seen that these modifications do not change the results of Theorem 2.4, provided that the condition on the update law is satisfied for all L_i and γ_i.

The class of systems considered is fairly general, and a closer examination of the results reveals that what is essential is that when we apply an input to the system, we can observe a corresponding output, and we act upon this output with the learning operator. The stability of the system may affect the convergence rate, but not the actual convergence of the learning algorithm.

It is important to remember that learning control is not a form of dynamic feedback. It cannot be used to stabilize a system nor to change its performance for a general trajectory. Therefore, in applications it is desirable to use a robust feedback controller to improve the system performance, and as explained earlier, this is the motivation for considering time-varying systems. Learning control iteratively updates a feed-forward term to provide a finer and finer "open loop" performance along a specific trajectory—it is not intended to make up for a poor feedback controller design.

In conclusion, we believe that the learning algorithm presented is applicable to a wide variety of problems. The stability of learning in the presence of disturbances and initial condition errors allows us to use the learning algorithm with confidence in applications. We further conjecture that these results can be extended to other update laws, allowing the differentiation to be replaced by say a lead filter with better noise response; this constitutes an interesting area for future research.

ACKNOWLEDGMENT

The authors would like to express their thanks to R. Minnichelli, R. Murray, and P. Jacobs for their help in the form of discussions and suggestions.

REFERENCES

[1] G. Heinzinger, D. Fenwick, B. Paden, and F. Miyazaki, "Robust learning control," in *Proc. 28th IEEE Conf. Decision Contr.*, Tampa, FL, Dec. 1989.
[2] S. Arimoto, S. Kawamura, and F. Miyazaki, "Bettering operation of robots by learning," *J. Robot. Syst.*, vol. 1, no 2, pp. 440-447, 1984.
[3] J. J. Craig, "Adaptive control of manipulators through repeated trials," in *Proc. Amer. Contr. Conf.*, San Diego, CA, June 1984, pp. 1566-1573.
[4] P. Bondi, G. Casalino, and L. Gambardella, "On the iterative learning control theory for robotic manipulators," *IEEE J. Robot. Automation*, vol. 4, pp. 14-22, Feb. 1988.
[5] C. G. Atkeson and J. McIntyre, "Robot trajectory learning through practice," in *Proc. IEEE Conf. Robot. Automation*, San Francisco, CA, Apr. 1986, pp. 1737-1742.
[6] J. Hauser, "Learning control for a class of nonlinear systems," in *Proc. 26th IEEE Conf. Decision Contr.*, Los Angeles, CA, Dec. 1987, pp. 859-860.
[7] T. Sugie and T. Ono, "On an iterative learning control law for dynamical systems," in *Proc. IFAC 10th World Conf.*, Munich, Germany, June 1987, pp. 339-344.
[8] S. Arimoto, S. Kawamura, and F. Miyazaki, "Convergence, stability, and robustness of learning control schemes for robot manipulators," in *Proc. Int. Symp. Robot Manipulators; Modeling, Contr., Education*, Albuquerque, NM, Nov. 1986.
[9] ——, "Mathematical theory of learning with applications to robot control," in *Adaptive and Learning Systems*, K. S. Narendra, Ed., New York: Plenum 1986.
[10] M. Togai and O. Yamano, "Learning control and its optimality: Analysis and its application to controlling industrial robots," in *Proc. 24th IEEE Conf. Decision Cont.* Fort Lauderdale, FL, Dec. 1985, pp. 248-253.
[11] T. M. Flett, *Differential Analysis*. Cambridge: Cambridge University Press, 1980.

