is not satisfied where Ψ_{max} is the largest subspace satisfying (2.1a) and (2.1c), what we can do is either increase α and adjust K suitably, or if R is the real field, just find the orthogonal projection L on $\Psi_{max}+\{U\}$, let $LK'=\bar{K}'$, and try to obtain a minimal order stable observer for $\bar{K}$. This way, for a fixed α, one can obtain an approximate K-observer.

Another approach to obtaining stable K-observers would be to parameterize the solutions possibly using the algorithm of [1] and then applying the decision methods of [10].

Among the possible applications of a stable K-observer would be inverse filters which can be used in input estimation in areas such as seismic data processing [6]–[8] and image processing [9]. Here we usually have a system

$$x_{k+1}=Fx_k+\bar{G}v_k$$

$$y_k=Hx_k+\bar{J}w_k$$

where v_k is to be estimated from the knowledge of F, G, $\bar{H}$, $\bar{J}$, and y_k, and w_k is the measurement noise. Then if we let

$$G=\left[\bar{G} \vdots 0\right], \quad J=\left[0 \vdots \bar{J}\right], \quad u_k=\begin{bmatrix} v_k \\ w_k \end{bmatrix},$$

we have $S=(F,G,H,J)$ for which a partial inverse to find only v_k is required. Thus, we see that if we can find such a stable K-observer, its output will estimate v_k asymptotically.

Other possible applications of K-observers could be decentralized control and regulator problems where information about the unknown inputs applied from the other channels and about the noise and their effects to the state could be obtained asymptotically.

References

[1] E. Emre, L. M. Silverman, and K. Glover, "Generalized dynamic covers for linear systems with applications to deterministic identification and realization problems," *IEEE Trans. Automat. Contr.*, vol. AC-22, Feb. 1977.

[2] F. M. Yuan, "Minimal dimension inverses of linear sequential circuits," *IEEE Trans. Automat. Contr.*, vol. AC-20, pp. 42–52, Feb. 1975.

[3] E. Emre and L. M. Silverman, "Minimal dynamic inverses for linear systems with arbitrary initial states," *IEEE Trans. Automat. Contr.*, vol. AC-21, pp. 766–769, Oct. 1976.

[4] W. M. Wonham and A. S. Morse, "Feedback invariants of linear multivariable systems," *Automatica*, vol. 8, pp. 93–100, 1972.

[5] T. E. Fortmann and D. Williamson, "Design of low-order observers for linear feedback control laws," *IEEE Trans. Automat. Contr.*, vol. AC-17, pp. 301–308, June 1972.

[6] J. M. Mendel, "White noise estimators for seismic data processing in oil exploration," Electron. Sci. Lab., Univ. Southern California, Los Angeles, Interim Rep., Nov. 1975.

[7] E. A. Robinson, "Predictive decomposition of seismic traces," *Geophysics*, vol. 22, pp. 767–778, Oct. 1957.

[8] J. F. Claerbout and E. A. Robinson, "The error in least squares inverse filtering," *Geophysics*, vol. 29, Feb. 1964.

[9] A. O. Aboutalib and L. M. Silverman, "Restoration of motion degraded images," *IEEE Trans. Circuits Syst.*, vol. CAS-22, pp. 278–286, Mar. 1975.

[10] B. D. O. Anderson, N. K. Bose, and E. I. Jury, "Output feedback stabilization and related problems via decision methods," *IEEE Trans. Automat. Contr.*, vol. AC-20, pp. 53–56, Feb. 1975.

[11] E. Emre and M. L. J. Hautus, "A polynomial characterization of (A, B)-invariant and reachability subspaces," *SIAM J. Contr. Optimiz.*, July 1980.

Lyapunov Techniques for the Exponential Stability of Linear Difference Equations with Random Coefficients

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

***Abstract*—We consider an approach to studying the exponential stability of linear difference equations with random coefficients through the use of Lyapunov stability techniques. The equations we study are of a form familiar from adaptive estimation algorithms, which motivates the examination. It is necessary to define the almost sure exponential convergence of a random process, and then to derive sufficient conditions on the coefficients of the difference equations to ensure the almost sure exponential convergence of the state. We consider, in particular, two very reasonable types of random coefficients—ergodic and stationary and ϕ-mixing and nonstationary—which would appear to encompass many engineering situations. An example of the power of the theory is given, where it is applied to a common adaptive filtering algorithm to derive mild conditions for exponential convergence with dependent random inputs.**

Manuscript received June 8, 1979; revised March 25, 1980. Paper recommended by P. E. Caines, Past Chairman of the Identification 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.

I. Introduction

In the field of model reference adaptive parametric estimation, the occurrence of linear error equations describing certain important algorithms is well known [1],[2], and the asymptotic stability of these error equations is identified with convergence of the estimation algorithm to the desired value. Noting this stability–convergence relationship, it transpires that many estimation algorithms are designed on the basis of Lyapunov stability of the error equations. This approach was first presented by Butchart and Shackloth [3] and Parks [4] and has subsequently been discussed by numerous others.

In the case where there exists a fixed value of the parameter which will produce identically zero model reference error for all inputs, the linear error equations above are homogeneous, while in the case where such a fixed value does not exist, the equations are simply altered by an additive driving term which accounts for time-variation, measurement noise, and undermodeling. The former case corresponds to there being an exact description of the system under consideration within the class of models proposed, and the latter case corresponds to the more reasonable engineering situation of inexact modeling. The importance of this distinction is that essentially all applications of model reference adaptive estimation will have nonhomogeneous error equations, as any modeling involves approximation, and usually the only *a priori* testable property of an estimation algorithm is its performance in the homogeneous case of exact modeling. The problem becomes one of relating the *a priori* testable behavior to the in-use performance so that we may ensure good results for tracking, smoothing, and modeling. That is, we require a robustness property of the solution to the homogeneous error equations.

