Performance of Adaptive Estimation Algorithms in Dependent Random Environments

ROBERT R. BITMEAD, MEMBER, IEEE, AND BRIAN D. O. ANDERSON, FELLOW, IEEE

Abstract—**We consider the convergence properties of certain algorithms arising in stochastic, discrete-time, adaptive estimation problems and operating in random environments of engineering significance. We demonstrate that the algorithms operating under ideal conditions are describable by homogeneous time-varying linear difference equations with dependent random coefficients, while in practical use, these equations are altered only through the addition of a driving term, accounting for time variation of system parameters, measurement noise, and system undermodeling. We present the concept of almost sure exponential convergence of the homogeneous difference equations as an *a priori* testable robustness property guaranteeing satisfactory performance in practice. For the three particular algorithms discussed, we present very mild conditions for the satisfaction of this property, and thus explain much of their observed behavior.**

I. Introduction

We consider here the study of the achievable performance of three important discrete-time stochastic adaptive estimation algorithms. While we discuss only three particular schemes, our approach does generalize, and consists of presenting an *a priori* testable property—exponential convergence—of the basic algorithms operating under ideal or "laboratory" conditions, and showing that this implies performance which is good, in a statistical sense, when the algorithms are used in real situations.

The underlying estimation problem that we consider is the following. Two measurable random time series, $\{y_k\}$ of scalars and $\{u_k\}$ of N-vectors, are known to be related in some way and we (perhaps arbitrarily) seek to model their relationship by postulating that they are, respectively, the output and input sequences of a linear dynamic system of a particular structure, parameterized by the finite-dimensional vector parameter w. We wish to recursively estimate the value w^* of w which minimizes some statistical criterion of fit of $\{\hat{y}_k(w)\}$, the output process of the parameterized *model* system with input $\{u_k\}$, to $\{y_k\}$, the "output" process of our hypothetical *plant* system with "input" $\{u_k\}$. The particular structures we consider are illustrated in Figs. 1 and 2.

The reason for choosing these two structures is that it can be shown, e.g., [1]–[3], that they represent canonical unknown blocks in the description of any linear finite-dimensional single-input single-output system. In these canonical realizations of unknown linear systems, we may often have feedback and filtering of $\{y_k\}$ contributing to $\{u_k\}$, although we shall not concern ourselves with the origins of $\{u_k\}$. These structures also are of separate interest for finite impulse response modeling and for adaptive control problems, as will be pointed out later.

Unless the two time series are stationary, it is not necessarily possible to find a fixed minimizing value w^* to solve the above estimation problem. The methods we consider are recursive stochastic *adaptive* estimation algorithms which, in case the optimum value w^* varies slowly with time, attempt to track the changing parameter. As is well known, it is often the case that there is some compromise in the performance of these algorithms when w^* is stationary in order to maintain the tracking ability in case w^* varies [4]. Usually, such algorithms are loosely based on stationary statistical estimation procedures, such as stochastic approximation, without certain gains going to zero.

Manuscript received August 20, 1979; revised February 18, 1980. Paper recommended by K. S. Narendra, Past Chairman of the Adaptive, Learning Systems, Pattern Recognition Committee. This work was supported by the Australian Research Grants Committee.

R. R. Bitmead was with the Department of Electrical Engineering, University of Newcastle, Newcastle, N.S.W., Australia. He is now with the Department of Electrical and Electronic Engineering, James Cook University of North Queensland, Queensland, Australia.

B. D. O. Anderson is with the Department of Electrical Engineering, University of Newcastle, Newcastle, N.S.W., Australia.

Fig. 1.

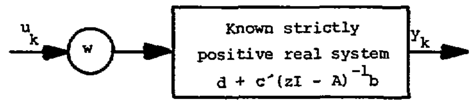

Fig. 2.

The equation describing the first structure for our parameterized model of the plant system (see Fig. 1) is

$$\hat{y}_k = u_k' w \tag{1}$$

and we consider the following two gradient-based algorithms for the recursive estimation of w^*. The first is the least mean-square (LMS) scheme of Widrow and Hoff [5] and others:

$$w_{k+1} = w_k + \mu u_k (y_k - \hat{y}_k) \tag{2}$$

where μ is a positive constant gain, which will be shown to be crucial to the convergence in the following analysis, and $\hat{y}_k$ is the output of the model parameterized by w_k. The correction term to the LMS algorithm has the interpretation as the estimated scaled negative gradient of the mean-squared error $E(y_k - u_k' w)^2$ at w_k, so that it resembles a steepest descent procedure [4].

The second algorithm is the scheme suggested by Albert and Gardner [6], Nagumo and Noda [7], and others which we shall call the normalized LMS (NLMS) scheme (rather than retaining the [6] terminology of "quick and dirty regression" scheme).

$$w_{k+1} = w_k + \mu \frac{u_k}{u_k' u_k} (y_k - \hat{y}_k). \tag{3}$$

This algorithm alters the magnitude of the correction without change in direction. It bypasses the problem of noise amplification for large u_k in LMS, but introduces problems for small u_k which can be overcome by using the alternate form:

$$w_{k+1} = w_k + \mu \frac{u_k}{\epsilon + u_k' u_k} (y_k - \hat{y}_k) \tag{4}$$

with $\epsilon > 0$.

For the second linear model structure of Fig. 2, we have

$$\begin{aligned} x_{k+1} &= A x_k + b u_k' w \\ y_k &= c' x_k + d u_k' w \end{aligned} \tag{5}$$

where $d + c'(zI - A)^{-1} b$ is a strictly positive real discrete-time transfer function with $\{A, b, c, d\}$ a minimal realization [8],[1] and we consider a more complicated update scheme which has its basis in Lyapunov design of model reference adaptive systems, originally presented by Butchart and Shackloth [9] and Parks [10] for continuous-time systems, and then derived for discrete-time systems by Landau [11] using hyperstability [12] and later by Ionescu and Monopoli [37] and Lin and Narendra [2] using Lyapunov functions. We use the notation of [2] since we shall be appealing to Lyapunov functions also. We have

$$\begin{aligned} \hat{x}_{k+1} &= A \hat{x}_k + b i_k \\ \hat{y}_k &= c' \hat{x}_k + d i_k \\ i_k &= u_k' w_k + (\delta + \alpha u_k' \Gamma u_k)(y_k - \hat{y}_k) \\ w_{k+1} &= w_k + \Gamma u_k (y_k - \hat{y}_k) \end{aligned} \tag{6}$$

[1]A *strictly* positive real discrete-time function is a positive real function [8] without poles or zeros on the unit circle, and with a strictly positive real part on the unit circle.

