The Annals of Mathematical Statistics
1969, Vol. 40, No. 6, 2064–2075

CONSISTENT ESTIMATES OF THE PARAMETERS OF A LINEAR SYSTEM[1]

BY WILLIAM N. ANDERSON, JR., GEORGE B. KLEINDORFER, PAUL R. KLEINDORFER, AND MICHAEL B. WOODROOFE

Carnegie-Mellon University

1. Introduction. We will be concerned with the following dynamic linear system which finds application in both economics and engineering, for example Aoki [3] and Griliches [6] have used this model.

$$x_{k+1} = Ax_k + v_k, \qquad k \geqq 0 \tag{1.1}$$

$$y_k = x_k + w_k, \qquad k \geqq 1. \tag{1.2}$$

In (1.1), the state equation, x_k is a p-dimensional column vector which represents the state of some system at time k; A is a $p \times p$ transition matrix; and v_k represents a random disturbance, or noise.

In (1.2), the observation equation, y_k represents an observation made on the system at time k, and w_k represents noise. We will assume that $v_0, v_1, \cdots$ and $w_1, w_2, \cdots$ are independent sequences of zero mean, independent and identically distributed random vectors with covariance matrices V and W respectively and that x_0 is independent of the v_i's and w_j's and has finite covariance matrix. We remark, in passing, that the superficially more general model in which (1.2) is replaced by

$$y_k = Mx_k + v_k, \qquad k \geqq 1,$$

where M is nonsingular, may be reduced to (1.2) by an appropriate change of bases.

When A, V, W, and the distribution of x_0 are known, linear least squares prediction and filtering may be done with the Kalman Filter [10], which provides a method for computing the projections, $x_{t \mid k}$ and $y_{t \mid k}$, of x_t and y_t on the Hilbert subspace spanned by $y_1, \cdots, y_k$. Specifically,

$$\begin{aligned} x_{k \mid k} &= (I - \Delta_k)Ax_{k-1 \mid k-1} + \Delta_k y_k, && k \geqq 1, \\ x_{t \mid k} &= A^{t-k}x_{k \mid k}, && \\ y_{t \mid k} &= x_{t \mid k}, && t > k, \end{aligned} \tag{1.3}$$

where I denotes the $p \times p$ identity matrix and $x_{0 \mid 0} = E[x_0]$. The matrix Δ_k appearing in (1.3) is determined by

Received 30 October 1968; revised 18 June 1969.

[1] This report was prepared as part of the activities of the Management Sciences Research Group, Carnegie-Mellon University, under Contract NONR 760(24) NR 047–048 with the U. S. Office of Naval Research and by the National Science Foundation through its grant GP 8910. Reproduction in whole or in part is permitted for any purpose of the U. S. Government.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Mathematical Statistics.* www.jstor.org

(1.4a) $$S_k = AP_{k-1}A' + V$$

(1.4b) $$\Delta_k = S_k(S_k + W)^+$$

(1.4c) $$P_k = (I - \Delta_k)S_k, \qquad k \geqq 1,$$

where $^+$ denotes pseudo-inverse, $'$ denotes transpose, and P_0 is the convariance matrix of x_0.

In practice, however, A, V, and W will often be unknown, so that two problems arise in connection with the Kalman Filter. First, the parameters A, V, and W must be estimated from the y_k's; and second, the effects of replacing A, V, and W by estimates in (1.4) should be considered. In this paper we will present estimates of A, V, and W, and show that they are strongly consistent when the system (1.1) is stable, that is when $\rho(A)$, the spectral radius of A, is less than one. We will then determine the asymptotic behavior as $k \to \infty$ of (1.4) and show that it is unchanged if A, V, and W are replaced by strongly consistent estimates. Our results are stated precisely in Section 2 and proved in Sections 3, 4 and 5. Theorem 2.3 and Section 4 are independent of the remainder of the paper.

Other approaches to the problem of parameter estimation in (1.1) and (1.2) and/or determining the effect of replacing A, V, and W by estimates in the Kalman Filter may be found in [3], [4], [7], [9], and [13]. These authors, however, have not been primarily concerned with analytical results; in fact, only [2] and [5] even consider the consistency of their estimates. Somewhat more theoretical work has been done on parameter estimation in linear-stochastic difference equations with independent inputs, of which (1.1) and (1.2) are a special case if $W = 0$ ([16], [17], and [19]). The presence of a non-zero W in (1.2), however, introduces major complications in the filtering and prediction problems (taking $W = 0$ in (1.4) yields $P_k = 0$, $S_k = V$, and $\Delta_k = VV^+$, $k \geqq 1$) as well as some complications in the parameter estimation problem. The main results of Section 4 on the asymptotic behavior of (1.4) have been proved by Kalman and Bucy [12] for the continuous case; i.e. when (1.1) is a differential instead of a difference equation. Theorem 2.3 has been proven via Lyapunov theory by Kalman ([11], page 371) under somewhat stronger conditions. However, the proof is not given explicitly for the discrete case, so we have included a proof here by other means.