The linear equations which are most tolerant of perturbations, in the sense of preserving qualitative properties of solutions, are those with an exponential stability property. Consequently, the analysis of exponential asymptotic stability (EAS) has been seen to be important [5],[6]. Here we deal exclusively with discrete-time systems and establish, early on, conditions for EAS of linear deterministic homogeneous difference equations which represent a direct extension of the continuous-time result of [7].

An important case in the application of these algorithms arises when the sequences driving the algorithm are stochastic in nature. In this situation, the randomness of the sequences manifests itself through the coefficients of the linear homogeneous error equation and the nonhomogeneous driving term being random. The *a priori* testable robustness property we consider for these linear difference equations with random coefficients is a modified form of random exponential asymptotic stability. An analysis of the in-use performance implications of this random EAS is presented in [27], although some results along these lines are also contained in [8].

While the random case is important in applications, the asymptotic stability of the algorithms has appeared to be quite difficult to analyze without restrictive assumptions on the random driving sequences such as independent and identically distributed inputs [9],[10], Markovian inputs [11], and M-dependent inputs [12]. Clearly, we need methods which are suitable for more general dependent sequences and for nonstationary environments, as this will then bring us more into the realm of real engineering and will enhance the usage value of the results.

We consider here two important cases of random coefficients, strictly stationary ergodic and nonstationary ϕ-mixing (defined later), which seem to include many feasible random situations. We establish weak sufficient conditions on the random coefficients for a linear homogeneous difference equation to be EAS in the random sense, and then

0018-9286/80/0800-0782$00.75

consider the ergodic and ϕ-mixing cases in particular. Our results follow from a careful study of Lyapunov techniques for linear difference equations and some familiarity with this theory is assumed.

As a demonstration of the power of these results, we establish very mild sufficient conditions for the EAS of a common adaptive estimation algorithm. Indeed, these conditions are almost necessary for stability as well.

The approach to the performance analysis of adaptive algorithms for the estimation of time-varying parameters, via the asymptotic stability of a time-varying linear homogeneous difference equation, appears to have been first suggested by Mendel [13, p. 290]. His arguments may be taken as some motivation for this current work, although we extend his consideration by examining EAS and allowing random coefficients.

In [27], we present a detailed application of the results of this paper to several other adaptive estimation schemes, as well as a discussion of the implications of EAS in the random case in applications.

II. Deterministic Linear Systems

We consider here the following discrete-time, time-varying free linear dynamic system:

$$x_{k+1}=F_kx_k \tag{1}$$

where the $\{x_i\}$'s are state N-vectors and the $\{F_i\}$'s are $N\times N$ matrices representing the coefficients of the N coupled difference equations. We shall be examining requirements for the exponential asymptotic stability (EAS) of (1) in cases where the $\{F_i\}$'s are random matrices, but first it is helpful to derive the requirements for deterministic $\{F_i\}$.

Since (1) is linear, we choose the familiar quadratic form $L(x,k)=x'P_kx$ for symmetric positive definite P_k as a candidate Lyapunov function. This is useful for two reasons: first, quadratic Lyapunov functions have been widely studied; and second, exponential convergence to zero of $L(x,k)$ along all trajectories of (1) is necessary and sufficient for exponential asymptotic stability of (1), provided that P_k and P_k^{-1} are bounded uniformly from above. We have the following result.

Theorem 1: Consider the system

$$x_{k+1}=F_kx_k. \tag{1}$$

If there exists a symmetric matrix function P_k and positive constants a, b, and c such that for all k

$$0<aI\leq P_k\leq bI<\infty, \tag{2}$$

and such that, for some matrix H_k,

$$F_k'P_{k+1}F_k-P_k=-H_kH_k' \tag{3}$$

with for all k and some n

$$W_n(k)=\sum_{i=0}^{n-1}\Phi'(k+i,k)H_{k+i}H'_{k+i}\Phi(k+i,k)\geq cI>0 \tag{4}$$

where $\Phi(\cdot,\cdot)$ is the transition matrix of the difference equation (1), then $L(x,k)=x'P_kx$ is a Lyapunov function for the system and the origin $x=0$ of the system is EAS.

Proof: From (2), $0<a|x|^2\leq L(x,k)\leq b|x|^2$ for $x\neq 0$, $L(0,k)=0$ for all k, and $L(x,k)$ is radially unbounded. We next evaluate the rate of change of $L(x,k)$ along the trajectories of (1):

$$\begin{aligned}\Delta_1L(x,k)&=x_{k+1}'P_{k+1}x_{k+1}-x_k'P_kx_k\\&=x_k'[\Phi'(k+1,k)P_{k+1}\Phi(k+1,k)-P_k]x_k\\&=x_k'[F_k'P_{k+1}F_k-P_k]x_k\\&=-x_k'H_kH_k'x_k.\end{aligned} \tag{5}$$

Evidently, $L(\cdot,\cdot)$ is nonincreasing along all trajectories, and hence is a Lyapunov function for (1). Furthermore,

$$\begin{aligned}\Delta_1L(x,k+1)&=x_{k+2}'P_{k+2}x_{k+2}-x_{k+1}'P_{k+1}x_{k+1}\\&=x'_{k+1}[F_{k+1}'P_{k+2}F_{k+1}-P_{k+1}]x_{k+1}\\&=-x_k'\Phi'(k+1,k)H_{k+1}H_{k+1}'\Phi(k+1,k)x_k\\\Delta_1L(x,k+i)&=-x_k'\Phi'(k+i,k)H_{k+i}H_{k+i}'\Phi(k+i,k)x_k.\end{aligned}$$