0018-9286/80/0800-0788$00.75 ©1980 IEEE

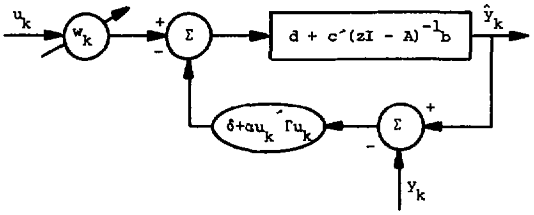

Fig. 3.

where δ is a small positive constant, α is a positive constant greater than $\frac{1}{2}$ (equal to one in [11]), and Γ is a constant positive definite gain matrix (this corresponds to proportional adaption in [11]). This adaption scheme, which we call the Lyapunov designed (LD) algorithm, resembles a copy of the model structure, with the insertion of an extra input at an intermediate point; see Fig. 3. It transpires that this extra term enters as a denominator as in NLMS, which is, in fact, a degenerate form of the above algorithm.[2]

We subdivide our performance analysis of the adaptive estimation schemes into two cases: 1) when there exists a fixed value w^* so that the plant system is exactly describable in terms of the model system parameterized by $w=w^*$, we call this the *homogeneous case* (for a reason which will subsequently become clear); 2) when such an exact description does not exist, we call this situation the *nonhomogeneous case*. The importance of this distinction is that essentially all applications are nonhomogeneous, as any modeling involves approximation, and usually the only *a priori* testable property of an estimation algorithm is its performance in the homogeneous case. The problem becomes one of relating the homogeneous performance to nonhomogeneous behavior, and then finding sufficient conditions for this former performance.

It should be remarked here that the nonhomogeneity of the application need not arise solely through unavoidable approximations, but may be an intentional consequence of the choice of feasible classes of model. This procedure of attempting to estimate the parameters of a model which is known to be very much simpler than the plant has recently been propounded by Young [13] for the modeling of badly defined systems and Goodwin and Ramadge [14] for the robust suboptimal control of linear systems.

The homogeneous performance property of the algorithms that we consider to be of particular importance is exponential convergence of the estimates w_k to w^*. We show that, subject to suitable definition for random variables, exponential convergence in the homogeneous case implies reasonable behavior in the nonhomogeneous case and, in an Appendix, we apply the recent results of the authors [15] to derive weak sufficient conditions for the $\{u_k\}$ input process to cause the algorithms to converge exponentially. This convergence property produces not only good asymptotic performance, but also adequate finite-time behavior.

The approach that we adopt here was initially suggested by Mendel [16, p. 290] for the deterministic case. We extend this by considering the random case and detailing more fully the effects of exponential asymptotic stability. We also treat a wider class of adaptive algorithms.

The broad variety of applications for these algorithms is indicated by the summaries of [14], [16], and [17], and it is apparent that this class of algorithm is of value when considering many cases of adaptive estimation and detection. We consider the study of these schemes worthwhile because of their applicability and also because there is a link between their performance and the finite time behavior of stochastic approximation schemes. Recently there has been some interest in improving the convergence rates of these latter schemes, and the results here may be of use.

II. Background

In the examination of the performance of the adaptive algorithms, it has been common to consider two modes of convergence. The first is convergence of the mean-squared parameter error from the optimum value, and the second is the convergence of the mean-squared output error $E(y_k-\hat{y}_k)^2$. Under mild assumptions, the statistical properties of the output error and those of the parameter error behave similarly so that we shall not place great importance on the distinction. We shall discuss both discrete-time and continuous-time results.

The LMS adaption algorithm has been analyzed by several people and, save for a handful, all have worked under the assumption of *independent random sequences* $\{u_k\}$. With this assumption plus stationarity, it is straightforward to prove convergence of the mean-squared parameter error and the mean-squared output error, as well as to tie an exponential rate to this convergence. This has been the approach of Widrow *et al.* [4], Gersho [18], Senne [19], and others.

Widrow *et al.* then use the exponential convergence rate of the algorithm to define an adaption time constant which characterizes the learning ability of the algorithm in a stationary environment. Furthermore, they show that the same constant determines the ability to track time-varying parameters in a nonstationary environment. They then examine the asymptotic performance in the stationary case (in terms of mean-squared output error) and demonstrate the tradeoff between this performance and tracking ability. They define a dimensionless measure of overall performance, misadjustment, and illustrate its dependence upon the gain μ. The correct design choice of the value of μ can then be regarded as an optimization problem for each particular application.

While their examination of the performance is heuristic and subject to very restrictive assumptions, it represents a major contribution to the application of the algorithm since it provides many useful rules of thumb for design which are known to work reasonably well in practice. Indeed, it is basically their approach which we emulate in part here, although our aim is to demonstrate that their harsh assumptions are not necessary and that the algorithms will perform adequately in dependent random environments. Our class of algorithms is larger also.

Most other authors concern themselves with the proof of convergence in the stationary case and neglect the analysis of nonstationary performance. The examination of NLMS by Nagumo and Noda falls into this class.

The LMS algorithm with stationary dependent input process $\{u_k\}$ has been studied by Davisson [20], Daniell [21], Kim and Davisson [22], Jones [23], and Farden, Goding, and Sayood [24].

Under the assumption that the mean-squared output error converges plus other conditions, Davisson establishes bounds on the output error variance for several different classes of dependent inputs. His conditions are difficult to verify, however, although in a subsequent paper with Kim, he demonstrates the convergence of the mean-squared parameter error and finds a bound for it. They consider M-dependent inputs and use a modified LMS algorithm which has a Markov structure for its derived input process.

Daniell, Jones, and Farden *et al.* all consider a general version of LMS and restrict their attention to the case of stationary almost surely bounded input processes (this is only partially true of Daniell) or to projected algorithms where it is known that the parameter vector lies in a given compact set. We also consider LMS only for this restricted class of inputs, although we realize that the need for this assumption represents a weakness of the current analytic techniques rather than a property of the algorithms. The available results for unbounded independent inputs [4], Markov inputs [23], and some dependent inputs [22], [24] seem to bear this idea out.

Daniell proves that, under a mild form of asymptotic independence of the inputs, for μ small enough, the parameter error variance will converge and may be made arbitrarily small with decreasing μ. Jones finds a bound on the asymptotic mean-squared output error, subject to convergence, and then proves convergence of both the mean-squared output error and mean-squared parameter error when the input sequence is a linear functional of a Markov process. He implicitly establishes a convergence rate of $1/t$ for the algorithm. Farden *et al.* establish bounds on, and prove convergence of, the asymptotic mean-squared parameter error in the case where the input is a full rank process with a covariance which decays at a specified rate. They show that this bound is approached exponentially fast and they examine its general μ-dependence for small μ. In each of the above five papers, it is established that decreasing μ

[2] While [2] and [11] do not have the arbitrary positive constant δ as we do, it is possible to consider their schemes at the same time since, as will be evident from Fig. 4, we may treat δ as an arbitrary small feedback to the strictly positive real system and this does not affect the strict positive real property.

decreases the mean-squared parameter error to an arbitrarily small value.

