

Note that R^{-1} corresponds to P_3 in the algorithm. Clearly, the same effect could be obtained by replacing $P_3(t)$ by some other approximation of the inverse of the second derivative matrix.

Instead of minimizing $V(\theta)$ given by (7.1) it may sometimes be of interest of minimize

$$V_1(\theta)=E\bar\epsilon^T(t;\theta)\bar S^{-1}(\theta)\bar\epsilon(t;\theta)+\log\det\bar S(\theta)\quad(7.14)$$

where $\bar S(\theta)$ is the assumed covariance matrix for the innovations $\bar\epsilon(t,\theta)$, given by (4.1). If the noise is supposed to be Gaussian, then (7.14) is the expected value of the negative log likelihood function.

Straightforward calculations give

$$\frac{d}{d\theta_i}V_1(\theta)=-2E\left[\bar\eta^{(i)}(t;\theta)\bar S^{-1}(\theta)\bar\epsilon(t;\theta)\right]$$
$$-E\left\{\bar\epsilon^T(t;\theta)\bar S^{-1}(\theta)\left[\frac{d}{d\theta_i}\bar S(\theta)\right]\right.$$
$$\left.\cdot\bar S^{-1}(\theta)\bar\epsilon(t;\theta)\right\}$$
$$+\operatorname{tr}\left[\bar S^{-1}(\theta)\frac{d}{d\theta_i}\bar S(\theta)\right].$$

[Note that the two last terms would cancel if indeed $\bar S(\theta)$ *were* the covariance matrix of $\bar\epsilon(t;\theta)$!] From this expression we see, as before, what modifications of the EKF are necessary in order to minimize $V_1(\theta)$. Hence, we have the following corollary to Theorem 7.1.

Corollary 7.1: Consider the algorithm (3.14)–(3.20′) with M_t replaced by M_t^* as in (7.9) and with the parameter updating algorithm (3.15) replaced by

$$\hat\theta(t+1)=\hat\theta(t)+L(t)\epsilon(t)+\zeta(t)\quad(7.15)$$

where the ith component of $\zeta(t)$ is

$$\zeta^{(i)}(t)=\frac{1}{2}\left[\epsilon^T(t)S_t^{-1}\sigma_t^{(i)}S_t^{-1}\epsilon(t)+\operatorname{tr}S_t^{-1}\sigma_t^{(i)}\right].\quad(7.16)$$

Here $\sigma_i^{(i)}$ is given by (7.8) and

$$\epsilon(t)=y(t)-C_t\hat x(t).\quad(7.17)$$

Assume further the same projection facility as in Theorem

7.1. Then the conclusion of this Theorem holds with the function $V(\theta)$ replaced by $V_1(\theta)$ defined by (7.14).

VIII. An Algorithm Based on Innovations Models

One disadvantage with the modification (7.9) of the EKF is that (7.8) will require an amount of computing that may be forbidding for higher order systems.

In the model there are certain assumptions associated with the noise covariance matrices, whether parameterized or not. It should be noted that the effect of these assumptions is in fact only to provide the Kalman filter gain. It is this gain that has the algorithmic importance and the noise assumptions are only vehicles to arrive at it. Therefore, it should in most cases be a good idea to parameterize the steady-state Kalman gain rather than the covariance matrices. This will normally involve fewer parameters (which means that usually the individual noise covariances are not identifiable—only the Kalman gain and the innovations covariance matrix are). The only cases, when this may be undesirable is when important *a priori* information of the noise structure in the form (3.1) *is* available or when it is important to have a time-varying Kalman gain for the initial part of the recorded data.

What has been said is that it is often a good idea to start with an innovations model

$$\begin{aligned}x(t+1)&=A(\theta)x(t)+B(\theta)u(t)+\overline{K}(\theta)\epsilon(t)\\ y(t)&=C(\theta)x(t)+\epsilon(t)\end{aligned}\tag{8.1}$$

with

$$E\epsilon(t)\epsilon^T(s)=\Lambda\delta_{ts},\qquad x(0)=0\tag{8.2}$$

instead of (3.1), (3.2). [In fact (8.1), (8.2) is just a special case of (3.1), (3.2) with

$$\begin{aligned}&Q^v(\theta)=\overline{K}(\theta)\Lambda\overline{K}^T(\theta);\\ &Q^c(\theta)=\overline{K}(\theta)\Lambda,\\ &Q^e(\theta)=\Lambda\end{aligned}$$

and $\Pi(\theta)=0$.]

If we are going to apply the EKF-idea to the model (8.1), where $\overline{K}(\theta)$ is explicitly parameterized, it is easy to include

$$\frac{\partial}{\partial\theta}\overline{K}(\theta)\epsilon$$

in the cross-coupling term M as discussed in Section VII.