2. Statement of the theorems. In order to state our results precisely, we will need the following notation. We will denote by R^p and $\mathfrak{R}^p$ respectively the real linear spaces of p-dimensional column vectors with real components and $p \times p$ matrices with real entries. The topologies in R^p and $\mathfrak{R}^p$ will be determined by the Euclidian norms.

$$|x| = (x'x)^{\frac{1}{2}}, \qquad x \in R^p$$

$$\|G\| = [\text{tr } (GG')]^{\frac{1}{2}}, \qquad G \in \mathfrak{R}^p.$$

If $G \in \mathfrak{R}^p$ is symmetric, then $G > 0$ and $G \geqq 0$ mean that G is positive definite (pd) and positive semi-definite (psd) respectively, and if $F, G \in \mathfrak{R}^p$ are symmetric, then $F \geqq G$ iff $F - G \geqq 0$. Finally, we will need the notion of parallel addition which is defined for psd matrices F, G by

$$F:G = F(F+G)^+G.$$

$F:G$ is called the parallel sum of F and G and is studied in detail by Anderson and Duffin [1], [2].

We will estimate the parameter A of (1.1) by

$$\hat{A}_n = \left(\sum_{k=3}^n y_k y_{k-2}'\right)\left(\sum_{k=3}^n y_{k-1} y_{k-2}'\right)^+, \qquad n \geqq 3. \tag{2.1}$$

The estimate $\hat{A}_n$ is suggested by the fact that $E\{y_k y_{k-2}'\} = AE\{y_{k-1}y_{k-2}'\}$, $k \geqq 3$, and enjoys the following consistency property.

THEOREM 2.1. *If $\rho(A) < 1$, and if A and V are nonsingular, then $\hat{A}_n$ is a strongly consistent estimate of A, that is, $\hat{A}_n \to A$ with probability one as $n \to \infty$.*

Theorem 2.1 will be proved in Section 3. Granting its validity for the moment, we may then estimate V and W as follows. Define

$$\begin{aligned} B_1 &= E\{(y_k - Ay_{k-1})(y_k - Ay_{k-1})'\} \\ &= V + W + AWA', \\ B_2 &= E\{(y_k - A^2y_{k-2})(y_k - A^2y_{k-2})'\} \\ &= V + W + AVA' + A^2WA^{2'}; \end{aligned}$$

then, if A is nonsingular, B_1, B_2, and A uniquely determine V and W by

$$\begin{aligned} W &= \tfrac{1}{2}\{B_1 + A^{-1}(B_1 - B_2)A^{-1'}\} \\ V &= B_1 - W - AWA'. \end{aligned}$$

Therefore, strongly consistent estimates of A, B_1, and B_2 determine strongly consistent estimates of V and W.

THEOREM 2.2. *If $\hat{A}_n$ is any strongly consistent estimate of A and if $\rho(A) < 1$, then*

$$B_{n,i} = 1/n \sum_{k=3}^n (y_k - \hat{A}_k^i y_{k-i})(y_k - \hat{A}_k^i y_{k-i})', \qquad n \geqq 3, \tag{2.2}$$

is a strongly consistent estimate of B_i, $i = 1, 2$. In particular if A and V are nonsingular, and if $\hat{A}_k$ is given by (2.1) then $B_{n,i}$ is a strongly consistent estimate of B_i.

REMARK. We have used $\hat{A}_k$ rather than $\hat{A}_n$ in (2.2) in order to make the computation of $B_{n,i}$ Markovian. It will be clear from the proof of Theorem 2.2 however, that $B_{n,i}$ would still be strongly consistent if $\hat{A}_k$ were replaced by $\hat{A}_n$ in (2.2).

Given any strongly consistent estimates $\hat{A}_n$, $\hat{V}_n$, and $\hat{W}_n$ of A, V, and W respectively it is natural to approximate the Kalman Filter by

$$\begin{aligned} \hat{S}_k &= \hat{A}_k\hat{P}_{k-1}\hat{A}_k' + \hat{V}_k \\ \hat{\Delta}_k &= \hat{S}_k(\hat{S}_k + \hat{W}_k)^+ \\ \hat{P}_k &= (I - \hat{\Delta}_k)\hat{S}_k \end{aligned} \tag{2.3}$$

$$\hat{x}_{k \mid k} = (1 - \hat{\Delta}_k)\hat{A}_k\hat{x}_{k-1 \mid k-1} + \hat{\Delta}_k y_k, \qquad k \geqq 1, \tag{2.4}$$

where I denotes the $p \times p$ identity matrix and $\hat{P}_0$ may be any psd matrix. A natural object of interest is then the asymptotic behavior of $\Delta_k - \hat{\Delta}_k$ and $x_{k \mid k} - \hat{x}_{k \mid k}$ as $k \to \infty$. Our analysis of this behavior requires knowledge of the asymptotic behavior of S_k as $k \to \infty$, which, of course, is of interest in its own right.

THEOREM 2.3. *Let V be >0 and let W be $\geqq 0$; define ϕ on the set $\mathcal{S}$ of* pd *matrices by*

$$\phi(S) = A(S:W)A' + V, \qquad S \varepsilon \mathcal{S}. \tag{2.5}$$

Then $S_k = \phi(S_{k-1})$, $k \geqq 2$. Moreover, ϕ has a unique positive definite fixed point S_0, and $\phi^n(S) \to S_0$ uniformly on $\mathcal{S}$ as $n \to \infty$, where ϕ^n denotes the nth *iterate of ϕ.*

COROLLARY 2.1. *Let $V > 0$, then $S_k \to S_0$, $\Delta_k \to \Delta_0 = S_0(S_0 + W)^{-1}$, and $P_k \to P_0 = S_0 - S_0(S_0 + W)^{-1}S_0$ as $k \to \infty$.*

COROLLARY 2.2. *For $A \varepsilon \mathcal{R}^p$, $W \geqq 0$, and $V > 0$, define $S_0(A, V, W)$ to be the unique positive definite fixed point of the function ϕ defined by* (2.5); *then $S_0(A, V, W)$ depends continuously on (A, V, W).*

REMARK. Theorem 2.3 and its corollaries have applications in the study of asymptotic properties of certain classes of optimal control problems via the duality theorem of Kalman [10].