The LD algorithm has been shown to converge in the homogeneous case (i.e., where there exists an exact description of the plant as a parameterized model) with deterministic measurements $\{u_k\}$ and $\{y_k\}$ by Landau [11] and Johnson [25] using hyperstability, and by Lin and Narendra [2] using Lyapunov stability. They have not, however, established exponential convergence rates. To our knowledge, there has been no analysis of the convergence of the LD algorithm with random inputs other than that of Landau [11] and Ljung [26] who consider the situation where the gain matrix Γ tends to zero with time.

In continuous time, the LMS and NLMS schemes coincide, and both are consequently members of the class of LD algorithms which take on a simpler structure; see, e.g., [27]. It is not then surprising to find that sometimes all the algorithms may be analyzed by the same methods.

The convergence of the differential equations arising in adaptive estimation with deterministic inputs for the homogeneous case has been studied by Yuan and Wonham [28], Anderson [27], and Morgan and Narendra [29], [30]. Their conditions for convergence [28] or exponential convergence [27], [29], [30] are remarkably similar and consist of a persistently exciting condition together with a requirement of well behavedness or regulation of the functions involved. It appears that this regulation of the functions is a considerable stumbling block to the straightforward extension of these results to the case of stochastic input processes. In discrete time, this problem of regulation does not arise.

For continuous time adaption algorithms with random inputs, the chief references are Kushner [31] and Sondhi and Mitra [32] who both consider the LMS scheme. Kushner demonstrates the convergence in probability in the homogeneous case of the differential equation when the inputs are derived from a certain class of Markov process.

In [32], the almost sure satisfaction of a persistently exciting condition is required, and then exponential convergence is demonstrated for the homogeneous case. This is a trivial extension of the deterministic results cited above, and the condition is not satisfied by almost all sample paths of a stationary ergodic process with positive definite covariance, as claimed by the authors. In spite of these shortcomings, they then proceed to examine the performance of the algorithm in the nonhomogeneous case given that the homogeneous algorithm is exponentially convergent. Our analysis of nonhomogeneous performance in Section IV parallels their approach quite closely.

Throughout the analyses above, one of the chief aims has been to characterize the in-use performance of the algorithms and, in order to do this, to present reasonable assumptions on the input process to guarantee this performance. We see our main contribution as greatly extending the class of allowable inputs to admit a very wide collection of feasible engineering situations, while still ensuring the performance characterization.

III. The Homogeneous Algorithms

In this section, we present the algorithms as they occur in the homogenous case in a form which will demonstrate the suitability of the *a priori* testable performance requirement of exponential asymptotic stability. This latter property is defined for the random difference equations which arise in this application.

Recall that the LMS algorithm is

$$w_{k+1}=w_k+\mu u_k(y_k-u_k'w_k). \tag{7}$$

We desire that asymptotically w_k approaches w^* where, in the homogeneous case, $y_k=u_k'w^*$. Writing $v_k=w_k-w^*$, the parameter error, (7) becomes

$$v_{k+1}=v_k-\mu u_k u_k' v_k$$

or

$$\boxed{\text{LMS} \qquad v_{k+1}=(I-\mu u_k u_k')v_k \tag{8}}$$

which is a linear free difference equation. Similarly, the NLMS algorithm may be written as a linear error equation:

$$\boxed{\text{NLMS} \qquad v_{k+1}=\left(I-\mu\frac{u_k u_k'}{\epsilon+u_k'u_k}\right)v_k. \tag{9}}$$

Now the LD algorithm is described by the following equations in the homogeneous case:

$$x_{k+1}=Ax_k+bu_k'w^* \quad y_k=c'x_k+du_k'w^* \tag{10}$$

$$\left.\begin{aligned} &\hat{x}_{k+1}=A\hat{x}_k+bi_k \quad \hat{y}_k=c'\hat{x}_k+di_k \\ &i_k=u_k'w_k-(\delta+\alpha u_k'\Gamma u_k)(\hat{y}_k-y_k) \quad w_{k+1}=w_k+\Gamma u_k(y_k-\hat{y}_k) \end{aligned}\right\}. \tag{11}$$

Equation (10) defines the plant system, while (11) describes an adaptive observer-type model. Defining the state error $e_k=\hat{x}_k-x_k$ and output error $e_{0k}=\hat{y}_k-y_k$, and with v_k the parameter error as above,

$$e_{k+1}=Ae_k+bu_k'v_k-b(\delta+\alpha u_k'\Gamma u_k)e_{0k}, \tag{12}$$

and hence,

$$e_{0k}=c'e_k+du_k'v_k-d(\delta+\alpha u_k'\Gamma u_k)e_{0k}, \tag{13}$$

and subtracting w^* from w_{k+1} above,

$$v_{k+1}=v_k-\Gamma u_k e_{0k}. \tag{14}$$

Error equations (12), (13), and (14) describe the linear system illustrated in Fig. 4. We can solve (13) for e_{0k} in terms of e_k and v_k and substitute back into (12) and (14) to yield

$$\boxed{\text{LD} \quad \begin{pmatrix} e_{k+1} \\ v_{k+1}\end{pmatrix}=\begin{bmatrix} A-\dfrac{\delta+\alpha u_k'\Gamma u_k}{1+d\alpha u_k'\Gamma u_k+d\delta}\cdot bc' & \dfrac{1}{1+d\alpha u_k'\Gamma u_k+d\delta}\cdot bu_k' \\ -\dfrac{1}{1+d\alpha u_k'\Gamma u_k+d\delta}\cdot\Gamma u_k c' & I-\dfrac{d}{1+d\alpha u_k'\Gamma u_k+d\delta}\cdot\Gamma u_k u_k'\end{bmatrix}\begin{pmatrix} e_k \\ v_k\end{pmatrix}. \tag{15}}$$

Putting $g_k'=(e_k',v_k')$, we can write (15) as

$$g_{k+1}=F_k g_k \tag{16}$$

which is again a linear, free, time-varying difference equation.

We have shown that, in the homogeneous case, the algorithms are described by homogeneous linear equations. In the next section, we demonstrate that, in the nonhomogeneous case, the algorithms are altered only by the inclusion of a driving term to the equations, while maintaining the same homogeneous part. In order that the estimation schemes be useful in practice, we seek a robustness property of the convergence of the unforced difference equations.