Hence, define

$$\overline{M}(\theta,x,u,\epsilon)=\frac{\partial}{\partial\theta}\big(A(\theta)\hat{x}+B(\theta)u+\overline{K}(\theta)\epsilon\big)\Big|_{\theta=\theta}.\tag{8.3}$$

Introduce for short

$$\overline{M}_t=\overline{M}\big(\hat{\theta}(t),\hat{x}(t),u(t),\epsilon(t)\big)\tag{8.4}$$

and

$$\overline{K}_t=\overline{K}\big(\hat{\theta}(t)\big).$$

From (3.14)–(3.20′) we now obtain the algorithm

$$\hat{x}(t+1)=A_t\hat{x}(t)+B_tu(t)+\overline{K}_t\epsilon(t)\tag{8.5}$$

$$\epsilon(t)=y(t)-C_t\hat{x}(t)\tag{8.6}$$

$$\hat{\theta}(t+1)=\hat{\theta}(t)+L(t)\epsilon(t)\tag{8.7}$$

$$L(t)=\big[P_2^T(t)C_t^T+P_3(t)D_t^T\big]\Lambda^{-1}\tag{8.8}$$

$$P_2(t+1)=A_tP_2(t)+\overline{M}_tP_3(t)-\overline{K}_t\Lambda L^T(t)\tag{8.9}$$

$$P_3(t+1)=P_3(t)-L(t)\Lambda L^T(t)-\delta P_3(t)P_3(t).\tag{8.10}$$

For this algorithm we have the following convergence result.

Theorem 8.1: Consider input–output data generated by the system (2.1), (2.2). Let the model be given by (8.1), (8.2). Suppose that the parameter vector θ is estimated by the algorithm (8.5)–(8.10). Assume further that the algorithm is complemented with a projection facility to keep $\hat{\theta}(t)$ in a compact subset of the set

$$D_s=\{\theta\,|\,\text{the matrix }A(\theta)-\overline{K}(\theta)C(\theta)\text{ is exponentially stable}\}.\tag{8.11}$$

Then the estimate $\hat{\theta}(t)$ converges with probability 1 to a stationary point of the function

$$V_2(\theta)=E\bar{\epsilon}^T(t;\theta)\Lambda^{-1}\bar{\epsilon}(t;\theta)\tag{8.12}$$

where

$$\begin{aligned}\bar{\epsilon}(t;\theta)=&\Big[C_0\big(qI-A_0+\overline{K}_0C_0\big)^{-1}B_0\\ &-C(\theta)\big(qI-A(\theta)+\overline{K}(\theta)C(\theta)\big)^{-1}B(\theta)\Big]u(t)\\ &+\Big[C_0\big(qI-A_0+\overline{K}_0C_0\big)^{-1}\overline{K}_0\\ &-C(\theta)\big(qI-A(\theta)+\overline{K}(\theta)C(\theta)\big)^{-1}K(\theta)\Big]y(t)\\ &+\epsilon_0(t)\end{aligned}\tag{8.13}$$

[q is the forward shift operator].

Furthermore, among isolated stationary points of $V(\theta)$ only local minima are possible convergence points. □

Proof: The proof is entirely analogous to that of Theorem 6.1.

If we would derive a Newton-type stochastic gradient algorithm for the minimization of $V_3(\theta)$ without relying upon the EKF, we would get an algorithm of the following type, cf. [19]–[22]:

$$\begin{aligned}\hat{x}(t+1)&=A_t\hat{x}(t)+B_tu(t)+\overline{K}_t\epsilon(t)\\ \epsilon(t)&=y(t)-C_t\hat{x}(t)\\ \hat{\theta}(t+1)&=\hat{\theta}(t)+R^{-1}(t)\psi(t)\Lambda^{-1}\epsilon(t)\\ \psi(t)&=w^T(t)C_t^T+D_t^T\\ w(t+1)&=\big(A_t-\overline{K}_tC_t\big)w(t)+\overline{M}_t-\overline{K}_tD_t\\ R(t+1)&=R(t)+\psi(t)\Lambda^{-1}\psi^T(t)+\delta I.\end{aligned}\tag{8.14}$$

This could be called a recursive prediction error algorithm. It is easy to see that the differential equation associated with (8.14) is the same as the one associated with (8.5)–(8.10). These two algorithms therefore have the same asymptotic convergence properties. They differ in fact only in the manner $R^{-1}(t)\psi(t)$ is calculated, and this difference is of a transient character. A further discussion of this point can be found in [18], [21], [22]. The question of which of the two algorithms exhibits the best transient convergence behavior must be left to simulation studies.

If also the covariance matrix Λ of the innovations is to be estimated it is natural to parameterize it independently and minimize

$$V_3(\theta,\Lambda)=E\bar{\epsilon}^T(t;\theta)\Lambda^{-1}\bar{\epsilon}(t;\theta)+\log\det\Lambda \qquad (8.15)$$

with respect to θ and Λ. This is, in fact (see, e.g., [23]), the same as minimizing

$$W(\theta)=\det E\bar{\epsilon}(t;\theta)\bar{\epsilon}^T(t;\theta)$$

with respect to θ, giving θ^* and then taking

$$\Lambda^*=E\bar{\epsilon}(t;\theta^*)\bar{\epsilon}^T(t;\theta^*).$$

Therefore, an obvious scheme to minimize (8.15) is to replace Λ in (6.8)–(6.10) by

$$\frac{1}{t}\sum_{1}^{t}\epsilon(k)\epsilon^T(k)\triangleq\hat{\Lambda}(t).$$

Corollary 8.1: Consider the algorithm (8.5)–(8.10) [or (8.14)] with Λ in (8.8)–(8.10) replaced by $\hat{\Lambda}(t)$, where

$$\hat{\Lambda}(t)=\hat{\Lambda}(t-1)+\frac{1}{t}\left[\epsilon(t)\epsilon^T(t)-\hat{\Lambda}(t-1)\right]. \qquad (8.16)$$

Assume further that the algorithm is complemented with a projection facility as described in Theorem 8.1. Then $(\hat{\theta}(t),\hat{\Lambda}(t))$ converges with probability 1 to a stationary point of the function $V_3(\theta,\Lambda)$ given by (8.15). Furthermore, among isolated stationary points of $V_3(\theta,\Lambda)$, only local minima are possible convergence points.