Discrete-Time Filtering for Linear Systems with Non-Gaussian Initial Conditions: Asymptotic Behavior of the Difference Between the MMSE and LMSE Estimates

Richard B. Sowers and Armand M. Makowski

Abstract—**We consider the one-step prediction problem for discrete-time linear systems in correlated plant and observation Gaussian white noises, with *non*-Gaussian initial conditions. We investigate the large time asymptotics of ϵ_t, the expected squared difference between the MMSE and LMSE (or Kalman) estimates of the state at time t given past observations. We characterize the *limit* of the error sequence $\{\epsilon_t, t = 0, 1, \cdots\}$ and obtain some related *rates* of convergence; a complete analysis is provided for the scalar case. The discussion is based on explicit representations for the MMSE and LMSE estimates, recently obtained by the authors, which display the dependence of these quantities on the initial distribution.**

I. INTRODUCTION

Consider the time-invariant linear discrete-time stochastic system

$$X_0^o = \xi, \; X_{t+1}^o = AX_t^o + W_{t+1}^o$$
$$Y_t = HX_t^o + V_{t+1}^o \qquad t = 0, 1, \cdots, \tag{1.1}$$

where the matrices A and H are of dimension $n \times n$ and $n \times k$, respectively. This system is defined on some underlying probability triple $(\Omega, \mathcal{F}, P)$ which carries all the random elements considered in this note, namely the $\mathbb{R}^n$-valued plant process $\{X_t^o, t = 0, 1, \cdots, \}$, the $\mathbb{R}^k$-valued observation process $\{Y_t, t = 0, 1, \cdots\}$ and the $\mathbb{R}^{n+k}$-valued noise process $\{(W_{t+1}^o, V_{t+1}^o), t = 0, 1, \cdots\}$. Throughout this note we make assumptions A.1-A.3, where:

A.1: the process $\{(W_{t+1}^o, V_{t+1}^o), t = 0, 1, \cdots\}$ is a stationary zero-mean $\mathbb{R}^{n+k}$-valued Gaussian white noise sequence [2, p. 22] with covariance structure Γ given by

$$\Gamma := \text{Cov}\begin{pmatrix} W_{t+1}^o \\ V_{t+1}^o \end{pmatrix} = \begin{pmatrix} \Gamma_w & \Gamma_{wv} \\ \Gamma_{vw} & \Gamma_v \end{pmatrix} \qquad t = 0, 1, \cdots; \tag{1.2}$$

A.2: the initial state ξ has distribution F with finite first and second moments μ and Δ, respectively, and is independent of the noise process $\{(W_{t+1}^o, V_{t+1}^o), t = 0, 1, \cdots\}$; and

A.3: the covariance matrices Γ_v and Δ are positive definite, thus invertible.

For each $t = 0, 1, \cdots$, we form the conditional mean $\hat{X}_{t+1} := E[X_{t+1}^o \mid Y_0, Y_1, \cdots, Y_t]$ or MMSE estimate of X_{t+1}^o on the basis of $\{Y_0, Y_1, \cdots, Y_t\}$. In general, $\hat{X}_{t+1}$ is a nonlinear function of $\{Y_0, Y_1, \cdots, Y_t\}$, in contrast to the corresponding LMSE or Kalman estimate of X_{t+1}^o which is by definition linear, and which we denote by $\hat{X}_{t+1}^K$. We then calculate $\epsilon_{t+1} := E[\| \hat{X}_{t+1} - \hat{X}_{t+1}^K \|^2]$[1]

Manuscript received March 30, 1989; revised July 9, 1990 and August 3, 1990. Paper recommended by Past Associate Editor, D. H. Owens. The work of R. B. Sowers was supported by an ONR Graduate Fellowship. The work of A. M. Makowski was supported in part by the Engineering Research Centers Programs NSF CDR 88-03012 and in part by the National Science Foundation under Grant ECS 83-51836.

R. B. Sowers was with the Department of Electrical Engineering and the Systems Research Center, University of Maryland, College Park, MD 20742. He is now with the Center for Applied Mathematical Sciences, Department of Mathematics, University of Southern California, Los Angeles, CA 90089-1113.

A. M. Makowski is with the Department of Electrical Engineering and the Systems Research Center, University of Maryland, College Park, MD 20742.