In the homogeneous case, the asymptotic convergence of the algorithm is usually all that is required, and this can be inferred from the global asymptotic stability of the difference equations. Now the difference equations which are most tolerant of perturbations, in the sense of preserving qualitative properties of solutions, are those with an exponential stability property—exponential stability implies exponential convergence rate. Consequently, we consider first a definition of exponential convergence for random processes, and then, in the next section, the performance implications of exponential convergence of the homogeneous algorithms. We concern ourselves with finding a suitable property of the homogeneous algorithms because usually, although not always, we have insufficient knowledge to be able to characterize the perturbations exactly, e.g., if we model a slightly nonlinear system with unknown nonlinearity by a finite-dimensional linear model. Furthermore, methods for analyzing the convergence of linear homogeneous difference equations with random coefficients have already been derived by the authors in [15] where the following definition of exponential convergence of a stochastic process is given.

Definition: By exponential convergence to zero (almost surely, in mean square, in probability) of a stochastic process $\{z_k\}$, we mean that

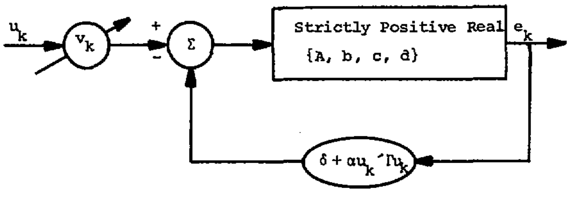

Fig. 4. LD error model.

the related stochastic process $\{(1+\beta)^k z_k\}$ for some $\beta>0$ independent of the realization converges to zero (almost surely, in mean square, in probability respectively) as $k\to\infty$. In this case, we say that $\{z_k\}$ converges exponentially faster than $(1+\beta)^{-k}$.

A difference equation with random state x_k is almost surely exponentially asymptotically stable if $\{x_k\}$ is almost surely exponentially convergent to zero.

This definition above differs from the usual uniform exponential convergence of deterministic sequences in that it refers to the average rate of convergence over arbitrarily long intervals rather than a stepwise rate. This is discussed further in [15].

Under the assumption of exponential convergence of the homogeneous part of the algorithms, we proceed in the next section to evaluate the nonhomogeneous performance. In the Appendix, we use the results of [15] to derive mild sufficient conditions for the exponential convergence of the homogeneous algorithms.

IV. The Nonhomogeneous Case—Performance

With the description of the homogeneous algorithms as derived in the previous section, we now address the problem of the description of the nonhomogeneous case algorithms. Equations (8), (9), and (15) will be referred to as the undriven error equations. In the next section, we shall present sufficient conditions for these equations to be exponentially stable, but for this section, we will assume that these conditions are satisfied in order that we can consider the performance implications of exponential convergence. We shall basically be appealing to the bounded input/bounded state property of exponentially stable linear systems, cf. [33, Theorem 3*] to make our claims concerning the performance of the schemes. This is akin to the approach suggested in [16].

In applying the results of [15] to prove homogeneous exponential convergence, we consider convergence with two basic classes of random input sequences: stationary ergodic and nonstationary ϕ-mixing (see [15] for a heuristic explanation of these technical concepts and references). These two classes encompass a very broad variety of feasible dependent random processes arising in engineering situations, requiring only mild asymptotic independence of widely separated events for their definition. The usefulness of the results to follow on nonhomogeneous performance stems from the mildness and robustness of these assumptions on the input sequence.

Recall that the nonhomogeneous case (which we associate with engineering applications of the adaptive estimators) occurs when there exists no exact, fixed representation of the plant in terms of the parameterized model. In this case, there exists no fixed w^* which causes $\hat{y}_k=y_k$ for all k and u_k, and indeed, if we rule out degenerate u_k, $\hat{y}_k\equiv y_k$ is, in general, not possible for any input. Therefore, we see that the correction terms in the update equations (7) and (11) never go identically to zero, and hence the algorithms need not converge.

Consistent with its definition in the homogeneous case, we redefine w_k^* to be that value of w which minimizes the mean-squared output error $E(y_k-\hat{y}_k)^2$ at time k. We redefine our parameter error $v_k=w_k-w_k^*$.

In examining the performance implications of exponential convergence of the estimates of the homogeneous algorithms, we classify the main perturbations, brought about in the nonhomogeneous application, into three groups. We assume their independence and treat each singly. This is usually justified because, as will be shown, the perturbations are introduced as driving terms adding on to the equations. As the equations are linear, we expect the responses to these perturbations to add. This is similar to the approach of Widrow *et al.* [4] and Sondhi and Mitra [32] for analyzing the behavior of discrete-time LMS and continuous-time NLMS algorithms under nonideal conditions.

Our discussion will necessarily be heuristic as, without accurately specifying the underlying probability space, we cannot be more precise. We defend this heuristic approach, however, by arguing that it provides an invaluable aid to thinking, particularly with regard to engineering practice where frequently the random processes met with are strongly mixing, bounded, and quite well-behaved, avoiding the sort of pathological situations which must be considered in an exact and general treatment. Finally, much of this heuristic argument is verified by computer simulation and real performance data, cf. [4], and can be formally validated in many cases [36].

We consider the three following general classes of nonhomogeneity.

1) Time Variation in the "True" Parameter Vector w^*: Here we suppose that w_k^* takes on different values at different times. There are two basic reasons for this nonstationarity—first, if the plant is completely describable in terms of the model with parameter w_k^*, and the plant itself changes with time so that w_k^* varies; and second, if the plant is fixed but is not in the model set, then a change in the statistical properties of the input and/or the output may cause w_k^* to vary, where w_k^* is the value of the parameter vector which minimizes the mean-squared output error at time k.

With this modified definition of $v_k=w_k-w_k^*$, the three update equations take on the form

$$v_{k+1}=G_k(v_k)-(w_{k+1}^*-w_k^*) \tag{17}$$

where G_k is a linear operator arising in the homogeneous algorithm.

$$\tilde{v}_{k+1}=G_k(\tilde{v}_k) \tag{18}$$

which is exponentially convergent. Equation (17) shows that the nonstationarity of w_k^* enters the error equations as an additive driving term.

2) Output Measurement Corrupted by Noise: Here we suppose that the sequence $\{y_k\}$ is not measurable, while the sequence $\{z_k\}$ where $z_k=y_n+n_k$ is measurable. Here $\{n_k\}$ is some noise sequence assumed zero-mean, white, and independent from $\{u_k\}$.

In this case, LMS becomes, from (7),

$$v_{k+1}=(I-\mu u_k u_k')v_k+\mu u_k n_k \tag{19}$$

and the other two algorithms are altered by the addition of $\mu u_k n_k(\epsilon+u_k'u_k)^{-1}$ and $-([\delta+\alpha u_k'\Gamma u_k]b'\ u_k'\Gamma u_k)' n_k$.