Proof: The proof is analogous to those of Theorems 6.1–8.1. This time $\hat{\Lambda}(t)$ is included in the $\xi(t)$ vector and the associated differential equation becomes

$$\begin{aligned}\dot{\theta}&=R^{-1}f(\theta,\Lambda)\\ \dot{\Lambda}&=H(\theta)-\Lambda\\ \dot{R}&=G(\theta)-R+\delta I\end{aligned} \qquad (8.17)$$

where

$$H(\theta)=E\bar{\epsilon}(t;\theta)\bar{\epsilon}^T(t;\theta)$$

and

$$f(\theta,\Lambda)=-\frac{1}{2}E\frac{d}{d\theta}\left[\bar{\epsilon}^T(t;\theta)\Lambda^{-1}\bar{\epsilon}(t;\theta)\right].$$

With the Lyapunov-function

$$V_3(\theta,\Lambda)=\frac{1}{2}E\bar{\epsilon}^T(t;\theta)\Lambda^{-1}\bar{\epsilon}(t;\theta)+\frac{1}{2}\log\det\Lambda$$

for (8.17) we find

$$\begin{aligned}\dot{V}=&-\frac{1}{2}E\frac{d}{d\theta}\left[\bar{\epsilon}^T(t;\theta)\Lambda^{-1}\bar{\epsilon}(t;\theta)\right]R^{-1}f(\theta,\Lambda)\\ &-\frac{1}{2}E\bar{\epsilon}^T(t;\theta)\Lambda^{-1}(H(\theta)-\Lambda)\Lambda^{-1}\bar{\epsilon}(t;\theta)\\ &+\frac{1}{2}\operatorname{tr}\left[\Lambda^{-1}(H(\theta)-\Lambda)\right]\\ =&-f^T(\theta,\Lambda)R^{-1}f(\theta,\Lambda)\\ &-\frac{1}{2}\operatorname{tr}\Lambda^{-1}(H(\theta)-\Lambda)\Lambda^{-1}(H(\theta)-\Lambda)\\ \leqslant&\ 0\end{aligned}$$

with equality only if $f(\theta,\Lambda)=0$ and $\Lambda=H(\theta)$, which proves the assertion. □

Let us conclude this section by applying the modified algorithm to the example of Section V-C.

Example 8.1: Instead of (5.6) the model will now be given by

$$\begin{aligned}x(t+1)&=ax(t)+k\epsilon(t)\\ y(t)&=x(t)+\epsilon(t)\end{aligned} \qquad (8.18)$$

and $\theta=\binom{a}{k}$.The corresponding input–output model is

$$\begin{aligned}y(t)&=\left(1+\frac{kq^{-1}}{1-aq^{-1}}\right)\epsilon(t)\\ &=\frac{1-(a-k)q^{-1}}{1-aq^{-1}}\epsilon(t)\end{aligned}$$

i.e., a first-order ARMA process. We estimate a and k with the scheme (8.5)–(8.10) in which now

$$\begin{aligned}D_t&=0\\ B_t&=0\\ A_t&=\hat{a}(t)\\ \overline{K}_t&=\hat{k}(t)\\ \overline{M}_t&=\left[\hat{x}(t)\quad \epsilon(t)\right].\end{aligned}$$

Then, according to Theorem 8.1, $\hat{a}(t)$ and $\hat{k}(t)$ will tend to a stationary point of

$$E\bar{\epsilon}^2(t,\theta)$$

where

$$\begin{aligned}\bar{\epsilon}(t,\theta)&=\frac{1-aq^{-1}}{1-(a-k)q^{-1}}y(t)\\ &=\frac{1-aq^{-1}}{1-a_0q^{-1}}\,\frac{1-(a_0-k_0)q^{-1}}{1-(a-k)q^{-1}}\epsilon_0(t).\end{aligned}$$

But it is proved in [24] that this function has only one stationary point, namely $a=a_0$, $k=k_0$. Consequently, the modified algorithm yields estimates that converge with probability 1 to the true values. This holds irrespective of the value λ in (5.5), and we have thus solve the bias problems as well as the convergence problems of Section V-C.