The proofs of Theorem 2.3 and Corollary 2.2 will be presented in Section 4 together with an example illustrating some difficulties which may arise if V is not pd. Corollary 2.1 is an obvious consequence of Theorem 2.3. We now consider the asymptotic behavior of $\mathcal{S}_k$ and $\hat{x}_{k \mid k}$. Theorems 2.4 and 2.5 (below) will be proved in Section 5; their corollaries are obvious.

THEOREM 2.4. *If $V > 0$, and if $\hat{A}_n$, $\hat{V}_n$, and $\hat{W}_n$ are strongly consistent estimates of A, V, and W for which $\hat{V}_n > 0$ and $\hat{W}_n \geqq 0$ for all $n \geqq 1$, then $\hat{S}_k \to S_0$ with probability one as $k \to \infty$.*

COROLLARY 2.3. *If the hypotheses of Theorem 2.4 are satisfied, then $\hat{\Delta}_k \to \Delta_0$ with probability one as $k \to \infty$.*

THEOREM 2.5. *If $V > 0$, if $\rho(A) < 1$, and if $\hat{A}_n$, $\hat{V}_n$, $\hat{W}_n$ are strongly consistent estimates of A, V, and W for which $\hat{V}_n > 0$ and $\hat{W}_n \geqq 0$ for all $n \geqq 1$, then*

$$\lim_{n\to\infty} n^{-1} \textstyle\sum_{k=1}^{n} |x_{k \mid k} - \hat{x}_{k \mid k}| = 0 \quad \textit{with probability one.}$$

COROLLARY 2.4. *If the hypotheses of Theorem 2.5 are satisfied, then*

$$\lim_{n\to\infty} n^{-1} \textstyle\sum_{k=1}^{n} |x_{k+r \mid k} - \hat{A}_k^{\,r}\hat{x}_{k \mid k}| = 0 \quad \textit{with probability one.}$$

for any $r \geqq 1$.

3. Consistency of the estimates. In this section we will establish Theorems 2.1 and 2.2; accordingly we assume throughout that $\rho(A) < 1$. Define

$$z_k = w_k + v_{k-1} - Aw_{k-1} = y_k - Ay_{k-1}, \qquad k \geqq 2$$

$$R_n = (\textstyle\sum_{k=3}^{n} z_k y_{k-2}')(\sum_{k=3}^{n} y_{k-1}y_{k-2}')^{+} \qquad n \geqq 3.$$

To prove Theorem 2.1, we will show that, with probability one,

$$n^{-1} \sum_{k=3}^{n} z_k y_{k-2}' \to 0 \tag{3.1}$$

$$n^{-1} \sum_{k=3}^{n} y_{k-1} y_{k-2}' \to \psi \tag{3.2}$$

where ψ is nonsingular. It then follows that for large n the left side of (3.2) is nonsingular, and thus for large n, by a simple computation, that $\hat{A}_n = A + R_n$. Thus, (3.1) and (3.2) suffice to prove Theorem 2.1.

To establish (3.1) we need first a bound on the covariance matrix of y_k. We have from (1.1) and (1.2)

$$y_k = w_k + \sum_{j=0}^{k-1} A^j v_{k-j-1} + A^k x_0 \tag{3.3}$$

from which it follows that

$$E(y_k) = A^k E(x_0) \tag{3.4a}$$

$$\begin{aligned} \Sigma_k &= \text{Cov}\ (y_k) \\ &= W + \sum_{j=0}^{k-1} A^j V A'^j + A^k \text{ Cov } (x_0) A'^k \qquad (3.4b) \\ &\to W + \sum_{j=0}^{\infty} A^j V A'^j = \Sigma \quad \text{say,} \end{aligned}$$

as $k \to \infty$. Here we have used the fact that $\lim_{n\to\infty} \|A^n\| n^{-1} = \rho(A) < 1$ ([15], page 75). Define

$$S_{n,1} = \sum_{k=3}^{n} k^{-1} w_k y_{k-2}'$$

$$S_{n,2} = \sum_{k=3}^{n} k^{-1} (v_{k-1} - A w_{k-1}) y_{k-2}', \qquad n \geqq 3,$$

and for $n \geqq 3$ let $\mathfrak{F}_n$ be the smallest σ algebra with respect to which $x_0, v_0, \cdots, v_n, w_1, \cdots, w_n$ are measurable. If $a, b \in R^p$ it is easily seen that $\{a' S_{n,1} b;\ \mathfrak{F}_n; n \geqq 3\}$ and $\{a' S_{n,2} b;\ \mathfrak{F}_{n-1}; n \geqq 3\}$ are martingales, $i = 1, 2$; for example

$$\begin{aligned} E(a' S_{n+1,1} b \mid \mathfrak{F}_n) - a' S_{n,1} b &= (n+1)^{-1} a' E(w_{n+1} y_{n-1}' \mid \mathfrak{F}_n) b \\ &= (n+1)^{-1} a' E(w_{n+1}) y_{n-1}' b = 0. \end{aligned}$$