IEEE Log Number 9103082.

[1] Here $\|v\|$ denotes the Euclidean norm of the vector v in $\mathbb{R}^n$.

0018-9286/92$03.00 © 1992 IEEE

which is an L^2-measure of the agreement between the MMSE and LMSE estimates of X_{t+1}^o on the basis of $\{Y_0, Y_1, \cdots, Y_t\}$.

The goal of this note is to study the asymptotic behavior of ϵ_t as the time parameter t tends to infinity. Noting the dependence

$$\epsilon_t = \epsilon_t((A, H, \Gamma), F), \qquad t = 1, 2, \cdots \tag{1.3}$$

we find it natural to parametrize our asymptotic analysis of ϵ_t in terms of the *system triple* (A, H, Γ) and of the initial distribution F. Of course, if F is Gaussian, the LMSE and MMSE estimates coincide and $\epsilon_t = 0$ for all $t = 1, 2, \cdots$, and any system triple (A, H, Γ).

We are interested in characterizing the *limit* of the error sequence $\{\epsilon_t, t = 0, 1, \cdots\}$ and in obtaining the corresponding *rate* of convergence (or bounds on it). In particular, we seek conditions under which the convergence $\lim_t \epsilon_t = 0$ takes place, and investigate the form of the corresponding rate of convergence and its dependence on the initial distribution F. Of special interest is the situation where exponential rates of convergence are available, i.e., $\lim_t \frac{1}{t} \log \epsilon_t = -I$ for some $I > 0$. Our most useful result along these lines is Theorem 5 below which is an immediate consequence of Theorem 4 and Proposition 3 discussed in Section III. To state this result, we need the $n \times n$ matrices $\bar{A}$ and $\bar{C}$ which are defined by

$$\bar{A} := A - \Gamma_{wv}\Gamma_v^{-1}H \quad \text{and} \quad \bar{C} := \Gamma_w - \Gamma_{wv}\Gamma_v^{-1}\Gamma_{vw}. \tag{1.4}$$

Since $\bar{C}$ is the common covariance matrix of the estimation errors $\{W_{t+1}^o - E[W_{t+1}^o \mid V_{t+1}^o], t = 0, 1, \cdots\}$, it is symmetric positive semidefinite, and its square root is thus well defined [2, Prop. D.1.3., p. 371]; let $\bar{C}^{1/2}$ denote any such square root of $\bar{C}$.

Theorem 5: Assume the pair (A, H) to be detectable and the pair $(\bar{A}, \bar{C}^{1/2})$ to be stabilizable. For any square-integrable distribution F on $\mathbb{R}^n$, we have

$$\lim_t \epsilon_t((A, H, \Gamma), F) = 0 \tag{1.5}$$

and

$$\overline{\lim}_t \frac{1}{t} \log \epsilon_t((A, H, \Gamma), F) \le 2 \log \rho(K_\infty) < 0 \tag{1.6}$$

where K_∞ is an asymptotically stable $n \times n$ matrix [given by (3.5)] and $\rho(K_\infty)$ denotes its spectral radius.

To the best of the authors' knowledge, no results have been reported in the literature on the large time asymptotics of ϵ_t for a general non-Gaussian initial distribution. Such a lack of results may be explained in part by the fact that the key representation result (Theorem 1) has been derived only relatively recently [4]–[7]. In any case, the work reported here provides a formal justification for the idea widely held by practitioners that short of first and second moment information, precise distributional assumptions of the initial condition can be dispensed with when estimating the state X_{t+1}^o on the basis of the observations $\{Y_0, Y_1, \cdots, Y_t\}$. This is a useful complement to Kalman filtering theory since in many applications, the initial distribution is a rather vaguely defined object.