We may also note that the same result holds for a general ARMA-process

$$y(t)+a_1y(t-1)+\cdots+a_ny(t-n)=\epsilon(t)+c_1\epsilon(t-1)+\cdots+c_n\epsilon(t-n) \quad (8.19)$$

for which we use the state model

$$x(t+1)=\begin{bmatrix} -a_1 & 1 & 0 & 0 \\ -a_2 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ -a_n & 0 & 0 & 0 \end{bmatrix}x(t)+\begin{bmatrix} k_1 \\ \vdots \\ k_n \end{bmatrix}\epsilon(t)$$

$$y(t)=(1 \quad 0 \quad \cdots \quad 0)x(t)+\epsilon(t) \quad (8.20)$$

where k_i corresponds to c_i-a_i or

$$x(t+1)=\begin{bmatrix} -a_1 & \cdots & \cdots & -a_n \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots 1 & 0 \end{bmatrix}x(t)+\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}\epsilon(t)$$

$$y(t)=(k_1 \quad \cdots \quad k_n)x(t)+\epsilon(t). \quad (8.21)$$

Application of the modified EKF-algorithm (8.5)–(8.10) to (8.20) or (8.21) consequently yields convergence almost surely to the true parameter values of (8.19). (The result of [24] shows that all stationary points of $V_2(\theta)$, for the model (8.20) or (8.21) give a correct description of (8.19).) This algorithm therefore is very powerful for modeling time series.

IX. Conclusions

The recursive parameter estimation problem for linear systems is inherently a nonlinear filtering problem, and as pointed out in [12] there are in principle no differences between parameter estimation and state estimation. For the nonlinear filtering problem, approximative techniques have to be used, and as remarked in [12], an essential difficulty with all approximation techniques is to establish convergence. Indeed, much of the work associated with nonlinear filtering concerns divergence problems. For the specific method under discussion here, the convergence analysis has illuminated, in an essential way, the possible causes of divergence and bias.

In short, the reason for divergence can be traced to the fact that the effect on the Kalman gain K of a change in θ is not taken care of. This lack of coupling between $K(t)$ and θ in the algorithm may, as demonstrated in Section V, lead to divergence of the estimates, even for simple cases, where only system parameters are unknown. This interpretation is consistent with the observations in practical applications [25], that the behavior of the EKF often is worse when the residuals are large and/or the inputs are small. In such cases the coupling term obviously becomes more important.

Also, for cases where the steady-state Kalman gain does not depend on θ, like for deterministic models, we will have good convergence properties. This was demonstrated in Section VI.

We have also found the remedy for the general case. If we include (an approximation of) a term

$$\left[\frac{\partial}{\partial\theta}\bar{K}(\theta)\right]\epsilon(t) \quad (9.1)$$

in the coupling term M_t as described in Sections VII and VIII, global convergence results are obtained, and the procedure can be interpreted as a minimization of the prediction error associated with model θ; $|\bar{\epsilon}(t;\theta)|^2$. Inclusion of a term like (9.1) is of course particularly simple if the model is given in the innovations form (8.1) with $\bar{K}(\theta)$ explicitly parameterized. However, it could be done also for the general case (3.1), (3.2) as shown in Section VII. In practical applications of the EKF, "manual" adjustments of the noise covariances are often used to make the algorithm work. This is the "tuning of the filter." The inclusion of parameters associated with the Kalman gain can thus be interpreted as automatic tuning of the filter. Also, the analysis has shown that *it is the innovations representation form rather than the original state-space form that should be the basis for the linearization procedure in the EKF.*

The convergence analysis of the EKF using the differential equation approach has thus given two different contributions. Results about the asymptotic behavior of the usual EKF algorithm that appear to be new have been obtained and discussed in Sections V and VI. In addition, a modification of the algorithm has been suggested that yields considerably improved convergence properties. The convergence properties of the modified algorithm are in fact equally good as those of off-line prediction error identification methods, e.g. [26]–[28] (and maximum likelihood methods). Notice that in the convergence results of Theorems 6.1–8.1 it is not necessary to assume that the true system can be described within the model parameterization. Convergence to a local minimum of the variance of the associated residuals is still obtained, and this result coincides with the general results on prediction error methods, [27]. If an exact description is possible, then it can always be checked whether the local minimum in which the algorithms stops is a global one (i.e., give the true description of the system) by performing a residual test. For some particular parameterizations it is *a priori* known that all stationary points indeed are global minima. Among these structures probably ARMA-models of time series are the most important ones.

The discussion has throughout been for the case of estimating parameters that are known to be constant. This is of course inherent in convergence analysis. It is reasonable to assume that the analysis is relevant also for the case of slowly drifting parameters. Then some small process noise would be added to the extended half of the state vector and the gain $L(t)$ would then not tend to zero but to a "small" value. For the genuine nonlinear filtering problem this situation corresponds to the case where some modes are considerably slower than the other ones.

Finally, it is clear that the EKF and its modified versions have obvious relationships with other suggested recursive parameter estimation schemes, [4]. These connections are discussed in [18] and [21]. In fact, the modified algorithm of Section VIII should be regarded as a recursive prediction error (RPE) algorithm, where the way of calculating P_3 is inspired by the EKF. For practical implementations of the algorithms discussed here, it should be noted that in order to achieve acceptable convergence rates some mechanism that gives a somewhat slower decrease of the gain $L(t)$ must be introduced. Some ways to do this are discussed in [4].

Appendix I
Proof of Lemma 4.1

In this Appendix we will have to make frequent references to [13], which is assumed to be available.

In order to verify that (3.14)–(3.20′) satisfies the conditions of Theorems 1 and 2 in [13], we shall first give the algorithm in an asymptotic form, where certain terms tending to zero are sorted out.