Again, we see that the perturbation manifests itself as an additive driving term to the exponentially stable difference equations. (Notice the practical role of ϵ in the denominator of NLMS—this prevents the noise from being amplified too greatly if u_k is very small.)

3) Model of Smaller Order than Plant: This situation arises when we do not allow sufficient flexibility in the class of parameterized models to completely describe the plant. In most applications, it is nearly always the case that slight nonlinearities are modeled as linear, saturation and hysteresis effects are ignored, etc., and the effective use of linearized models for a description of these systems is reliant on the robustness of the models. We aim to show here that the estimation schemes for identifying these linear models are also robust.

Clearly, this modeling error can be regarded as manifesting itself either as a rapid time variation of the parameter w_k^* or as an additive output error term with essentially static w_k^* which may be incorporated into equations similar to (19).

The general linear equation governing the nonhomogeneous algorithms with driving terms from sources 1), 2), and 3) is then seen to consist of an exponentially stable homogeneous part with additive driving terms. We write this general equation as

$$v_{k+1}=F_k v_k+\mu G_k n_k \tag{20}$$

where n_k is the driving term and, denoting the transition matrix of (20) as $\Phi(\cdot,\cdot)$, we see that

$$v_{k+1}=\Phi(k+1,0)v_0+\sum_{i=1}^{k}\mu\Phi(k+1,i)G_i n_i. \tag{21}$$

We may next take the variance and mean of (21) and invoke the mean-square exponential convergence of $\Phi(\cdot,\cdot)$ to demonstrate that

$$E(v_k) \doteq K_1(1-\beta)^k E(v_0) + \sum_{i=1}^{k} \mu K_1 (1-\lambda)^{k-i} E(G_i n_i) \tag{22}$$

and

$$E|v_{k+1}|^2 \leqslant K_2(1-\beta)^{2k} E|v_0|^2 + \sum_{i=1}^{k} \mu^2 K_2 (1-\beta)^{2(k-i)} E(n_i^2 G_i' G_i) \tag{23}$$

where $(1-\beta)^k$ is the convergence rate of the homogeneous equation. (We have used our heuristic license here in separating many of the expectations. This can be formally justified in the ϕ-mixing and other cases [36].)

Thus, we see from (22) that the initial condition offset of the mean is exponentially vanishing and, in the stationary case at least, there is a limiting bias of the estimate v_k of the order $\mu K_1 \beta^{-1} E(G_i n_i)$. For transversal filters with LMS or NLMS, this last expression is readily shown to be approximately equal to the difference between the filter consisting of the first N values of the plant system impulse response and the best "Wiener" solution. In case $\{n_k\}$ is zero-mean and independent from $\{u_k\}$, we have the estimates exponentially asymptotically unbiased—a slight improvement on Theorem 2.5 of Mendel [16].

From (23), provided we have bounded variance of the disturbances $\{n_k\}$, we are assured of bounded variance of the estimates under the mixing assumption. This limiting bound is approximately $\mu^2 K_2 \beta^{-1} E(n_i^2 G_i' G_i)$. As a general rule, true for independent and identically distributed $\{u_k\}$, β is proportional to μ(see [4], [32], and [34]) so that the bounds on the parameter variance indicate that this variance is linear in μ, to a first approximation at least. This result may also be gleaned from the work of Daniell [21], Davisson [20], and Jones [23], although their analyses concern operation under perturbations 2) and 3) only. This shows that we may force the parameter error variance due to noise and mismatch to be arbitrarily small by sacrificing the convergence rate and hence the tracking performance. There is obviously a compromise as illustrated in [4].

To summarize this section, we point out that exponential convergence of the homogeneous algorithm, even in the dependent input case, provides a guarantee of nonhomogeneous performance. Further, the choice of convergence rate is determined by a compromise between tracking speed and stationary error performance.

V. Sufficient Conditions for Good Performance

So far, we have described our operating adaptive estimation algorithms by nonhomogeneous linear time-varying difference equations, and we have proposed almost sure exponential asymptotic stability of the homogeneous part of these equations as a measure of good performance in applications. This was then examined in some detail in Section IV for the three classes of nonhomogeneity. With this characterization of the in-use performance via homogeneous exponential stability, we now go on to study sufficient conditions for this latter stability. The proofs are contained in the Appendix and are based on the results of [15] which have their origin in Lyapunov stability.

We consider the NLMS and LD algorithms with unbounded inputs and the more widely studied LMS algorithm with almost surely bounded inputs, i.e., we assume that there exists a number M such that for all k and almost all $\omega \in \Omega$, $|u_k(\omega)| \leqslant M$. This latter requirement, while being unpopular in theoretical analyses, is often easily justified in practice where there are always physical limits to the signals occurring and where, say, pseudorandom binary noise or periodic signals are used as "training" sequences in the initial adjustment of the estimator to the neighborhood of its operating point, e.g., [35]. Furthermore, as we shall see, the fact that the larger the bound M is, the smaller is the maximum admissible value of μ, does not appear as an unnatural byproduct of the analysis. Indeed, as was indicated earlier, in practical situations with slow time variation, we will want μ to be quite small because the misalignment due to extra signals driving our error equations is essentially linearly dependent upon μ and we desire to make this misalignment small while still being able to track slow time variations. A typical practical value of μ in communication networks is around $0.001 \div$ average input signal power [4], and the associated bound on the input, equivalent to this value of μ, is greatly above the possible level of any signal.

We have the following.

Theorem 1: With stationary ergodic input $\{u_k\}$ having positive definite covariance:

1) the homogeneous LMS algorithm is almost surely and mean-square exponentially convergent provided there exists a uniform a.s. bound M such that $|u_k| \leqslant M < \infty$ a.s. for all k and $\mu \in (0, 2/M^2)$;

2) the homogeneous NLMS algorithm is almost surely and mean-square exponentially convergent provided $\mu \in (0,2)$;

3) the homogeneous LD algorithm is almost surely and mean-square exponentially convergent provided $\alpha > \frac{1}{2}$ and $\Gamma = \Gamma' > 0$.

We now impose restrictions on the asymptotic dependence of the inputs by replacing ergodicity by ϕ-mixing, but consequently admit the inclusion of nonstationary processes. As remarked before, in a theoretical analysis of algorithms designed for use in nonstationary environment, it is necessary to derive convergence results which remain valid under nonstationarity.