Moreover, by the mutual independence of $w_1, w_2, \cdots$, and the independence of $w_k, w_{k+1}, \cdots$, from $y_1, \cdots, y_{k-1}$, we find

$$\begin{aligned} E\{(a' S_{n,1} b)^2\} &= \sum_{k=3}^{n} k^{-2} E\{(a' w_k y_{k-2}' b)^2\} \\ &= \sum_{k=3}^{n} k^{-2} (a' W a) b' E\{y_{k-2} y_{k-2}'\} b \end{aligned}$$

and (similarly)

$$E\{(a' S_{n,2} b)^2\} = \sum_{k=3}^{n} k^{-2} [a'(V + AWA')a] b' E\{y_{k-2} y_{k-2}'\} b$$

are bounded for $n \geqq 3$ by (3.4). The martingale convergence theorem ([5], page 319) therefore asserts that $\lim_{n\to\infty} a' S_{n,i} b$ exists and is finite with probability one, $i = 1, 2$. It now follows from the Kronecker Lemma ([14], page 238) (which asserts that if $\sum a_n$ converges and $b_n \to \infty$, then $b_n^{-1} \sum_{k=1}^{n} b_k a_k \to 0$) that

$$a'(n^{-1} \textstyle\sum_{k=3}^{n} w_k y_{k-2}')b \to 0$$
$$a'(n^{-1} \textstyle\sum_{k=3}^{n} (v_{k-1} - A w_{k-1}) y_{k-2}')b \to 0$$

with probability one. Equation (3.1) now follows by the arbitrariness of a and b.

To establish (3.2) we will take advantage of the fact that y_k in (3.3) is almost a moving average of the v_i's and w_j's. Let $v_{-1}v_{-2}, \cdots$ be a sequence of independent random vectors which have the same distribution as the v_i's and are mutually independent of $x_0, v_0, v_1, \cdots$, and $w_1, w_2, \cdots$; such a sequence may always be found by possibly enlarging the probability space ([5], page 71). We now define random vectors u_k and q_k as

$$y_k = u_k - q_k, \tag{3.5}$$
$$u_k = w_k + \textstyle\sum_{j=0}^{\infty} A^j v_{k-j-1}, \qquad k \geqq 1, \tag{3.6}$$
$$q_k = A^k(\textstyle\sum_{j=0}^{\infty} A^j v_{-j-1} - x_0) = A^k q_0, \qquad k \geqq 0. \tag{3.7}$$

Using $\rho(A) < 1$, it may be shown by the Three Series Theorem ([3], page 111) that u_k and q_k are well-defined random vectors. Here u_k, $k \geqq 1$, is a moving average of the v_i's and w_j's and, therefore, a metrically transitive, strictly stationary process ([5], page 460); and $|q_k| \to 0$ with probability one and mean square as $k \to \infty$. Equation (3.2) is now a special case ($i = 1$) of the following lemmas.

Lemma 3.1. *Let $i \geqq 0$ be an integer. Then*

$$\text{(i)} \qquad E\{\|u_k u_{k-i}'\|\} \leqq E\{|u_k|^2\} = \operatorname{tr}(\Sigma) < \infty;$$
$$\text{(ii)} \qquad E\{u_k u_{k-i}'\} = A^i(\Sigma - W) + \delta_{i,0} W$$

where $\delta_{i,j}$ is the Kronecker δ. If A and V are nonsingular, then so is $E\{u_k u_{k-i}'\}$.

Lemma 3.2. *Let $i \geqq 0$ be an integer; then*

$$\lim_{n\to\infty} n^{-1} \textstyle\sum_{k=i+1}^{n} y_k y_{k-i}' = \lim_{n\to\infty} n^{-1} \sum_{k=i+1}^{n} u_k u_{k-i}' = E\{u_k u_{k-i}'\} \tag{3.8a}$$
$$\lim_{n\to\infty} n^{-1} \textstyle\sum_{k=i+1}^{n} \|y_k y_{k-i}'\| = \lim_{n\to\infty} n^{-1} \sum_{k=i+1}^{n} \|u_k u_{k-i}'\| = E\{\|u_k u_{k-i}'\|\}. \tag{3.8b}$$

Proof. Equation (ii) of Lemma 3.1 follows from (3.3) and (3.4) by a routine computation and the remark that $\Sigma - W$ is pd if V is nonsingular. Thereafter, (i) follows from

$$\begin{aligned} E\{\|u_k u_{k-i}'\|\} &= E\{|u_k|\,|u_{k-i}|\} \\ &\leqq E\{|u_k|^2\} \\ &= E\{\operatorname{tr}(u_k u_k')\} = \operatorname{tr}(\Sigma). \end{aligned}$$

The final equalities in (3.8a) and (3.8b) follow from the ergodic theorem and Lemma 3.1, since $u_k u_{k-i}'$, $k \geqq i$, is again a metrically transitive, strictly stationary process. Therefore, Lemma 3.2 would follow from

$$u_k u_{k-i}' - y_k y_{k-i}' = u_k q_{k-i}' + q_k u_{k-i}' + q_k q_{k-i}' \to 0 \tag{3.9}$$

with probability one as $k \to \infty$. Since $\|A^k\| \to 0$ exponentially fast as $k \to \infty$, (3.9) follows from (3.7) and the fact that $\sup_{k \geqq 1} k^{-1} |u_k| \leqq \sup_{N \geqq 1} N^{-1} \sum_{j=1}^{N} |u_j| < \infty$ with probability one. □