From the matrix inversion lemma (3.20′) can be rewritten as

$$\begin{aligned}P_3^{-1}(t+1)=P_3^{-1}(t)+P_3^{-1}(t)L(t)S_t^{-1}\\ \times\left[Q_e+C_t\{P_1(t)-P_2(t)P_3^{-1}(t)P_2(t)\}C_t^T\right]^{-1}\\ \times S_t^{-1}L^T(t)P_3^{-1}(t)+\delta I.\end{aligned}$$

Since the expression within brackets is positive definite, we find that

$$P_3^{-1}(t)\geqslant\delta t\cdot I. \tag{A.1}$$

Therefore, introduce

$$\tilde{P}_3(t)=t\cdot P_3(t). \tag{A.2a}$$

From (3.19′) it follows that $P_2(t)$ is of the same order of magnitude as $P_3(t)$, as long as $[A_t-K(t)C_t]$ is stable. Therefore, introduce also

$$\tilde{P}_2(t)=t\cdot P_2(t) \tag{A.2b}$$

$$\tilde{L}(t)=t\cdot L(t). \tag{A.2c}$$

We can now rewrite (3.15) and (3.20′) as

$$\hat{\theta}(t+1)=\hat{\theta}(t)+\frac{1}{t}\tilde{L}(t)\left[y(t)-C_t\hat{x}(t)\right] \tag{A.3a}$$

$$\begin{aligned}\tilde{P}_3^{-1}(t+1)=\tilde{P}_3^{-1}(t)\\ +\frac{1}{t+1}\Bigg[\tilde{P}_3^{-1}(t)\tilde{L}(t)S_t\Bigg[Q_e+C_t\\ \times\left\{P_1(t)-\frac{1}{t}\tilde{P}_2(t)\tilde{P}_3^{-1}(t)\tilde{P}_2(t)\right\}C_t^T\Bigg]^{-1}\\ \times S_t\tilde{L}(t)\tilde{P}_3^{-1}(t)+\delta I-\tilde{P}_3^{-1}(t)\Bigg].\end{aligned} \tag{A.3b}$$

These two quantities correspond to the estimate vector in the general recursive algorithm of [13]. Let this estimate vector $\xi(t)$ (in [13] called $x(t)$) be

$$\xi(t)=\begin{bmatrix}\hat{\theta}(t)\\ \operatorname{Col}\tilde{P}_3^{-1}(t)\end{bmatrix}$$

and let the observation vector of [13] be

$$\varphi(t)=\begin{bmatrix}\hat{x}(t)\\ \operatorname{Col}\tilde{P}_2(t)\end{bmatrix}$$

where "Col" denotes some way to convert a matrix to a column vector. The updating of the ξ-vector is given by (A.3)

$$\xi(t+1)=\xi(t)+\frac{1}{t}Q(\xi(t),\varphi(t),S_t) \tag{A.4}$$

with a slightly complex, but straightforward definition of $Q(\cdot,\cdot,\cdot)$. Moreover, from (3.19′) and (3.14) we have

$$\begin{aligned}\varphi(t+1)=\left[\begin{array}{c:c}A_t-K(t)C_t & 0\\ \hdashline \zeta\left(\hat{\theta}(t),K(t),\tilde{P}_3(t),t\right) & A_t-K(t)C_t\end{array}\right]\varphi(t)\\ +\beta\left(\hat{\theta}(t),K(t),\tilde{P}_3(t)\right)\begin{bmatrix}u(t)\\ y(t)\end{bmatrix}\end{aligned} \tag{A.5}$$

where the matrix $\zeta(\cdot,\cdot,\cdot,\cdot)$ is obtained from (3.19′). The relationship is linear since M_t and D_t are linear functions of $\hat{x}(t)$ and $u(t)$. Denote the dynamics matrix of (A.5) by

$$\alpha\left(\hat{\theta}(t),K(t),\tilde{P}_3(t),t\right).$$

Its stability properties obviously coincide with those of $A(\theta)-\bar{K}(\theta)C(\theta)$ for a constant θ. This matrix is stable if $\theta\in D_s$.

To verify the B-conditions of [13] we note that the quadratic structure of $Q(\cdot,\cdot,\cdot)$ and the assumed differentiability of $C(\theta)$, $D(\theta,\hat{x})$, etc. assure that B.3 holds. As the Lipschitz constant we may take

$$K_1(\xi,\varphi,\zeta,v)=C^M(|\theta|+\zeta)(1+|\varphi|+v)^2(1+|\tilde{P}_3|+\zeta)$$

where

$$C^M=\sup_\theta\left\{\left|\frac{\partial}{\partial\theta}A(\theta)\right|+\left|\frac{\partial}{\partial\theta}B(\theta)\right|+\left|\frac{\partial}{\partial\theta}C(\theta)\right|\right\}.$$

With K_1 as above, clearly also B.4 holds. Since all moments of $v(t)$, $e(t)$, and $u(t)$ are assumed to exist, so will also all moments of Q, K_1, and K_2, which verifies condition B.7. Conditions B.8–B.11 are all satisfied for $\gamma(t)=1/t$.