Summing, we have

$$\begin{aligned}\Delta_nL(x,k)&=x_{k+n-1}'P_{k+n-1}x_{k+n-1}-x_k'P_kx_k\\&=\sum_{i=0}^{n-1}\Delta_1L(x,k+i)\\&=-x_k'W_n(k)x_k\end{aligned}$$

and (4) implies that $\Delta_nL(x,k)\leq -c|x_k|^2$. That is, along all trajectories of (1), the Lyapunov function satisfies

$$\frac{\Delta_nL(x,k)}{L(x,k)}\leq\frac{-c}{b}<0, \tag{6}$$

and hence, $L(x,k)$ is exponentially convergent to zero. Since $L(x,k)$ is quadratic in the state x, it follows that (1) is globally EAS. △△△

This theorem represents a direct extension of the continuous-time result of Anderson and Moore [7]. Once we find a quadratic Lyapunov function, the condition (4) appears as the property which determines the exponential rate. This condition is part of the property of uniform complete observability of $[F_k,H_k]$, and we shall simply refer to this type of requirement, in general, as an observability condition. As we proceed, it will be this observability condition which will play an important role in the development of more powerful results.

As it stands, this theorem applies equally well to deterministic and random matrices F_k. That is, should random F_k and thus random H_k satisfy the lemma requirements almost surely, say, then the convergence is exponential in the deterministic sense along almost all sample paths. The definition of almost sure exponential convergence applicable to more general random matrices F_k, to be given below, is much weaker than this, and it will be clearly necessary to modify the theorem to weaken the requirements on F_k and admit a wider class. This we shall now do.

III. Linear Systems with Random Coefficients

Now we suppose F_k to be random matrices over the probability space $(\Omega,\mathfrak{B},P)$ with elements $\omega\in\Omega$. All expectations, denoted $E(\cdot)$, when used, are assumed to exist. We derive sufficient conditions for the almost sure exponential convergence of the linear difference equation (1), satisfying the following definition.

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

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

We remark here that Kushner [11] has used a more restricted definition of almost sure exponential convergence of a stochastic process which is equivalent to demanding that $\{(1+\beta)^kz_k\}$ be a positive supermartingale.

The definition of exponential convergence of a random process given above differs from uniform exponential convergence of deterministic sequences in that it refers to the average rate of convergence over a long interval rather than to the ability to always bound the norm of the transition matrix $\Phi(t,s)$ by $c_1\exp[-c_2(t-s)]$ for all given times t and s and some positive constrants c_1,c_2. Thus, we argue that the long sample run behavior of the algorithm is the property which determines the exponential convergence. This use of average rate of convergence over a large number of steps, as opposed to the stepwise rate of convergence, is familiar from the study of deterministic optimization algorithms [14, see p. 129]. Note also that while we do not require uniformity of convergence rate over each finite time interval, we do require uniformity of average convergence rate over almost all sample paths $\omega\in\Omega$; this means that it is not an admissible proof of almost sure exponential convergence to show that for each ω, the average convergence rate is exponential unless we also show that the a.s. supremum of these rates is exponential.

Suppose now that we have a random system and let us reexamine the proof of Theorem 1 with an eye towards proving almost sure EAS. We suppose that (2) and (3) are satisfied and we concentrate on the observability condition. In (4), for fixed n and random $F_k(\omega)$, c becomes $c_k(\omega)$, a random variable, and (6) yields (suppressing the x argument)

$$L[(k+1)n] \leq [1-d_k(\omega)]L(kn)$$

where $d_k = c_{kn}/b$. Clearly, the bounds on $L(k)$ ensure that $d_k \in [0,1]$. Taking logarithms

$$\log L[(k+1)n] - \log L(kn) \leq \log[1-d_k](\omega)$$

and summing yields

$$\log L[(k+m)n] - \log L(kn) \leq \sum_{i=0}^{m-1} \log[1-d_{k+i}]. \tag{7}$$

Now suppose that the summation on the right above diverges to $-\infty$ as $m\to\infty$ faster than $m\nu$ for some $\nu>0$, i.e., as $m\to\infty$

$$m\nu + \sum_{i=0}^{m-1} \log[1-d_{k+i}] \to -\infty \tag{8}$$

almost surely. Then assuming $L(kn)$ to be almost surely finite, we have that, combining (7) and (8),

$$\log L[(k+m)n] - \log L[kn] + m\nu \to -\infty$$

almost surely as $m\to\infty$, whence

$$[e^{\nu}]^m L[(k+m)n] \to 0$$

almost surely as $m\to\infty$. That is, the system is almost surely globally exponentially asymptotically stable.

The case where $d_j = 1$ for some j needs to be singled out for special mention here as it may cause problems when using $\log[1-d_j]$ as in (7). In case this happens, it is apparent from the above that $L[(j+k)n] \equiv 0$ for all $k \geq 1$ so that we have convergence in a finite number of steps, and the rate is consequently faster than exponential. (This situation could occur when $F_i = 0$ for some i.) Since $d_j \in [0,1]$, we assign $\log(1-d_j)|_{d_j=1}$ the value $\lim_{d_j\to 1} \inf \log(1-d_j) = -\infty$; the analysis carries through without problems and the sum in (8) diverges for finite m. Indeed, it is evident that the main problem of establishing minimum convergence rates is associated with values of d_j which are arbitrarily close to zero rather than those close to one.

That the convergence proof relies upon the observability condition is evident from requirement (8) because d_k is related to the minimum eigenvalue of the observability Gramian $W_n(k)$. Condition (8) is, in fact, the key condition to establishing almost sure exponential convergence, and following a statement of the random Lyapunov theorem and an invariance rule, we go on to present two cases where (8) can be easily satisfied.