Theorem 2: With nonstationary ϕ-mixing input $\{u_k\}$, with mixing constants $\{\phi_i\}$ satisfying $\sum_{i=1}^{\infty} \phi_i^{1/2} < \infty$, and having $\lim_{n\to\infty} \inf (1/n) \sum_{i=1}^{n} E[u_i u_i']$ positive definite, conclusions 1)–3) of Theorem 1 remain valid.

The requirements of ergodicity and ϕ-mixing as set out above are really to ensure that the large sample behavior is almost surely determined by the expected values. The conditions on the covariances are then perceived as stochastic analogs of the nondegeneracy conditions of [27]–[30]—the persistently spanning conditions now refer to a.s. sample average behavior. That these requirements are so easily related to the known deterministic conditions for exponential convergence would seem to indicate the success of the approach.

These results are pleasing from several viewpoints. First, it is the first time that exponential convergence (deterministic or stochastic) has been proved for the LD algorithm, although this is known in the continuous-time version. Second, it is a useful result concerning the algorithms which allows broad classes of realistic random inputs.

VI. Conclusion

The stochastic adaptive estimation problem has been presented, and we have discussed three commonly proposed algorithms for its solution. We have demonstrated that these algorithms are described by homogeneous time-varying linear difference equations in the ideal case where reality is assumed to conform precisely to our model structure. We then showed that, while performance in this ideal homogeneous case was usually all we could test *a priori*, the in-use behavior could be characterized by the homogeneous performance because the only perturbation to the describing equations was the introduction of a driving term accounting for time variation of system parameters, measurement noise, and/or system undermodeling.

Almost sure and mean-square exponential convergence of the homogeneous algorithms were then defined and proposed as a useful testable measure guaranteeing satisfactory nonhomogeneous performance. This latter property was demonstrated by appealing to the bounded-input bounded-output nature of exponentially stable linear equations. In making these observations, we have specifically tried to avoid unnecessary technical assumptions in order to enhance their practical usefulness.

In the previous section, we presented very mild conditions on the inputs to the algorithms which ensure almost sure and mean-square exponential convergence of their homogeneous parts. In line with our previous remarks, we have endeavored to make assumptions which are practical and feasible in engineering situations. To this end, we have requirements which admit a very broad class of dependent random inputs, both stationary and nonstationary.

The work embodied in this paper explains much of the observed behavior of these algorithms in use, particularly with regard to the robustness of their performance in poorly modeled applications and in situations of time variation, and should provide useful guides to thinking in the implementation of these schemes.

Appendix
Proofs of Theorems 1 and 2

In Section III, we showed that the three algorithms, LMS, NLMS, and, LD, may be described in the homogeneous case by homogeneous linear difference equations of the form [see (8), (9), and (15)]

$$x_{k+1}=F_k x_k$$

where F_k is a matrix of random coefficients. We shall use the results of [15] to prove a.s. exponential asymptotic stability (EAS) for LMS with bounded inputs, NLMS and LD. We refer the reader to [15] for details, and remark here that a.s. exponential asymptotic stability and mean-square exponentially asymptotic stability may be readily shown to be equivalent.

LMS

With $P_k=I$ for all k, the solution of the Lyapunov matrix equation

$$F_k' P_{k+1} F_k - P_k = -H_k H_k' \tag{23}$$

satisfies $-H_kH_k' \leq -\mu u_k u_k'(2-\mu M^2)$ for the LMS scheme where M is the a.s. uniform bound on $|u_k|$. For a.s. EAS, we need to consider the observability of the pair $[F_k, H_k]$ which is implied by the observability of $[F_k, u_k]$. Using [15, Theorem 3], we may choose bounded feedback matrix $K_k=\mu u_k$ and study the observability of $[F_k+K_kH_k', u_k]=[I,u_k]$. Now the observability Gramian of $[I,u_k]$ is

$$W_n(kn)=\sum_{i=0}^{n-1} u_{kn+i}u_{kn+i}'$$

and the requirements for almost sure exponential stability of [15] are satisfied if the conditions of Theorems 1 and 2, part 1) hold. This follows analogously to the proof of [15, Theorems 7 and 8].

NLMS

Solving the Lyapunov matrix equation (23) with $P_k=I$, we have

$$H_kH_k' = 2\mu\frac{u_ku_k'}{\epsilon+u_k'u_k} - \mu^2\frac{u_ku_k'}{(\epsilon+u_k'u_k)^2}\cdot u_k'u_k$$

$$=\mu(2-\mu)\frac{u_ku_k'}{\epsilon+u_k'u_k} + \mu^2\epsilon\frac{u_ku_k'}{(\epsilon+u_k'u_k)^2}$$

and a.s. EAS is determined by the observability of the pair $[F_k, H_k]$ which, in turn, is determined by that of the pair $[F_k+K_kH_k', H_k]$ for bounded K_k. Here we have

$$F_k = I-\mu\frac{u_ku_k'}{\epsilon+u_k'u_k}, \quad H_k=\left[[\mu(2-\mu)]^{1/2}\frac{u_k}{(\epsilon+u_k'u_k)^{1/2}} \quad \mu\epsilon^{1/2}\frac{u_k}{\epsilon+u_k'u_k}\right]$$

and choose

$$K_k=\left[\left(\frac{\mu}{2-\mu}\right)^{1/2}\frac{u_k}{(\epsilon+u_k'u_k)^{1/2}} \quad 0\right]$$

to yield $F_k+K_kH_k'=I$. The proof then follows similarly to that of [15, Theorem 7 and 8], yielding sufficient conditions as in the current Theorems 1 and 2, part 2).

LD

Here we shall use the Kalman–Yacubovich lemma from [8], which states that if $\{A,b,c,d\}$ is a minimal realization of a discrete-time strictly positive real function, then there exists a real symmetric positive definite matrix P, a real vector L, and real numbers $\sigma, \nu>0$ such that

$$A'PA-P=-LL'-\sigma^2 I$$
$$A'Pb=c+\nu L$$
$$2d-b'Pb=\nu^2$$

with $[A,L]$ completely observable.

Using the matrix F_k of (15), we consider the solution of the Lyapunov matrix equation (23) with $P_k=P\oplus\Gamma^{-1}$ where P is as above and Γ is the positive definite gain matrix of the algorithm. We may factor H_kH_k' as

$$H_k=\begin{pmatrix} \sigma I & L+\nu(\delta+\alpha u_k'\Gamma u_k)\chi_k c & \xi_k\chi_k c \\ 0 & -\chi_k\nu u_k & \xi_k\chi_k d u_k\end{pmatrix}$$

where $\chi_k=(1+d\delta+d\alpha u_k'\Gamma u_k)^{-1}$, $\xi_k=[2\delta+(2\alpha-1)u_k'\Gamma u_k]^{1/2}$.