To establish Theorem 2.2, define

$$\bar{B}_{n,i} = n^{-1} \sum_{k=3}^{n} (y_k - A^i y_{k-i})(y_k - A^i y_{k-i})', \qquad n \geqq 3,\ i = 1, 2;$$

then by the ergodic theorem $\bar{B}_{n,i} \to B_i$ with probability one as $n \to \infty$ because the sequences $(y_k - A^i y_{k-i})(y_k - A^i y_{k-i})'$ are strictly stationary and $(i + 1)$-dependent and therefore metrically transitive. Therefore, it will be sufficient to show that

$$\begin{aligned} B_{n,i} - \bar{B}_{n,i} = n^{-1} \textstyle\sum_{k=3}^{n} y_k y_{k-i}'(A - \hat{A}_k)' + n^{-1} \sum_{k=3}^{n} (A - \hat{A}_k) y_{k-i} y_k' \\ + n^{-1} \textstyle\sum_{k=3}^{n} (\hat{A}_k - A) y_{k-i} y_{k-i}' A' \\ + n^{-1} \textstyle\sum_{k=3}^{n} \hat{A}_k y_{k-i} y_{k-i}' (\hat{A}_k - A)' \end{aligned} \tag{3.10}$$

converges to zero as $n \to \infty$, $i = 0, 1$. This, however, is an easy consequence of Lemma 3.2. For example, it follows from Lemma 3.2 that for any $m \geqq 3$

$$\begin{aligned} \lim_{n\to\infty} \sup \|n^{-1} \textstyle\sum_{k=3}^{n} y_k y_{k-i}'(A - \hat{A}_k)'\| \\ \leqq \sup_{k \geqq m} \|A - \hat{A}_k\| \lim_{n\to\infty} \sup n^{-1} \textstyle\sum_{k=m}^{n} \|y_k y_{k-i}'\| \\ = \sup_{k \geqq m} \|A - \hat{A}_k\| E\{\|u_k u_{k-i}'\|\} \end{aligned}$$

which may be made arbitrarily small by proper choice of m. The other sums in (3.10) may be handled similarly, thus completing the proof of Theorem 2.2.

4. Asymptotic behavior of S_k. In this section we will prove Theorem 2.3, which asserts the existence of a unique fixed point for the ϕ defined by $\phi(S) = A(S:W)A' + V$, $S \in \mathcal{S}$, the set of pd matrices. For this purpose we will obviously need to know some properties of the parallel sum $(S:W) = S(S + W)^{+}W$. Since we consider parallel addition only when one of the summands is pd, the pseudo-inverse appearing in its definition is really a true inverse. This fact simplifies the proof of the following lemmas considerably (cf [2]).

LEMMA 4.1. *Let $\mathcal{G}$ be the set of $(F, G) \in \mathcal{R}^p \times \mathcal{R}^p$ for which $F \geqq 0$, $G \geqq 0$, and $F + G > 0$: then*

(i) *parallel addition is continuous when restricted to $\mathcal{G}$;*
(ii) $F:G = G:F \geqq 0$, $(F, G) \in \mathcal{G}$;
(iii) *if $(F, G) \in \mathcal{G}$ and $F \leqq H$, then $(F:G) \leqq (H:G)$; and*
(iv) $F:G = F - F(F + G)^{-1}F \leqq F$, $(F, G) \in \mathcal{G}$.

PROOF. (i) is obvious since matrix inversion and multiplication are continuous operations. In the special case that F and G are pd, (ii) and (iii) are also obvious from $(F:G)^{-1} = F^{-1} + G^{-1}$; and the general case follows from the special one by considering $F_\epsilon = F + \epsilon I$ and $G_\epsilon = G + \epsilon I$ as $\epsilon \to 0$. Finally, (iv) follows from

$$F:G = F(F + G)^{-1}(F + G - F) = F - F(F + G)^{-1}F. \ \square$$

We will also need the following lemma, which is the easy half of a theorem due to Stein (see [8]).

LEMMA 4.2. *Let* $D \varepsilon \mathcal{R}^p$. *If there exists a* pd *matrix* F *for which* $F - D'FD$ *is* pd, *then* $\rho(D) < 1$.

PROOF. Let F be such a matrix and let λ be any (possibly complex) eigenvalue of D; then there is an $x \varepsilon R^p$ for which $Dx = \lambda x$ and, consequently,

$$x'(F - D'FD)x = (1 - |\lambda|^2)x'Fx > 0.$$

It follows that $|\lambda| < 1$ and, therefore, that $\rho(D) < 1$. □

The first step in the proof of Theorem 2.3 will be to verify that if $V > 0$, then $S_k = \phi(S_{k-1})$, $k \geqq 2$. If S_{k-1} is pd, then from (1.4) and Lemma 4.1 (iv)

$$S_k = A[S_{k-1} - S_{k-1}(S_{k-1} + W)^{-1}S_{k-1}]A' + V \tag{4.1}$$
$$= A(S_{k-1}:W)A' + V = \phi(S_{k-1}),$$

which is again pd by Lemma 4.1 (ii). Therefore, since S_1 is pd by (1.4), (4.1) must hold for $k \geqq 2$.

Next we show that ϕ has at most one fixed point. Toward this end we observe that if T_1 and T_2 are any two fixed points of ϕ, then by parts (ii) and (iv) of Lemma 4.1

$$T_1 - T_2 = A\{(T_1:W) - (T_2:W)\}A' = AW\{(T_2 + W)^{-1} - (T_1 + W)^{-1}\}WA'$$
$$= AW(T_1 + W)^{-1}\{T_1 - T_2\}(T_2 + W)^{-1}WA' = D_1^n(T_1 - T_2)D_2'^n,$$
$$n \geqq 1,$$

where $D_i = AW(T_i + W)^{-1}$; $i = 1, 2$. Therefore, it will suffice to show that $\rho(D_i) < 1$, $i = 1, 2$. For later reference we state this fact as

LEMMA 4.3. *Let* V *be* pd; *let* T *be any fixed point of* ϕ, *and let*

$$D = AW(T + W)^{-1};$$

then $\rho(D) < 1$.