Theorem 2 (Random Lyapunov Theorem for Exponential Stability): Consider the discrete-time, free dynamic system with random coefficients

$$x_{k+1} = F_k(\omega)x_k. \tag{9}$$

Suppose that there exists a symmetric positive definite matrix function $P_k(\omega)$ and positive constants a and b such that for all k

$$0 < aI \leq P_k \leq bI < \infty \qquad \text{a.s.} \tag{10}$$

and such that, for some matrix $H_k(\omega)$,

$$F_k' P_{k+1} F_k - P_k = -H_k H_k'. \tag{11}$$

Now define

$$W_n(kn) = \sum_{i=0}^{n-1} \Phi'(kn+i,kn) H_{kn+i} H_{kn+i}' \Phi(kn+i,kn) \tag{12}$$

with $\Phi(\cdot,\cdot)$ the transition matrix of (9), and define $d_k(\omega)$ as the minimum eigenvalue of $W_n(kn)$ divided by b of (10); observe that $0 \leq d_k \leq 1$.

Then, if there exists a constant $\nu > 0$ such that almost surely

$$m\nu + \sum_{i=0}^{m-1} \log[1-d_{k+i}] \to -\infty \tag{13}$$

as $m\to\infty$, the dynamic system is almost surely globally exponentially asymptotically stable.

From the study so far, we see that the linear difference equation is exponentially stable provided that we can find $P_k(\omega)>0$ and $H_k(\omega)$ to satisfy (10) and (11), and the pair $[F_k, H_k]$ is observable in the sense that $W_n(kn)(\omega)$ has the desired properties as illustrated in Theorem 2. Conditions for this will be obtained in later sections. However, often $W_n(kn)$ is difficult to evaluate because of the need to introduce the transition matrix of the difference equation. We now demonstrate a result that can overcome this difficulty in application to the effect that exponential convergence can be determined by studying the (possibly simpler) observability Gramian of the pair $[F_k + K_k H_k', H_k]$ for uniformly bounded output feedback matrix $K_k(\omega)$. (From (10) and (11), we see that F_k and H_k are bounded as $H_k H_k' \leq bI$ and $F_k' F_k \leq (b/a)I$.)

Theorem 3 (Invariance of Observability under Bounded Output Feedback): Let $W_n(k)$ be the observability Gramian associated with the uniformly bounded pair $[F_k, H_k]$, i.e., with

$$\Phi(k+i,k) = F_{k+i-1}F_{k+i-2}\cdots F_k$$

$$W_n(k) = \sum_{i=0}^{n-1} \Phi'(k+i,k) H_{k+i} H_{k+i}' \Phi(k+i,k)$$

and let $\overline{W}_n(k)$ be the observability Gramian associated with the pair $[F_k + K_k H_k', H_k]$ for uniformly bounded matrices K_k. Then there exist positive constants α and β (dependent on n) such that

$$\alpha \overline{W}_n(k) \leq W_n(k) \leq \beta \overline{W}_n(k).$$

That is, $W_n(k)$ will satisfy an observability condition if and only if $\overline{W}_n(k)$ does also, so that Theorem 2 still provides a sufficient test for exponential stability when requirement (13) on $W_n(kn)$ is replaced by a similar requirement on $\overline{W}_n(kn)$. The indicated rate of convergence is not invariant, however.

Proof: We may write the two observability Gramians as $W_n(k) = LL'$ and $\overline{W}_n(k) = LTT'L'$ where

$$L = \begin{bmatrix} H_k & F_k' H_{k+1} & F_k' F_{k+1}' H_{k+2} \cdots \end{bmatrix}$$

$$T = \begin{bmatrix} I & K_k' H_{k+1} & K_k' F_{k+1}' H_{k+2} + K_k' H_{k+1} K_{k+1}' H_{k+2} & \cdots \\ 0 & I & K_{k+1}' H_{k+2} & \cdots \\ 0 & 0 & I & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

T is upper triangular, nonsingular, and has uniformly bounded entries. Similarly, T^{-1} is upper triangular, nonsingular, and has uniformly bounded entries. Thus, T and T^{-1} both have uniformly bounded eigenvalues so that

$$\alpha I \leq TT' \leq \beta I$$

for some $\alpha, \beta > 0$. The result follows. △△△

The analysis so far can be seen as approaching the question of: given a Lyapunov-stable linear time-varying difference equation, what are the extra conditions necessary to ensure EAS in either the deterministic or random framework? The importance of answering this question lies in the fact that many adaption algorithms are designed for the deterministic problem on the basis of Lyapunov stability [15]. Asymptotic stability, corresponding to convergence of the algorithm to the desired value, is then inferred by nondegeneracy of the coefficients typically expressed using a persistently exciting or another spanning condition [5],[6]. We provide an extension to the present theory which allows EAS to be established in the random case.

Next we consider two important instances where the particular random nature of the coefficients $\{F_k(\omega)\}$ allows us to satisfy (13) under very mild assumptions. This is followed by an example of the method's application to an often used algorithm which yields very weak sufficient conditions for EAS.

IV. Ergodic Coefficients

Suppose that $\{F_k(\omega)\}$ is an ergodic sequence of random matrices. By this we mean that, in the probability space $(\Omega, \mathfrak{B}, P)$, the random variable $F_0(\omega)$ is $\mathfrak{B}$-measurable, and there exists an invertible, measure-preserving transformation T of Ω into Ω with $F_k(\omega)=F_0(T^k\omega)$. The ergodic theorem then ensures that almost surely

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} F_i(\omega) \to E(F_0) \tag{14}$$

provided that this expectation exists. The properties of ergodic transformations of probability spaces are discussed at length by Billingsley [16] and Halmos [17].