Rather than directly consider the observability of $[F_k, H_k]$, we consider the uniformly bounded output feedback matrix

$$K_k=\begin{pmatrix} 0 & \nu^{-1}b & 0 \\ 0 & 0 & \xi_k^{-1}\Gamma u_k\end{pmatrix}$$

and study the pair $[F_k+K_kH_k', H_k]=[(A+\nu^{-1}bL')\oplus I, H_k]$. Furthermore, we may write

$$H_k=\begin{pmatrix}\sigma I & 0 & 0 \\ 0 & 0 & u_k\end{pmatrix}\begin{bmatrix} I & 0 & 0 \\ -\sigma^{-1}(d\nu^{-1}L+c) & \nu(\delta+\alpha u_k'\Gamma u_k)\chi_k I & \xi_k\chi_k I \\ \sigma^{-1}\nu^{-1}L & -\chi_k\nu I & \xi_k\chi_k d I\end{bmatrix}$$

where this second matrix and its inverse have uniformly bounded entries. Consequently, $H_kH_k' \geq a_1(\sigma I\oplus u_k)(\sigma I\oplus u_k)'$ for some positive constant a_1 so that $[F_k, H_k]$ will be observable if $[A+\nu^{-1}bL', \sigma I]$ and $[I, u_k]$ are observable. The first pair is always observable, while the second is if the conditions of parts 3) of Theorems 1 and 2 are satisfied.

References

[1] B. D. O. Anderson, "An approach to multivariable system identification," *Automatica*, vol. 13, pp. 401–408, 1977.
[2] Y. -H. Lin and K. S. Narendra, "A new error model for discrete systems and its application to adaptive identification and control," Syst. and Inform. Sci., Yale Univ., New Haven, CT, Rep. 7802, Oct. 1978.
[3] G. Lüders and K. S. Narendra, "Stable adaptive schemes for state estimation and identification of linear systems," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 841–847, Dec. 1974.
[4] B. Widrow, J. M. McCool, M. G. Larimore, and C. R. Johnson, Jr., "Stationary and nonstationary learning characteristics of the LMS adaptive filter," *Proc. IEEE*, vol. 64, pp. 1151–1162, Aug. 1976.
[5] B. Widrow and M. E. Hoff, Jr., "Adaptive switching circuits," in *IRE Wescon Conv. Rec.*, part. 4, 1960, pp. 96–104.
[6] A. E. Albert and L. S. Gardner, Jr., *Stochastic Approximation and Nonlinear Regression*. Cambridge, MA: M.I.T. Press, 1967.
[7] J. -I. Nagumo and A. Noda, "A learning method for system identification," *IEEE Trans. Automat. Contr.*, vol. AC-12, pp. 282–287, June 1967.
[8] K. L. Hitz and B. D. O. Anderson, "Discrete positive-real functions and their application to system stability," *Proc. IEE*, vol. 116, pp. 153–155, Jan. 1969.
[9] R. L. Butchart and B. Shackloth, "Synthesis of model reference adaptive control systems by Lyapunov's second method," in *Proc. 1965 IFAC Symp. Adaptive Contr.*, Teddington, England.
[10] P. C. Parks, "Liapunov redesign of model reference adaptive control systems," *IEEE Trans. Automat. Contr.*, vol. AC-11, pp. 362–367, July 1966.
[11] I. D. Landau, "Unbiased recursive identification using model reference adaptive techniques," *IEEE Trans. Automat. Contr.*, vol. AC-21, pp. 194–202, Apr. 1976.
[12] V. M. Popov, *Hyperstability of Control Systems*. New York: Springer-Verlag, 1973.
[13] P. C. Young, "A general theory of modeling for badly defined systems," in *Modeling and Simulation of Land, Air and Water Resources Systems*, G. C. Vansteenkiste, Ed. Amsterdam: North Holland/American Elsevier, 1978.
[14] G. C. Goodwin and P. J. Ramadge, "Design of restricted complexity adaptive regulators," in *Proc. 17th IEEE Conf. Decision Contr.*, San Diego, CA, Jan. 1979.
[15] R. R. Bitmead and B. D. O. Anderson, "Lyapunov techniques for the exponential stability of linear difference equations with random coefficients," this issue, pp. 782–787.
[16] J. M. Mendel, *Discrete Techniques of Parameter Estimation: The Equation Error Formulation*. New York: Marcel Dekker, 1973.
[17] I. D. Landau, "A survey of model reference adaptive techniques—Theory and applications," *Automatica*, vol. 10, pp. 353–379, 1974.
[18] A. Gersho, "Adaptive equalization of highly dispersive channels for data transmission," *Bell Syst. Tech. J.*, vol. 48, pp. 55–70, Jan. 1969.
[19] K. D. Senne, "An exact solution to an adaptive linear estimation problem," U. S. Air Force Syst. Command, Frank J. Seiler Res. Lab., Tech. Rep. SRL-TR-70-0014, Sept. 1970.
[20] L. D. Davisson, "Steady-state error in adaptive mean-square minimization," *IEEE Trans. Inform. Theory*, vol. IT-16, pp. 382–385, July 1970.
[21] T. P. Daniell, "Adaptive estimation with mutually correlated training sequences," *IEEE Trans. Syst. Sci. Cybern.*, vol. SSC-6, pp. 12–19, Jan. 1970.
[22] J. -K. Kim and L. D. Davisson, "Adaptive linear estimation for stationary M-dependent processes," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 23–31, Jan. 1975.
[23] S. K. Jones, "Adaptive filtering with correlated training samples: A mean square error analysis," Ph.D. dissertation, Dep. Elec. Eng., Southern Methodist Univ., Dallas, TX, Jan. 1973.