PROOF. Since $\rho(D) = \rho(D')$ it will suffice by Lemma 4.2 to exhibit a pd matrix F for which $F - DFD'$ is pd; but $F = T$ is such a matrix, for

$$T - DTD' = AW(T + W)^{-1}TA' + V - AW(T + W)^{-1}T(T + W)^{-1}WA'$$
$$= AW(T + W)^{-1}T\{I - (T + W)^{-1}W\}A' + V$$
$$= AW(T + W)^{-1}T(T + W)^{-1}TA' + V$$
$$= AT(T + W)^{-1}W(T + W)^{-1}TA' + V. \quad \square$$

To complete the proof of Theorem 2.3, we observe first that for any $S \varepsilon \mathcal{S}$,

$$V \leqq \phi(S) = A(S:W)A' + V = A\{W - W(S + W)^{-1}W\}A' + V \tag{4.2}$$
$$\leqq AWA' + V$$

by Lemma 4.1 (iv). In particular, $V \leqq \phi(V)$, from which it follows by induction from Lemma 4.1 (iii) that

$$\phi^n(V) \leqq \phi^{n+1}(V) \leqq AWA' + V, \qquad n \geqq 1.$$

Therefore, $\lim_{n\to\infty} \phi^n(V) = S_0$ exists, ([18] page 263) and since ϕ is continuous on $\mathcal{S}$, S_0 must be a fixed point. Similarly, $\lim_{n\to\infty} \phi^n(AWA' + V)$ is a fixed point which must, therefore, equal S_0. The uniformity statement in Theorem 2.3 now follows from (4.2); indeed,

$$\phi^n(V) \leqq \phi^{n+1}(S) \leqq \phi^n(AWA' + V)$$

for all $S \varepsilon \mathcal{S}$ and $n \geqq 1$. □

To establish Corollary 2.2 let $A_n \to A$, $V_n \to V > 0$, and $W_n \to W$ with $V_n > 0$ and $W_n \geqq 0$ for all $n \geqq 1$; then, setting $S_0^n = S_0(A_n, V_n, W_n)$, $n \geqq 1$,

$$S_0^n = A_n(S_0^n : W_n)A_n' + V_n \leqq A_n W_n A_n' + V_n, \qquad n \geqq 1, \tag{4.3}$$

by (4.2). Therefore, S_0^n is bounded. Moreover, if S is any limit point of S_0, then $S = A(S:W)A' + V$ by Lemma 4.1 (i). Therefore, $S_0(A, V, W)$ is the unique limit point of S_0^n. □

Finally, we remark that if (2.5) were used to define ϕ on the set of all psd matrices, and if the requirement that V be pd were dropped, then the extended ϕ need not have a unique fixed point. For example, let $A = 2I$,

$$V = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad \text{and} \quad W = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix};$$

then it is easily verified that V and $V + 3W$ are both solutions of the equation $S = 4(S:W) + V$.

5. Asymptotic behavior of $\hat{S}_k$ and $\hat{x}_{k|k}$. In this section we will prove Theorems 2.4 and 2.5, which compare the asymptotic behaviors of $\hat{S}_k$ and $\hat{x}_{k|k}$ with those of S_k and $x_{k|k}$ respectively. To establish Theorem 2.4 it will clearly suffice to show that if $\hat{A}_n$, $\hat{V}_n$, and $\hat{W}_n$ are any fixed sequences of matrices for which $\hat{A}_n \to \hat{A}$, $\hat{V}_n \to V > 0$ and $\hat{W}_n \to W \geqq 0$, with $\hat{V}_n > 0$ and $\hat{W}_n \geqq 0$ for all $n \geqq 1$, then $\hat{S}_k \to S_0$, where $\hat{S}_k$ is defined by (2.3) and S_0 is as in Theorem 2.3. Let $\hat{A}_n$, $\hat{V}_n$, and $\hat{W}_n$ be such sequences and define ϕ_n, $n \geqq 1$ by

$$\phi_n(S) = \hat{A}_n(S:\hat{W}_{n-1})\hat{A}_n' + \hat{V}_n, \qquad S \varepsilon \mathcal{S};$$

then by (4.3) there is a compact subset $\mathcal{S}_0 \subseteq \mathcal{S}$ for which $\phi_n(\mathcal{S}) \subseteq \mathcal{S}_0$ $n \geqq 1$, and by Lemma 4.1 $\phi_n \to \phi$ uniformly on $\mathcal{S}_0$. We now observe that the estimate $\hat{S}_n$ of S_n may be written

$$\hat{S}_n = \phi_n \circ \cdots \circ \phi_2(\hat{S}_1), \qquad n \geqq 1,$$

where $\circ$ denotes composition, $\hat{S}_1 > 0$ by (2.3), and $\hat{S}_k \varepsilon \mathcal{S}_0$, $k \geqq 2$. Let $\epsilon > 0$; then by Theorem 2.3 there is an integer $r = r_\epsilon$ for which $\|S_0 - \phi^r(S)\| \leqq \epsilon$

for all $S \varepsilon \mathcal{S}$; and since $\mathcal{S}_0$ is compact, we may select a subsequence k_i, $i \geqq 1$, for which

$$\lim_{i\to\infty} \|\hat{S}_{k_i} - S_0\| = \lim_{k\to\infty} \sup \|\hat{S}_k - S_0\|, \qquad \lim_{i\to\infty} \hat{S}_{k_i - r} = T \varepsilon \mathcal{S}_0 .$$