The organization of this note is as follows. In Section II we summarize a representation result for $\{\epsilon_t, t = 0, 1, \cdots\}$ which constitutes the basis for the analysis presented here. In Section III, we investigate the asymptotic behavior of $\{\epsilon_t, t = 0, 1, \cdots\}$ for a general multivariable system. Section IV is devoted to the derivation of a key technical result which is used in the discussion of Section III. This is followed in Section V by a more complete analysis of the scalar case (i.e., $n = k = 1$).

The following notation is used throughout. Elements of $\mathbb{R}^n$ are viewed as column vectors and transposition is denoted by $'$, so that $\|v\|^2 = v'v$ for every v in $\mathbb{R}^n$. For any positive integer n, we denote by $\mathcal{M}_n$ the space of $n \times n$ real matrices and by $\mathcal{Q}_n$ the cone of $n \times n$ positive semidefinite matrices. Moreover, let I_n and O_n be the unit and zero elements in $\mathcal{M}_n$, respectively. For any matrix K in $\mathcal{M}_n$, with sp (K) denoting the set of all eigenvalues of K, we set $\lambda_{\min}(K) := \min\{|\lambda| : \lambda \in \text{sp}(K)\}$ and $\lambda_{\max}(K) := \max\{|\lambda| : \lambda \in \text{sp}(K)\}$; it is customary to call $\lambda_{\max}(K)$ the spectral radius of K and denote it by $\rho(K)$. The mapping $\mathcal{M}_n \to \mathbb{R}_+$ given by

$$\|K\|_{op} := \sup_{v \neq 0} \frac{\|Kv\|}{\|v\|}, \qquad K \in \mathcal{M}_n \tag{1.7}$$

defines the norm on $\mathcal{M}_n$ induced by the Euclidean norm on $\mathbb{R}^n$. However, since all norms are equivalent on $\mathcal{M}_n$, all limiting operations involving matrices can be safely understood entrywise.

We denote by $\mathcal{E}_n$ the set of all square-integrable probability distribution functions on $\mathbb{R}^n$ with positive definite covariance matrix, and by $\mathcal{D}_n$ the set of those distributions in $\mathcal{E}_n$ which have zero mean. For each matrix R in $\mathcal{Q}_n$, let G_R denote the distribution of a zero-mean $\mathbb{R}^n$-valued Gaussian random variable with covariance R.

II. A Representation Result

The basis for our analysis is a representation result for the sequence $\{\epsilon_t, t = 0, 1, \cdots\}$ obtained in [5], [6]. However, before stating this result, we find it useful to observe that there is no loss of generality in assuming $E[\xi] = 0$ or equivalently, in restricting attention to distributions F in $\mathcal{D}_n$. Indeed, a simple translation argument [5, Sect. VI.1] shows that for any square-integrable distribution F in $\mathcal{E}_n$ with mean μ, the relation

$$\epsilon_t((A, H, \Gamma), F) = \epsilon_t((A, H, \Gamma), \tilde{F}) \qquad t = 0, 1, \cdots, \tag{2.1}$$

holds for any triple (A, H, Γ) where $\tilde{F}$ is the element of $\mathcal{D}_n$ given by $\tilde{F}(x) := F(x - \mu)$ for each x in $\mathbb{R}^n$.

We now can state the needed representation result, the proof of which is found in [5], [6].

Theorem 1: Define the $\mathcal{Q}_n$-valued sequence $\{P_t, t = 0, 1, \cdots\}$ by the recursions

$$P_0 = O_n, \; P_{t+1} = AP_tA' - [AP_tH' + \Gamma_{wv}][HP_tH' + \Gamma_v]^{-1} \cdot [AP_tH' + \Gamma_{wv}]' + \Gamma_w \qquad t = 0, 1, \cdots. \tag{2.2}$$