We choose ergodicity for the $\{F_k\}$'s rather than just the satisfaction of the law of large numbers (14) because it is easily extended to related random variables. In particular, it is apparent that if P_k is fixed independent of k and ω, then the sequence $\{H_kH_k'(\omega)\}$ is ergodic because H_kH_k' is a random variable measurable with respect to the sigma field of events generated by F_k. Furthermore, the random sequence $\{W_n(kn)\}$ of observability matrices over nonoverlapping intervals of length n is ergodic as is $\{d_k(\omega)\}$, the sequence of random variables proportional to the minimum eigenvalue of these observability matrices, as is any function of d_k. Let $1>\epsilon>0$ and define $\hat{d}_k=(1-\epsilon)d_k$. Evidently, $\hat{d}_k(\omega)\in[0,1-\epsilon]$ so that $\{\log(1-\hat{d}_i)\}$ is ergodic with $E|\log(1-\hat{d}_1)|<\infty$. Hence,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\log(1-d_i) &\le \frac{1}{n}\sum_{i=1}^{n}\log\left(1-\hat{d}_i\right) \quad \text{a.s.} \\ &\to E\{\log[1-(1-\epsilon)d_1]\} \quad \text{a.s. as } n\to\infty \\ &\le \log[1-(1-\epsilon)Ed_1]\end{aligned} \tag{15}$$

by Jensen's inequality [26].

Since $\epsilon>0$ is arbitrary, we see that a.s. as $n\to\infty$,

$$\frac{1}{n}\sum_{i=1}^{n}\log(1-d_i) \le \log[1-Ed_1]$$

and taking any $\nu < -\log[1-Ed_1]$, the requirement for satisfaction of (13) is that $Ed_1>0$. We have the following.

Theorem 4: With $\{F_k\}$ an ergodic sequence of random matrices and with P_k constant, a sufficient condition for the exponential convergence of the system

$$x_{k+1}=F_kx_k$$

is that $E[d_1]>0$ where d_1 is defined as in Theorem 2.

For the moment, we leave in abeyance the next logical question: what further property, additional to ergodicity of $\{F_k\}$, is sufficient to ensure $E[d_1]>0$? To answer this question, we need more precise knowledge of $\{F_k\}$ and $\{H_k\}$, and we will present an example of these additional conditions shortly. Suffice it to say that these extra requirements appear to be quite mild for the cases studied by the authors.

The class of ergodic coefficients includes the class of independent and identically distributed processes, as well as many stationary dependent processes. For a stationary process to be ergodic, the dependence between events widely separated in time should diminish (in a certain weak sense) as the separation increases, see, e.g., the chapter on mixing by Halmos [17]. Consequently, if we know that widely separated events are asymptotically independent, e.g., for every pair of $\mathfrak{B}$-measurable sets A and B,

$$\lim_{m\to\infty} |P(T^mA\cap B)-P(A)P(B)| \to 0, \tag{16}$$

then we have ergodicity. This is called strong mixing [17] or complete regularity of the stationary process [18]. Also, ergodicity is implied by weak mixing where we replace the convergence in (16) by the weaker Césaro convergence,

$$\lim_{m\to\infty} \frac{1}{m}\sum_{i=1}^{m} |P(T^iA\cap B)-P(A)P(B)| \to 0. \tag{17}$$

Ibragimov and Linnik [18] present several conditions for Gaussian sequences to be mixing in various senses, and, hence, to be ergodic.

Finally, we remark that perhaps the greatest restriction associated with the ergodicity assumption is that the $\{F_k\}$ must be stationary. As remarked in the Introduction, one of the reasons we desire exponential convergence of our algorithms is so that in nonstationary environments we can still use the algorithms to estimate and track time-varying parameters. This would appear to be precluded by the stationarity assumption. All is not lost, however, since, although ergodicity is difficult to extend to any other than asymptotically mean stationary sequences, the weaker properties of mixing are widely applicable to nonstationary processes, particularly with regard to proving laws of large numbers.

V. Weakly Dependent and ϕ-Mixing Nonstationary Coefficients

As indicated above, mixing properties like (16) and (17) extend to the study of nonstationary processes. Generally, a mixing property is a form of asymptotic independence of events with increasing time separation, and consequently will apply to independent and M-dependent stochastic processes. However, there are many results using mixing properties which simply require widely separated events to be weakly dependent in a certain sense, and which allow consideration of stochastic processes which are dependent over arbitrarily large time intervals.

The particular mixing property that we consider is ϕ-mixing [19], which will be defined below. This property says that the distant future of the stochastic process is weakly dependent upon the present and past, and it would appear that this is a very reasonable type of condition to demand in random nonstationary engineering or physical situations. One apparent drawback of the use of ϕ-mixing in establishing results is that from the definition below of the mixing constants ϕ_n, it would appear to be all but impossible to determine whether the ϕ-mixing requirements for convergence are satisfied in a particular engineering application.

Our counter to this argument is that the willingness to accept the ϕ-mixing conditions as holding is affected by the familiarity with the concepts. Indeed, it would appear that acceptance of the satisfaction of the ϕ-mixing conditions of Theorem 6 below in an engineering application is less presumptuous that the acceptance of, say, the central limit theorem. The property of ϕ-mixing then appears as a very general notion which is quite justifiable in engineering contexts and which, as will be shown, allows extraordinary flexibility in the handling of nonstationary processes. We now go on to discuss ϕ-mixing in detail.