By the uniform convergence of ϕ_n to ϕ on $\mathcal{S}_0$, we must then have $\lim_{i\to\infty} \hat{S}_k = \phi^r(T)$ and, therefore,

$$\lim_{k\to\infty} \sup \|\hat{S}_k - S_0\| = \lim_{i\to\infty} \|\hat{S}_{k_i} - S_0\| = \|\phi^r(T) - S_0\| \leqq \epsilon.$$

Since ϵ is arbitrary, Theorem 2.4 follows. □

Finally, to prove Theorem 2.5 we write

$$x_{k|k} = (\textstyle\prod_{i=1}^{k} G_i) x_{0|0} + \sum_{j=1}^{k} (\prod_{i=j+1}^{k} G_i) \Delta_j y_j$$

$$\hat{x}_{k|k} = (\textstyle\prod_{i=1}^{k} \hat{G}_i) \hat{x}_{0|0} + \sum_{j=1}^{k} (\prod_{i=j+1}^{k} \hat{G}_i) \hat{\Delta}_j y_j$$

where $G_k = (I - \Delta_k)A$ and $\hat{G}_k = (I - \hat{\Delta}_k)\hat{A}_k$, $k \geqq 1$. Now under the hypothesis of Theorem 2.5

$$G_k \to G = (I - \Delta_0)A = W(S_0 + W)^{-1}A, \quad \text{as} \quad k \to \infty,$$

where $\rho(G) = \rho(W(S_0 + W)^{-1}A) = \rho(AW(S_0 + W)^{-1}) < 1$ by Lemma 4.3. Therefore, there is an $r \geqq 1$ for which $\|G^r\| < 1$, and since $G_k \to G$ and $\hat{G}_k - G_k \to 0$ with probability one by Corollaries 2.1 and 2.3 respectively, there exist $\rho_0 < 1$ and a (random) integer $k_0 \geqq 1$ such that with probability one

$$\max \{ \|\textstyle\prod_{i=j}^{j+r} G_i\|, \|\prod_{i=j}^{j+r} \hat{G}_i\| \} \leqq \rho_0$$

whenever $j \geqq k_0$. In particular,

$$\max \{ \|\textstyle\prod_{i=1}^{k} G_i\|, \|\prod_{i=1}^{k} \hat{G}_i\| \} \to 0 \quad \text{with probability one}$$

as $k \to \infty$. It follows that for any $j_0 \geqq 1$

$$\begin{aligned} &\lim_{n\to\infty} \sup n^{-1} \textstyle\sum_{k=1}^{n} |\hat{x}_{k|k} - x_{k|k}| \\ &\leqq \lim_{n\to\infty} \sup n^{-1} \textstyle\sum_{j=1}^{n} \sum_{k=j}^{n} \|\prod_{i=j+1}^{k} \hat{G}_i \Delta_j - \prod_{i=j+1}^{k} G_i \Delta_j\| |y_j| \\ (5.1) \qquad &\leqq \lim_{n\to\infty} \sup n^{-1} \textstyle\sum_{j=j_0}^{n} \sum_{k=j}^{\infty} \prod_{i=j+1}^{k} \hat{G}_i \hat{\Delta}_j - \prod_{i=j+1}^{k} G_i \Delta_j\| |y_j| \\ &\leqq \sup_{j \geqq j_0} (\textstyle\sum_{k=j} \|\prod_{i=j+1}^{k} \hat{G}_i \hat{\Delta}_j - \prod_{i=j+1}^{k} G_i \Delta_j\|) (\lim n^{-1} \sum_{j=j_0}^{n} |y_j|) \\ &\leqq \sup_{j \geqq j_0} (\textstyle\sum_{k=j} \|\prod_{i=j+1}^{k} \hat{G}_i \Delta_j - \prod_{i=j+1}^{k} G_i \Delta_j\|) E\{|u_1|\} \end{aligned}$$

where the final inequality follows as in (3.8). Moreover, since $\hat{\Delta}_k \to \Delta_0 \leftarrow \Delta_k$ and $\hat{G}_k \to G \leftarrow G_k$ as $k \to \infty$, we have, for any fixed $j' \geqq 1$ that

$$\lim_{j\to\infty} \|\textstyle\prod_{i=j+1}^{j+j'} \hat{G}_i \Delta_j - \prod_{i=j+1}^{j+j'} G_i \Delta_j\| = 0.$$

It follows immediately that for any $s \geqq 1$

$$\begin{aligned} &\lim \sup_{j\to\infty} (\textstyle\sum_{k=j}^{\infty} \|\prod_{i=j+1}^{k} \hat{G}_i \hat{\Delta}_j - \prod_{i=j+1}^{k} G_i \Delta_j\|) \\ (5.2) \qquad &\leqq \lim \sup_{j\to\infty} \textstyle\sum_{k=j+rs}^{\infty} \|\prod_{i=j+1}^{k} \hat{G}_i \hat{\Delta}_j - \prod_{i=j+1}^{k} G_i \Delta_j\| \\ &\leqq 2\|\Delta_0\| r \rho_0^s / (1 - \rho_0), \quad \text{with probability one} \end{aligned}$$

which may be made arbitrarily small by proper choice of s. Theorem 2.5 follows easily from (5.1) and (5.2). ▯

6. Numerical results. A computer program embodying the estimators described above gave the results in Table 1. In this program the linear system (1.1) and (1.2) was scalar with normal noise and parameters $A = 0.9$, $V = 4.0$, and $W = 1.0$. The initial condition on x_k was $x_0 = 100.0$. The program simulated the system (1.1) and (1.2) and computed the estimators $\hat{A}_k$, $\hat{V}_k$, $\hat{W}_k$, and $\hat{\Delta}_k$ over periods of time of length 20, 40, 60, 80, 100, and 200. Fifty runs were made for each of these time periods.