Moreover, let the deterministic sequences $\{Q_t^*, t = 0, 1, \cdots\}$ and $\{R_t^*, t = 0, 1, \cdots\}$ in $\mathcal{M}_n$ and $\mathcal{Q}_n$, respectively, be defined recursively by

$$Q_0^* = I_n, \; Q_{t+1}^* = [A - [AP_tH' + \Gamma_{wv}] \cdot [HP_tH' + \Gamma_v]^{-1}H]Q_t^* \qquad t = 0, 1, \cdots, \tag{2.3}$$

and

$$R_0^* = O_n, \; R_{t+1}^* = R_t^* + Q_t^{*\prime}H'[HP_tH' + \Gamma_v]^{-1}HQ_t^* \qquad t = 0, 1, \cdots. \tag{2.4}$$

For any distribution F in $\mathcal{D}_n$, the representation

$$\epsilon_{t+1} = \int_{\mathbb{R}^n} \frac{\left\| Q_{t+1}^* \int_{\mathbb{R}^n} \{z - [R_{t+1}^* + \Delta^{-1}]^{-1} b\} \exp\left[z'b - \frac{1}{2} z' R_{t+1}^* z\right] dF(z) \right\|^2}{\int_{\mathbb{R}^n} \exp\left[z'b - \frac{1}{2} z' R_{t+1}^* z\right] dF(z)} \, dG_{R_{t+1}^*}(b) \tag{2.5}$$

holds true for each $t = 0, 1, \cdots$.

In order to rewrite (2.5) in a more manageable form, we associate with each distribution F in $\mathscr{D}_n$, the mapping $I_F: \mathscr{M}_n \times \mathscr{Q}_n \to \mathbb{R}$ defined by

$$I_F(K, R) := \int_{\mathbb{R}^n} \frac{\left\| K \int_{\mathbb{R}^n} \{z - [R + \Delta^{-1}]^{-1} b\} \exp\left[z'b - \frac{1}{2} z' R z\right] dF(z) \right\|^2}{\int_{\mathbb{R}^n} \exp\left[z'b - \frac{1}{2} z' R z\right] dF(z)} dG_R(b) \tag{2.6}$$

for all K in $\mathscr{M}_n$ and R in $\mathscr{Q}_n$. We show in Proposition 1 below that (2.6) is always well defined and finite owing to the finite second assumption A.2 on ξ.

With this notation, (2.5) may be rewritten as

$$\epsilon_t = I_F(Q_t^*, R_t^*) \qquad t = 1, 2, \cdots. \tag{2.7}$$

This representation clearly separates the dependence of ϵ_t on the system triple (A, H, Γ) from the dependence on the initial distribution F; the distribution F affects ϵ_t only through the *structure* of the functional I_F, whereas the system triple and time affect ϵ_t only through Q_t^* and R_t^*.

We conclude this section by showing that (2.6) is indeed well defined and finite. For ease of exposition, we set

$$\phi(z, b; R) := \exp\left[z'b - \tfrac{1}{2} z' R z\right] \tag{2.8a}$$

and

$$\Phi(b; R) := \int_{\mathbb{R}^n} \phi(z, b; R)\, dF(z) \tag{2.8b}$$

for all b, z in $\mathbb{R}^n$ and all R in $\mathscr{Q}_n$.

Proposition 1: Let F be a distribution in $\mathscr{D}_n$. For all K in $\mathscr{M}_n$ and R in $\mathscr{Q}_n$, the quantity $I_F(K, R)$ is well defined and finite, with alternate representation

$$I_F(K, R) = \int_{\mathbb{R}^n} \left\| K \int_{\mathbb{R}^n} \{z - [R + \Delta^{-1}]^{-1} b\} \cdot \frac{\phi(z, b; R)}{\Phi(b; R)} dF(z) \right\|^2 \Phi(b; R)\, dG_R(b) < \infty. \tag{2.9}$$