For the sequence of matrices $\{F_k\}$, $k=1,2,\cdots$, let $\mathfrak{M}_a^b$ for $1\le a\le b$ denote the sigma field of events generated by the random variables $F_a, F_{a+1}, \ldots, F_b$ with the obvious extension for $\mathfrak{M}_a^\infty$. There have been several attempts to define dependence coefficients relating the dependence of events in different sigma fields $\mathfrak{M}_1$ and $\mathfrak{M}_2 \subset \mathfrak{B}$. We only consider the following one due to Ibragimov [20], Iosifescu and Theodorescu [21], and Billingsley [19]:

$$\phi(\mathfrak{M}_1,\mathfrak{M}_2) = \sup_{B\in\mathfrak{M}_2}\left[\operatorname*{ess\,sup}_{\omega\in\Omega} |P(B|\mathfrak{M}_1)(\omega)-P(B)|\right]$$

(where ess sup refers to the essential supremum of a strictly bounded function).

The interpretation of these numbers as dependence coefficients arises because, if the $\{F_k\}$'s are independent, then $\phi(\mathfrak{M}_a^k, \mathfrak{M}_{k+n}^b)=0$ for $n\ge 1$, while if they are highly dependent, the coefficients will be close to one for a large number of values of n. We say that the sequence $\{F_k\}$ is ϕ-mixing if there exists a sequence of numbers ϕ_n such that, for all $k\ge 1$,

$$\phi(\mathfrak{M}_1^k, \mathfrak{M}_{k+n}^\infty) \le \phi_n$$

and we shall be concerned with the properties of nonstationary ϕ-mixing sequences with nontrivial $\{\phi_n\}$. (An example of trivial ϕ_n is $\phi_n=1$ for all n, which does not allow anything to be proved.)

In general, the most interesting cases of ϕ-mixing sequences are when we have $\phi_n\to 0$ as $n\to\infty$ (corresponding to asymptotic independence of the distant future on the present), and more particularly when $\phi_n\to 0$ at some prescribed rate. For these sequences of dependent random variables, it is often possible to establish properties usually associated with sequences of independent random variables; see particularly [21]. We shall appeal chiefly to the following result of [21] attributed to Cohn, which is a strong law of large numbers.

Theorem 5 [21]: Let the random variables f_i be $\mathfrak{M}_i^i$-measurable, $i = 1,2,\cdots$ and ϕ-mixing with mixing constants $\{\phi_n\}$.

If

$$\sum_{n=1}^{\infty} \phi_n^{1/2} < \infty \quad \text{and} \quad \sum_{n=1}^{\infty} \frac{\text{var}(f_n)}{n^2} < \infty,$$

then as $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^{n} f_i \to \frac{1}{n}\sum_{i=1}^{n} E(f_i)$$

almost surely.

Returning to the case at hand of the exponential convergence of the discrete-time dynamic system with random coefficients, we notice that if the sequence $\{F_k(\omega)\}$ is ϕ-mixing and $\{P_k\}$ is constant, then the sequence $\{H_k H_k'(\omega)\}$ is also ϕ-mixing. This follows because $H_k H_k'$ is $\mathfrak{M}_k^k$-measurable and ϕ-mixing is a property concerned with the sequence of σ-fields $\{\mathfrak{M}_i^j\}$ rather than just with a particular sequence of random variables.

By a chain of inference similar to the ergodic case, we can see that the sequence of scalars $\{d_k(\omega)\}$ is ϕ-mixing, as is any sequence of a function of d_k. Now suppose that for the $\{\phi_n\}$'s associated with $\{F_k\}$, we have $\sum_{n=1}^{\infty}\phi_n^{1/2} < \infty$;[1] then for $1 > \epsilon > 0$, defining $\hat{d}_k = (1-\epsilon)d_k$ and noting that $\text{var}[\log(1-\hat{d}_k)]$ is bounded,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\log(1-d_i) &\le \frac{1}{n}\sum_{i=1}^{n}\log\left(1-\hat{d}_i\right) \quad \text{a.s.}\\
&\to \frac{1}{n}\sum_{i=1}^{n} E\left\{\log\left(1-\hat{d}_i\right)\right\} \quad \text{a.s. as } n\to\infty\\
&\le \frac{1}{n}\sum_{i=1}^{n}\log\left[1-E\left(\hat{d}_i\right)\right] \quad \text{by Jensen's inequality}\\
&\le \frac{-1}{n}\sum_{i=1}^{n} E\left(\hat{d}_i\right) \quad \text{using } \log x \le x-1 \text{ for } x>0,\\
&= \frac{-(1-\epsilon)}{n}\sum_{i=1}^{n} E(d_i).
\end{aligned}$$

Again paralleling the ergodic case, condition (13) of Theorem 2 is satisfied if $\lim_{n\to\infty}\inf(1/n)\sum_{i=1}^{n}E(d_i) > 0$. This yields the following.

Theorem 6: With $\{F_k\}$ a nonstationary ϕ-mixing sequence of random matrices, with $\sum_{i=1}^{\infty}\phi_i^{1/2} < \infty$, a sufficient condition for the exponential convergence of the system

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

is that $\lim_{n\to\infty}\inf(1/n)\sum_{i=1}^{n}E(d_i) > 0$ where d_i is defined as in Theorem 2.

The comparison between this result and Theorem 4 is apparent, and it shows that if we allow a slight weakening of ergodicity to ϕ-mixing, with a guaranteed rate of decrease of ϕ_n, then we robustify the exponential convergence result to include certain nonstationary systems.