The only problem in the application of the theorem is associated with condition B.5. As discussed in [13, Appendix V], (A.2) is more general than (1) and (2) of [13], since $\alpha(\cdot,\cdot,\cdot,\cdot)$ is a function of $K(t)$, which is not a direct function of θ. It will, however, be close to $\bar{K}(\hat{\theta}(t))$ for large t. In order to apply the results of [13], we therefore have to consider

$$|K(t)-\bar{K}(\theta)|$$

where $\bar{K}(\theta)$ is defined by (4.1).

Now consider the situation when $\hat{\theta}(t)$ belongs to a small enough neighborhood of $\bar{\theta}\in D_s$ and $K(t)$ belongs to a small enough neighborhood of $\bar{K}(\bar{\theta})$ for $n\leqslant t\leqslant j-1$ and $|\bar{x}(n)|<C$, $|\tilde{P}_2(n)|<C$ (which is the situation from which the proofs in [13] are built up). Then $[A_t-K(t)C_t]$ is exponentially stable for $n\leqslant t\leqslant j$ and we find that

$$|\varphi(j)|\leqslant C\tilde{\lambda}^{j-n}+v(j,\tilde{\lambda},c)$$

as in (I.14) of [13] with $v(\cdot,\cdot,\cdot)$ defined in [13]. From (3.16) and (4.1) we find that

$$\begin{aligned}|K(j)-K(\bar{\theta})|\leqslant{}&|A_jP_1(j)C_j-A(\bar{\theta})\bar{P}_1(\bar{\theta})C(\bar{\theta})|\\&+C\cdot\frac{1}{j}|\varphi(j)|.\qquad\text{(A.6)}\end{aligned}$$

Moreover,

$$\begin{aligned}|P_1(j)-\bar{P}_1(\bar{\theta})|\leqslant{}&C\cdot\max_{n\leqslant k\leqslant j}|\theta(k)-\bar{\theta}|\\&+C\cdot\frac{1}{n}\max_{n\leqslant k\leqslant j}\left(1+|\varphi(k)|^3\right)\\&+C\lambda^{j-n}|P_1(n)-\bar{P}_1(\bar{\theta})|\qquad\text{(A.7)}\end{aligned}$$

which follows from the fact that the Riccati equation (3.18) is stable with respect to disturbances in its parameters, and exponentially stable with respect to initial conditions, [1].

The effect of (A.6) with (A.7) in the proofs of [13] will be that in the expressions following (I.18), p. 566, a term should be added. In the final expression (I.27) this term takes the form

$$\frac{1}{n}\sum_{n}^{j}\gamma(k)v(k,\lambda,c)\qquad\text{(A.8)}$$

and Step 3c of [13] is still valid.

A similar problem arises from the fact that in (A.4) S_t enters, which is not a direct function of $\xi(t)$. However, for large t, $S_t-\bar{S}(\hat{\theta}(t))$ will be small and bounded by the same quantities as in (A.6), (A.7).

It now remains only to show that the associated differential equation in fact is (4.8). With ξ fixed to the values θ and $\tilde{P}_3^{-1}$, we find that the upper part of the $\bar{\varphi}$-vector will be $\bar{\hat{x}}(t;\theta)$, given by (4.2), (4.3). Similarly, comparing (3.19′) with (4.4) we find that the lower part of $\bar{\varphi}(t;\xi)$ is

$$\tilde{P}_2(t;\theta,\tilde{P}_3)=\bar{w}(t;\theta)\tilde{P}_3$$

where $\bar{w}(t;\theta)$ is given by (4.4), and that consequently (cf. (3.17))

$$\begin{aligned}\bar{\tilde{L}}(t;\theta,\tilde{P}_3)={}&\tilde{P}_3\Big[\bar{w}^T(t;\theta)C^T(\theta)\\&+D^T\big(\theta,\bar{\hat{x}}(t;\theta)\big)\Big]\bar{S}^{-1}(\theta)\\={}&\tilde{P}_3\bar{\psi}(t;\theta)\bar{S}^{-1}(\theta).\qquad\text{(A.9)}\end{aligned}$$

Hence, since $1/t\tilde{L}(t)\epsilon(t)$ updates $\hat{\theta}(t)$ (see A.3a) the right-hand side of the differential equation (d.e.) associated with θ will be

$$E\bar{\tilde{L}}(t;\theta,\tilde{P}_3)\bar{\epsilon}(t;\theta)=\tilde{P}_3f(\theta)$$

where $f(\theta)$ is given by (4.6). Similarily, from (A.3b), what is updating $\tilde{P}_3^{-1}$ is asymptotically

$$\tilde{P}_3^{-1}\tilde{L}S_t[Q_e+C_tP_1C_t^T]^{-1}S_t\tilde{L}\tilde{P}_3^{-1}+\delta I-\tilde{P}_3^{-1}.$$

Now

$$\begin{aligned}&E\tilde{P}_3^{-1}\bar{\tilde{L}}(t;\theta,\tilde{P}_3)\bar{S}(\theta)[Q_e+C(\theta)\bar{P}_1(\theta)C^T(\theta)]^{-1}\\&\times\bar{S}(\theta)\bar{\tilde{L}}(t;\theta,\tilde{P}_3)=E\bar{\psi}(t;\theta)\bar{S}^{-1}(\theta)\bar{\psi}^T(t;\theta)\\&=G(\theta)\end{aligned}$$

according to (A.9), (4.1b), (4.7). Hence, the right-hand side of the d.e. associated with $\tilde{P}_3^{-1}$ is $G(\theta)+\delta I-\tilde{P}_3^{-1}$. Introducing the notation R for $\tilde{P}_3^{-1}$ now gives the d.e. (4.8).