[24] D. C. Farden, J. C. Goding, and K. Sayood, "On the 'desired behavior' of adaptive signal processing algorithms," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Washington, DC, Apr. 1979.
[25] C. R. Johnson, Jr., "A convergence proof for a hyperstable adaptive recursive filter: HARF," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 745–749, Nov. 1979.
[26] L. Ljung, "On positive real transfer functions and the convergence of some recursive schemes," *IEEE Trans. Automat. Contr.*, vol. AC-22, pp. 539–551, Aug. 1977.
[27] B. D. O. Anderson, "Exponential stability of linear equations arising in adaptive identification," *IEEE Trans. Automat. Contr.*, vol. AC-22, pp. 83–88, Feb. 1977.
[28] J. S.-C. Yuan and W. M. Wonham, "Probing signals for model reference identification," *IEEE Trans. Automat. Contr.*, vol. AC-22, pp. 530–538, Aug. 1977.
[29] A. P. Morgan and K. S. Narendra, "On the uniform asymptotic stability of certain linear nonautonomous differential equations," *SIAM J. Contr. Optimiz.*, vol. 15, pp. 5–24, Jan. 1977.
[30] ——, "On the stability of nonautonomous differential equations $\dot{x}=[A+B(t)]x$, with skew symmetric matrix $B(t)$," *SIAM J. Contr. Optimiz.*, vol. 15, pp. 163–176, Jan. 1977.
[31] H. J. Kushner, "On the convergence of Lion's identification method with random inputs," *IEEE Trans. Automat. Contr.*, vol. AC-15, pp. 652–654, Dec. 1970.
[32] M. M. Sondhi and D. Mitra, "New results on the performance of a well-known class of adaptive filters," *Proc. IEEE*, vol. 64, pp. 1583–1597, Nov. 1976.
[33] R. E. Kalman and J. E. Bertram, "Control system analysis and design via the 'second method' of Lyapunov, II. Discrete-time systems," *Trans. ASME, Ser. D., J. Basic Eng.*, vol. 82, pp. 394–400, 1960.
[34] R. R. Bitmead and B. D. O. Anderson, "Exponentially convergent behavior of simple stochastic estimation algorithms," in *Proc. 17th IEEE Conf. Decision Contr.*, San Diego, CA, 1979.
[35] S. H. Qureshi, "Fast start-up equalization with periodic training sequences," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 553–563, Sept. 1977.
[36] R. R. Bitmead, "Convergence in distribution of LMS-type adaptive parameter estimates," Dep. Elec. Eng., Univ. Newcastle, Australia, Tech. Rep. EE7912, June 1979.
[37] T. Ionescu and R. Monopoli, "Discrete model reference adaptive control with an augmented error signal," *Automatica*, vol. 13, pp. 507–517, 1977.

Scattering Theory and Linear Least-Squares Estimation, Part III: The Estimates

GEORGE VERGHESE, MEMBER, IEEE, BENJAMIN FRIEDLANDER, MEMBER, IEEE, AND THOMAS KAILATH, FELLOW, IEEE

***Abstract*—We extend our earlier work on providing a scattering model for state-space estimation, by interpreting the smoothed state estimate as a propagating wave in a scattering medium, with the observed system outputs acting as internal sources in the medium. The value of this interpretation in providing simple pictorial derivations, in guessing new results, and in visualizing and organizing old and new estimation results is demonstrated by analyzing the effects of altered *a priori* information, and of partitioning of data and/or computations.**

I. Introduction

Scattering theory analyzes the propagation and interaction of waves in a medium, through the study of the transmission and reflection properties of the medium. The fact that Riccati equations arise naturally in this analysis motivated us, in Parts I and II, [1], [2], to consider the interpretation of certain equations of least-squares state-space estimation theory (where the Riccati equation also plays an important role) in terms of a scattering model. This scattering framework was then used (see also [3]–[7] and the survey [8]) to simply derive and organize known results in estimation theory, for both continuous- and discrete-time state-space systems. Furthermore, several new results were suggested by and established in this framework [3]–[8].

The emphasis in the papers [1]–[6], however, was on the medium parameters because the interpretation of the waves themselves was not clear. In this paper we extend our earlier picture by focusing on the propagating quantities. We begin in Section II with the Hamiltonian form of the solution to the problem of determining smoothed estimates of the state of a continuous-time dynamic system from observations of its output [9], [10]. The Hamiltonian, better known in control than in estimation theory, is based on the introduction of an adjoint system that converts the original problem to a two-point boundary-value problem. We show that this solution leads immediately to a scattering model, with the smoothed state estimate and the adjoint state as propagating quantities and the given observations as internal sources in the scattering medium.

Having identified a scattering model in the estimation context, we proceed in Section III to develop a small part of the calculus of scattering operators, following Redheffer [11], and demonstrate its use in obtaining the basic state-space estimation formulas. The most significant feature of this approach is that it enables quick and simple "pictorial" derivations, with only elementary algebraic manipulations.

This is further illustrated in Section IV. We show that the scattering framework provides an efficient way to determine the effect on estimates of altering *a priori* data in the estimation problem. We also bring out the fact that many of the formulas in the literature correspond merely to different ways of looking at propagation through sections of a scattering medium. One of these is actually via an isomorphic framework based on the so-called transmission operator; we briefly describe this in the Appendix.

The main point we wish to illustrate in this paper is the ease and naturalness with which the scattering picture, and simple flow graph manipulations, lead to simple derivations of results that are far from transparent in other approaches.

As regards related work, Nicholson [12], [13], in relating some of Kron's ideas to those of scattering theory, embedded certain solutions of the smoothing problem in a scattering framework, but did not significantly exploit this framework for estimation. More recently Sidhu and Desai [14] obtained some of the results of [1], [2] by essentially the reverse approach: following Reid [15], who developed his ideas in parallel with Redheffer, they study two-point boundary value problems via a higher dimensional extended Riccati equation. The interpretation of the estimates as propagating waves in a scattering medium was presented by us previously in [7] and [16] using a slightly different derivation than the one given here.

Finally, we wish to note that our scattering analysis of the Hamiltonian equations follows lines similar to the method of invariant imbedding, as developed and applied (in several fields) by R. Bellman and his colleagues, see the references in the book [17]. Our own work is based more explicitly on Redheffer's comprehensive formulation of transmission line scattering theory, but we are aware that the rich variety of applications noted in [17] can be the source of further insights and extensions in our study of estimation problems.

II. Hamiltonian Equations and a Scattering Picture for State Estimates

We shall be interested in linear, least-mean-square-error, filtered and smoothed estimates of the state $x(\cdot)$ of a linear dynamic system driven by white noise, given noise-corrupted observations $y(\cdot)$ of the system output. Thus,

$$\dot{x}(s)=F(s)x(s)+u(s) \tag{1a}$$

and

$$y(s)=H(s)x(s)+v(s) \tag{1b}$$

where $u(\cdot)$ and $v(\cdot)$ are zero-mean, white-noise processes, with

Manuscript received December, 9, 1977; revised December 5, 1979, and March 3, 1980. Paper recommended by A. H. Haddad, Past Chairman of the Estimation Committee. This work was supported by the Air Force Office of Scientific Research, U.S. Air Force Systems Command, under Contract AF49-620-79-C-0058, and by the Joint Services Electronics Program under Contract DAAG-29-79-C-0047.

G. Verghese is with the Electric Power Systems Engineering Laboratory, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139.
B. Friedlander is with Systems Control, Inc., Palo Alto, CA 94304.
T. Kailath is with the Information Systems Laboratory, Department of Electrical Engineering, Stanford University, Stanford, CA 94305.

0018-9286/80/0800-0794$00.75 ©1980 IEEE