We remark here that there are conditions, other than ϕ-mixing or ergodicity, on processes which allow the satisfaction of the strong law of large numbers as in Theorem 5; see, e.g., Révész [22] or Cramer and Leadbetter [23, p. 94] for particularly interesting examples. The principal advantage of ϕ-mixing and ergodicity is that it is hereditary in the sense that a function of a ϕ-mixing process is itself ϕ-mixing. The covariance decay-rate condition of [23] is more difficult to carry over, although under the assumption of ϕ-mixing, it can readily be verified [21],[19].

[1] Once again, we remark that the problem is a pedagogical one here with the summability condition needing to be believed rather than verified. We refer to the earlier discussion where it was argued that the assumption of ϕ-mixing with this summability is entirely justifiable for engineering processes and often is considerably weaker than assuming more familiar conditions such as ergodicity, whiteness, or that the central limit theorem holds.

VI. An Example in Adaptive Filtering

Here we consider the EAS of the error equation arising in an adaptive estimation algorithm proposed by Albert and Gardner [24] and others [25] as a "quick and dirty regression scheme" or adaptive filter algorithm. The homogeneous algorithm arises when we have a measurable random input sequence of N-vectors $\{u_k\}$ producing a measurable random output sequence of scalars $\{y_k\}$ according to

$$y_k = u_k' w^* \tag{18}$$

and, using measurements of $\{y_k, u_k\}$, we recursively estimate w^* by

$$\begin{aligned} w_{k+1} &= w_k + \mu\frac{u_k}{u_k'u_k}(y_k - \hat{y}_k) \\ \hat{y}_k &= u_k' w_k \end{aligned} \tag{19}$$

where μ is a positive constant. In application, (18) is never exact, and the estimation scheme (19) is performing a linear regression or finding a transversal filter model of the relationship between u_k and y_k. Also, in practice it is often necessary to include a small adaptive constant in the denominator of (19) to avoid problems where $u_k = 0$—this modified algorithm is examined in [27].

Writing $v_k = w_k - w^*$, the parameter error, and substituting from (18),(19) becomes

$$v_{k+1} = \left(I - \mu\frac{u_k u_k'}{u_k' u_k}\right)v_k. \tag{20}$$

We propose the candidate Lyapunov function

$$L(v,k) = v'v = L(v)$$

which is of the general form of the quadratic Lyapunov function proposed earlier for the stability analysis of linear equations with $P_k = I$. We next consider $\Delta_n L(v_0) = L(v_n) - L(v_0)$ by studying the equation

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

with

$$F_k = \left(I - \mu\frac{u_k u_k'}{u_k' u_k}\right) \quad \text{and} \quad P_k = I.$$

Clearly,

$$H_k H_k' = [1-(1-\mu)^2]\frac{u_k u_k'}{u_k' u_k}$$

and we may choose

$$H_k = [\mu(2-\mu)]^{1/2}|u_k|^{-1}u_k.$$

Here we see that μ must belong to the interval $(0,2)$.

From here, we could proceed directly as indicated by Theorem 2 and evaluate the observability Gramian of $[F_k, H_k]$. Things become cumbersome, however, when we calculate the transition matrix. Therefore, we invoke Theorem 3 and consider the observability Gramian of $[F_k + K_k H_k', H_k]$ with

$$K_k = \mu[\mu(2-\mu)]^{-1/2}|u_k|^{-1}u_k.$$

F_k, K_k, and H_k are now uniformly bounded and

$$F_k + K_k H_k' = I$$

so that the transition matrix is also I and

$$\overline{W}_k(kn)=\sum_{i=0}^{n-1} H_{kn+i}H'_{kn+i}$$
$$=\mu(2-\mu)\sum_{i=kn}^{kn+n-1}\frac{u_iu_i'}{u_i'u_i}.$$

Returning to Theorem 2 and 3, our next task is to define

$$\bar{d}_k=\lambda_{\min}\left(\sum_{i=kn}^{kn+n-1}\frac{u_iu_i'}{u_i'u_i}\right)\mu(2-\mu)$$

and to consider what it implies about $d_k=\lambda_{\min}[W_n(kn)]$ satisfying (13).

If $\{u_k\}$, and hence $\{F_k\}$, is ergodic, then Theorem 4 gives $E(d_1)>0$ as a sufficient condition for exponential stability, and from Theorem 3, this holds if $E(\bar{d}_1)>0$.

Theorem 7: With ergodic inputs $\{u_k\}$, the algorithm (20) is almost surely exponentially convergent for $\mu\in(0,2)$ if there exists a finite integer n such that

$$E\left\{\lambda_{\min}\left(\sum_{i=0}^{n-1}\frac{u_iu_i'}{u_i'u_i}\right)\right\}>0. \tag{21}$$

Furthermore, this condition holds provided $\{u_k\}$ has full rank covariance and its fourth moment exists.

Proof: The former claim is established with the argument preceding the theorem statement. The latter claim follows from the ergodicity—if $\{u_k\}$ is ergodic, then the summation (21) behaves like n times the expected value as n gets large. Thus, if $E(u_iu_i'/u_i'u_i)$ is full rank, then for some n, we satisfy (21).

To prove the second claim, we show that the existence of $x\neq 0$ such that $x'E(u_iu_i'/u_i'u_i)x=0$ necessitates $x'E(u_iu_i')x=0$.

$$x'E(u_iu_i')x=E\left[(x'u_i)^2\right]$$
$$=E\left[\frac{(x'u_i)}{|u_i|}\cdot(x'u_i)|u_i|\right].$$

Now use the Cauchy–Schwarz inequality

$$x'E(u_iu_i')x<E\left[\frac{(x'u_i)^2}{|u_i|^2}\right]^{1/2}E\left[(x'u_i)^2|u_i|^2\right]^{1/2}$$
$$=\left[x'E\left(\frac{u_iu_i'}{u_i'u_i}\right)x\right]^{1/2}\left[x'E(u_iu_i'u_iu_i')x\right]^{1/2}$$

so that, provided $E[u_iu_i'u_iu_i']$ exists, the result follows. △△△

Notice the double role of ergodicity in this proof. It is used to prove the convergence of the sum of logarithms and also to prove that $\bar{d}_1$ behaves correctly.