Appendix II
Proof of Theorem 6.1

Let us consider the differential equation (4.1)–(4.8) in case $Q^v=0$. Then (4.1) gives that $\bar{P}_1(\theta)=0$, $\bar{K}(\theta)=0$, and $\bar{S}(\theta)=Q^e$ for all $\theta\in D_s$. Since $\bar{K}(\theta)=0$ in (4.2) we find that, in fact,

$$\bar{w}(t;\theta)=\frac{\partial}{\partial\theta}\bar{\hat{x}}(t;\theta)$$

and hence

$$\bar{\psi}(t;\theta)=-\frac{\partial}{\partial\theta}\bar{\epsilon}(t;\theta).$$

The last expression implies, via (4.6), that

$$f(\theta)=-\frac{1}{2}\cdot\frac{d}{d\theta}V(\theta)$$

(interchanging differentiation and expectation) where $V(\theta)$ is defined in the theorem. Along a solution $\theta(\tau)$ to (4.8) we have

$$\begin{aligned}\frac{d}{d\tau}V(\theta(\tau))&=\left[\frac{d}{d\theta}V(\theta)\Big|_{\theta=\theta(\tau)}\right]^T\frac{d}{d\tau}\theta(\tau)\\&=-\frac{1}{2}\cdot f^T(\theta(\tau))R^{-1}(\tau)f(\theta(\tau)).\end{aligned}$$

Since $R^{-1}(\tau)$ is a positive definite matrix, the function V

is a Lyapunov function for (4.8), i.e., it is decreasing outside the set

$$D_c = \{\theta | f(\theta) = 0\}$$

which coincides with the set of stationary points of (5.1). As long as $\hat{\theta}(t)$ remains in D_s, we have consequently shown that the estimate $\hat{\theta}(t)$ converges with probability one to the set D_c. In fact, using result 2) of Lemma 4.1, it follows that among the isolated points in D_c only the local minima of $V(\theta)$ are possible convergence points of $\{\theta(t)\}$.

We now turn to the question of the effect of projecting the estimates $\hat{\theta}(t)$ into the set $D_s = \{\theta | A(\theta) \text{ stable}\}$. Since the function $V(\theta)$ tends to infinity as θ approaches the boundary of D_s, the trajectories of (5.1) which point "downhill" cannot cross the boundary. Therefore, Lemma 4.1:3) implies convergence of the estimates to D_c. Finally, we shall somewhat comment upon the set D_c. Suppose that there exists a $\theta_0 \in D_s$ such that (6.3) holds that is, θ_0 gives the true input–output transfer function for the model. From the representation (2.10) of the true system we find that

$$\begin{aligned} y(t) &= C_0(qI - A_0)^{-1}B_0u(t) \\ &\quad + C_0(qI - A_0)^{-1}\bar{K}_0\epsilon_0(t) + \epsilon_0(t) \\ &= C(\theta_0)\bar{\hat{x}}(t;\theta_0) \\ &\quad + \left(C_0(qI - A_0)^{-1}\bar{K}_0 + I\right)\epsilon_0(t) \end{aligned}$$

where the last equality follows from (6.1). We then have

$$\begin{aligned} \bar{\epsilon}(t,\theta) = C(\theta_0)\bar{\hat{x}}(t;\theta_0) - C(\theta)\bar{\hat{x}}(t;\theta) \\ + \left(C_0(qI - A_0)^{-1}\bar{K}_0 + I\right)\epsilon_0(t). \quad \text{(B.1)} \end{aligned}$$

The estimate $\bar{\hat{x}}(t;\theta)$ is formed entirely from $u(s)$, $s<t$, and if the system operates *in open loop (u(t)* independent of $v_0(s)$, $e_0(s)$, all s), then the first two terms in (B.1) are independent of the last one. This means that θ_0 gives the *global minimum* of the function $E\bar{\epsilon}^T(t,\theta)(Q^e)\bar{\epsilon}^{1-}(t,\theta)$, and in particular, that $f(\theta_0) = 0$.

If the system operates in *closed loop*, where the input is partly determined from output feedback ($u(t)$ independent of $v_0(s)$, $e_0(s)$, $s>t$) then we have to require that $\bar{K}_0 = 0$ in order to draw the same conclusions.

We have thus concluded that $\theta_0 \in D_c$. Whether this set contains more points is a question of the parameterization, and has to be studied separately.

References

[1] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. New York: Academic, 1970.

[2] G. N. Saridis, "Comparison of six on-line identification algorithms," *Automatica*, vol. 10, no. 1, pp. 69–80, 1974.

[3] R. Isermann, U. Baur, W. Bamberger, P. Kneppo and H. Siebert, "Comparison of six on-line identification and parameter estimation algorithms," *Automatica*, vol. 10, no. 1, pp. 81–104, 1974.

[4] T. Söderström, L. Ljung, and I. Gustavsson, "A theoretical analysis of recursive identification algorithms," *Automatica*, vol. 14, pp. 231–244, 1978.