TABLE 1

Time Period n	$\hat{A}_n$ mean	$\hat{A}_n$ variance	$\hat{V}_n$ mean	$\hat{V}_n$ variance
20	.899	$.152 \times 10^{-3}$	$.274 \times 10^{1}$	$.481 \times 10^{1}$
40	.899	$.110 \times 10^{-3}$	$.331 \times 10^{1}$	$.264 \times 10^{1}$
60	.900	$.141 \times 10^{-3}$	$.348 \times 10^{1}$	$.170 \times 10^{1}$
80	.899	$.547 \times 10^{-4}$	$.344 \times 10^{1}$	$.126 \times 10^{1}$
100	.900	$.138 \times 10^{-3}$	$.362 \times 10^{1}$	$.170 \times 10^{1}$
200	.901	$.768 \times 10^{-4}$	$.378 \times 10^{1}$	.624

Time Period	$\hat{W}_n$ mean	$\hat{W}_n$ variance	$\hat{\Delta}_n$ mean	$\hat{\Delta}_n$ variance
20	$.174 \times 10^{1}$	$.289 \times 10^{1}$	.639	$.961 \times 10^{-1}$
40	$.143 \times 10^{1}$	.903	.730	$.320 \times 10^{-1}$
60	$.140 \times 10^{1}$	.722	.748	$.212 \times 10^{-1}$
80	$.128 \times 10^{1}$	.569	.762	$.214 \times 10^{-1}$
100	$.124 \times 10^{1}$	.815	.780	$.223 \times 10^{-1}$
200	$.110 \times 10^{1}$	.232	.802	$.669 \times 10^{-2}$

Computation of Δ_k showed that it was stationary at the end of 20 time periods at $\Delta_k = 0.824$.

It was found that the parameter estimators are sensitive to the initial condition of the linear system. Occasionally when the system is at $x_0 = 0$ the fluctuations in the initial values of $\hat{A}_k$ cause the $B_{n,i}$ to assume extremely high values so that the corresponding means and variances of $\hat{V}_k$, and $\hat{W}_k$ display large dispersion. This problem does not arise when the initial conditions of the process differ from zero enough to give initial stability to $\hat{A}_k$.

REFERENCES

[1] ANDERSON, W. N. (1968). Series and parallel addition of operators. Ph.D. dissertation, Carnegie-Mellon Univ.

[2] ANDERSON, W. N. and DUFFIN, R. J. (1969). Series and parallel addition of matrices. *J. Math. Anal. Appl.* **26** 576–594.

[3] AOKI, M. (1967). *Optimization of Stochastic Systems*. Academic Press, New York.

[4] Cox, H. (1964). On the estimation of state variables and parameters for noisey dynamic systems. *IEEE Trans. Automatic Control.* **9** 9–12.
[5] Doob, J. L. (1953). *Stochastic Processes.* Wiley, New York.
[6] Griliches, Z. (1967). Distributed lags: a survey. *Econometrica* **37** 16–49.
[7] Ho, Y. C. and Lee, R. C. K. (1965). Identification of linear dynamic systems. *Information and Control* **8** 93–110.
[8] Householder, A. S. (1958). The approximate solution of matrix problems. *J. Assoc. Comput. Mach.* **5** 205–243.
[9] Jones, R. H. (1966). Exponential smoothing for multivariate time series. *J. Roy. Statist. Soc. Ser.* B **28** 286–293.
[10] Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Trans. ASME Ser. D. J. Basic Engineering* **82** 35–45.
[11] Kalman, R. E. (1963). New methods in Wiener filtering theory. *Proceedings First Symposium on Engineering Applications of Random Function Theory and Probability.* Wiley, New York.
[12] Kalman, R. E. and Bucy, R. S. (1960). New results in linear filtering and prediction problems. *Trans. ASME Ser. D. J. Basic Engineering* **82** 35–45.
[13] Kumar, K. S. and Sridhar, R. (1964). On the identification of linear systems. *Proc. Joint Automatic Control Conf. Palo Alto* 361–365.
[14] Loève, M. (1953). *Probability Theory.* Van Nostrand, Princeton.
[15] Loomis, L. (1953). *Abstract Harmonic Analysis.* Van Nostrand, Princeton.
[16] Mann, H. B. and Wald, H. (1943). On the statistical treatment of linear stochastic difference equations. *Econometrica* **11** 173–220.
[17] Rao, M. M. (1961). Consistency and limit distributions of estimators of parameters in explosive stochastic difference equations. *Ann. Math. Statist.* **32** 195–218.
[18] Riesz, F. and Sz-Nagy, B. (1955). *Functional Analysis.* Ungar, New York.
[19] Rubin, H. (1950). Consistency of maximum likelihood estimation in the explosive case. *Statistical Inference in Dynamic Economic Models.* Cowles Commission Monograph No. 10, Wiley, New York.