We may easily extend Theorem 7 to include ϕ-mixing inputs by drawing upon Theorem 6.

Theorem 8: With nonstationary ϕ-mixing inputs $\{u_k\}$, with $\sum_{i=1}^{\infty}\phi_i^{1/2}<\infty$, the algorithm (20) with $\mu\in(0,2)$ is almost surely exponentially convergent if for some finite integer n

$$\lim_{m\to\infty}\frac{1}{m}\sum_{i=0}^{m-1}E\left\{\lambda_{\min}\left(\sum_{j=in}^{in+n-1}\frac{u_iu_i'}{u_i'u_i}\right)\right\}>0. \tag{22}$$

Furthermore, this holds provided $\lim_{m\to\infty}(1/m)\sum_{i=0}^{m-1}E(u_iu_i')$ has full rank and fourth moments exist.

The proof of this theorem is analogous to that of Theorem 7.

These two theorems provide very strong convergence properties for this algorithm. They greatly extend the analysis of Nagumo and Noda [25] who considered only independent and identically distributed $\{u_k\}$'s. However, one drawback of our approach using Theorem 3 is that our estimate of convergence rate is difficult—one needs to know the constants relating the observability Gramians under feedback as well as eigenvalue information. Despite this, however, we have derived an estimate of convergence rate for this algorithm in [8] using an approach which does not appeal to Theorem 3, but which is consequently much more difficult.

VII. Conclusion

We have presented the problem of determining the exponential asymptotic stability (EAS) of linear homogeneous difference equations with random coefficients. The problem is motivated by the study of adaptive estimation algorithms.

We provide an approach to the solution of the problem via Lyapunov stability theory, and produce strong results for the EAS of deterministic linear difference equations and of such equations with stationary ergodic or nonstationary ϕ-mixing random coefficients. While the major theorems appear quite complicated, it is shown by the example of an adaptive filtering algorithm that the sufficient conditions for EAS may be easily satisfied in many common algorithms.

In a companion work [27], we more fully motivate the EAS for homogeneous adaptive algorithms and then present a detailed use of the results given here to establish the EAS, and hence good in-use performance for other important estimation schemes.

References

[1] B. D. O. Anderson, "An approach of multivariable system identification," *Automatica*, vol. 13, pp. 401–408, 1977.
[2] K. S. Narendra, "Stable identification schemes," in *System Identification: Advances and Case Studies*, R. K. Mehra and D. G. Lainiotis, Eds. New York: Academic, 1976.
[3] 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.
[4] P. C. Parks, "Liapunov redesign of model reference adaptive control system," *IEEE Trans. Automat. Contr.*, vol. AC-11, pp. 362–367, July 1966.
[5] 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.
[6] 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.
[7] B. D. O. Anderson and J. B. Moore, "New results in linear system stability," *SIAM J. Contr.*, vol. 7, pp. 398–414, Aug. 1969.
[8] 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.
[9] 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.
[10] A. Gersho, "Adaptive equalization of highly dispersive channels for data transmission," *Bell Syst. Tech. J.*, vol. 48, pp. 55–70, Jan. 1969.
[11] H. J. Kushner, *Introduction to Stochastic Control*. New York: Holt, Rinehart and Winston, 1971.
[12] 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.
[13] J. M. Mendel, *Discrete Techniques of Parameter Estimation: The Equation Error Formulation*. New York: Marcel Dekker, 1973.
[14] D. G. Luenberger, *Introduction to Linear and Nonlinear Programming*. Reading MA: Addison-Wesley, 1973.
[15] I. D. Landau, "A survey of model reference adaptive techniques—Theory and applications," *Automatica*, vol. 10, pp. 353–379, 1974.
[16] P. Billingsley, *Ergodic Theory and Information*. New York: Wiley, 1965.
[17] P. R. Halmos, *Lectures on Ergodic Theory*. New York: Chelsea, 1956.
[18] I. A. Ibragimov and Yu. V. Linnik, *Independent and Stationary Sequences of Random Variables*. Groningen: Wolters-Noordhoff, 1971.
[19] P. Billingsley, *Convergence of Probability Measures*. New York: Wiley, 1968.
[20] I. A. Ibragimov, "Some limit theorems for stochastic processes stationary in the strict sense," *Dokl. Akad. Nauk SSSR*, vol. 125, pp. 711–714, 1959.
[21] M. Iosifescu and R. Theodorescu, *Random Processes and Learning*. New York: Springer-Verlag, 1969.
[22] P. Révész, *The Laws of Large Numbers*. New York: Academic, 1967.
[23] H. Cramér and M. R. Leadbetter, *Stationary and Related Stochastic Processes*. New York: Wiley, 1967.
[24] A. E. Albert and L. A. Gardner, Jr., *Stochastic Approximation and Nonlinear Regression*. Cambridge, MA: M.I.T. Press, 1967.
[25] J.-I. Nagumo and A. Noda, "A learning method for system identification," *IEEE Trans. Automat. Contr.*, vol. AC-12, pp. 282–287, June 1967.
[26] R. B. Ash, *Real Analysis and Probability*. New York: Academic, 1972.
[27] R. R. Bitmead and B. D. O. Anderson, "Performance of adaptive estimation algorithms in dependent random environments," this issue, pp. 788–794.