[5] R. E. Kopp and R. J. Orford, "Linear regression applied to system identification for adaptive control systems," *AIAA J.* vol. 1, no. 10, pp. 2300–2306, 1963.

[6] H. Cox: "On the estimation of state variables and parameters for noisy dynamic systems," *IEEE Trans. Automat. Contr.*, vol. AC-9, pp. 5–12, Feb. 1964.

[7] A. P. Sage and C. D. Wakefield, "Maximum likelihood identification of time varying and random system parameters," *Int. J. Contr.*, vol. 16, no. 1, pp. 81–100, 1972.

[8] V. P. Leung and L. Padmanabhan, "Improved estimation algorithms using smoothing and relinearization," *Chem. Eng. J.*, vol. 5, pp. 197–208, 1973.

[9] L. W. Nelson and E. Stear, "The simultaneous on-line estimation of parameters and states in linear systems," *IEEE Trans. Automat. Contr.*, vol. AC-21, pp. 94–98, Feb. 1976.

[10] J. B. Farison, R. E. Graham, and R. C. Shelton, "Identification and control of linear discrete systems," *IEEE Trans. Automat. Contr.*, vol AC-12, pp. 438–442, Aug. 1967.

[11] D. G. Orlhac, M. Athans, J. Speyer, and P. K. Houpt, "Dynamic stochastic control of freeway corridor systems, Volume IV—Estimation of traffic variables via extended Kalman filter methods," MIT, Cambridge, MA, Rep. ESL-R-611, Sept. 1975.

[12] K. J. Åström and P Eykhoff, "System Identification-A survey," *Automatica*, vol. 7, pp. 123–162, 1971.

[13] L. Ljung, "Analysis of recursive, stochastic algorithms," *IEEE Trans. Automat. Contr.*, vol. AC-22, pp. 551–575, Aug. 1977.

[14] A. H. Jazwinski, "Adaptive filtering," *Automatica*, vol. 5, pp. 475–485, 1969.

[15] R. K. Mehra, "On the identification of variances and adaptive Kalman filtering," *IEEE Trans. Automat. Contr.*, vol. AC-15, pp. 175–184, Apr. 1970.

[16] R. F. Ohap and A. R. Stubberud, "Adaptive minimum variance estimation in discrete-time linear systems," in *Control and Dynamic Systems, Advances in Theory and Applications*, vol. 12. New York: Academic, 1976, pp. 583–624.

[17] P. R. Bélanger, "Estimation of noise covariance matrices for a linear time-varying stochastic process," *Automatica*, vol. 10, pp. 267–277, 1974.

[18] L. Ljung, "The extended Kalman filter as a parameter estimator for linear systems," Dep. Elec. Eng. Linköping Univ., Linköping, Sweden, LiTH-ISY-I-0154, May 1977.

[19] T. Söderström, "An on-line algorithm for approximate maximum likelihood identification of linear dynamic systems," Div. Automat. Contr., Lund Inst. of Techn., Lund, Sweden, Rep. 7308, 1973.

[20] L. Ljung, "Some basic ideas in recursive identification," *Journees d'Automatique*, IRISA, Rennes, France, Sept 1977; IRISA Publication Interne, No. 92, pp. 36–49.

[21] ——, "On recursive prediction error identification algorithms," Dep. Elec. Eng., Linköping Univ., Sweden, Rep. LiTH-ISY-I-0226, Aug. 1978.

[22] J. B. Moore and H. Weiss, "Recursive prediction error methods for adaptive estimation," *Preprint*, Dep. Elec. Eng., Univ. Newcastle, N.S.W., Australia, Feb. 1978.

[23] H. Akaike, "Maximum likelihood identification of Gaussian autoregressive moving average models," *Biometrika*, vol. 60, pp. 255–265, 1973.

[24] K. J. Åström and T. Söderström, "Uniqueness of the maximum likelihood estimates of the parameters of an ARMA model," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 768–774, Dec. 1974.

[25] G. Stein, private communication.

[26] L. Ljung, "Prediction error identification methods," Dep. Elec. Eng., Linköping Univ., Linköping, Sweden, Rep. LiTH-ISY-o139, March 1977.

[27] ——, "Convergence analysis of parametric identification methods," *IEEE Trans. Automat. Contr.*, Vol. AC-23, No. 5, 1978.

[28] P. E. Caines, "Prediction error identification methods for stationary stochastic processes," *IEEE Trans. Automat. Contr.*, vol. AC-21, pp. 500–506, Aug. 1976.

Lennart Ljung (S'74–M'75) was born in Malmö, Sweden, on September 13, 1946. He received the B.A., M.S., and Ph.D. degrees in 1967, 1970, and 1974, respectively, all from Lund University, Sweden.

From 1970 to 1976 he held various teaching and research positions at the Department of Automatic Control, Lund Institute of Technology, Sweden. In 1972 he spent six months with the Laboratory for Adaptive Systems at the Institute for Control Problems (IPU) in Moscow, USSR. In 1974–1975 he was a Research Associate at the Information Systems Laboratory at Stanford University. Since 1976 he has been a Professor of Automatic Control in the Department of Electrical Engineering, Linköping University, Linköping, Sweden. His scientific interests include aspects on identification, estimation, and adaptive control.