Proof: Fix K in $\mathscr{M}_n$ and R in $\mathscr{Q}_n$. Observe that whenever b lies in the range $\operatorname{Im}(R)$ of R, the quadratic form in the exponent of ϕ in (2.8) is amenable to a completion of squares, namely

$$z'b - \tfrac{1}{2} z' R z = \tfrac{1}{2} b' R^{\#} b - \frac{1}{2}(z - R^{\#} b)' R (z - R^{\#} b), \qquad z \in \mathbb{R}^n,\ b \in \operatorname{Im}(R) \tag{2.10}$$

where $R^{\#}$ denotes the Moore–Penrose pseudoinverse of R [1, pp. 329–330]. Consequently, $\Phi(b; R)$ is finite for each b in $\operatorname{Im}(R)$ since $\phi(z, b; R) < \exp[\frac{1}{2} b' R^{\#} b]$ for b in $\operatorname{Im}(R)$ and z in $\mathbb{R}^n$. This bound and the finite second moment assumption A.2 on ξ together imply that the inner integral in (2.6) is well defined and finite for each b in $\operatorname{Im}(R)$. Therefore, since the support of the Gaussian distribution G_R is exactly $\operatorname{Im}(R)$ and since $\Phi(b; R) < \infty$ for b in $\operatorname{Im}(R)$, we conclude that $I_F(K, R)$ is indeed well defined and that the representation (2.9) holds.

To show that $I_F(K, R)$ is finite, we first observe from Jensen's inequality that

$$\left\| K \int_{\mathbb{R}^n} \{z - [R + \Delta^{-1}]^{-1} b\} \frac{\phi(z, b; R)}{\Phi(b; R)} dF(z) \right\|^2 \le \|K\|_{op}^2 \int_{\mathbb{R}^n} \|z - [R + \Delta^{-1}]^{-1} b\|^2 \frac{\phi(z, b; R)}{\Phi(b R)} dF(z), \qquad b \in \operatorname{Im}(R). \tag{2.11}$$

Next, we set

$$J_F(R) = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \|z - [R + \Delta^{-1}]^{-1} b\|^2 \cdot \phi(z, b; R)\, dF(z)\, dG_R(b)$$
$$= \int_{\mathbb{R}^n} \left[\int_{\mathbb{R}^n} \|z - [R + \Delta^{-1}]^{-1} b\|^2 \cdot \exp[z'b]\, dG_R(b) \right] \exp\left[-\tfrac{1}{2} z' R z\right] dF(z) \tag{2.12}$$

where the last equality follows from Tonelli's theorem. It is now plain from (2.9) and (2.11) that

$$I_F(K, R) \le \|K\|_{op}^2 J_F(R). \tag{2.13}$$

However, after some tedious calculations, we find that

$$J_F(R) = \operatorname{tr}\left([R + \Delta^{-1}]^{-1} R [R + \Delta^{-1}]^{-1}\right) + \int_{\mathbb{R}^n} z' \Delta^{-1} [R + \Delta^{-1}]^{-1} [R + \Delta^{-1}]^{-1} \Delta^{-1} z\, dF(z) < \infty \tag{2.14}$$

since ξ has finite second moments, whence $J_F(R)$ is finite and so is $I_F(K, R)$ as a result of (2.13). ■

III. Some Convergence Estimates

We shall analyze the asymptotic behavior of $\{\epsilon_t, t = 0, 1, \cdots\}$ by making use of the representation (2.7). This requires that we study the behavior of I_F under the *joint* asymptotic behavior of $\{Q_t^*, t = 0, 1, \cdots\}$ and $\{R_t^*, t = 0, 1, \cdots\}$. However, defining the mapping $I_F^*: \mathscr{Q}_n \to \mathbb{R}$ by

$$I_F^*(R) := I_F(I_n, R), \qquad R \in \mathscr{Q}_n \tag{3.1}$$

we observe the inequalities

$$\lambda_{\min}(Q_t^{*\prime} Q_t^*) I_F^*(R_t^*) \le \epsilon_t((A, H, \Gamma), F) \le \lambda_{\min}(Q_t^{*\prime} q_t^*) I_F^*(R_t^*) \qquad t = 1, 2, \cdots. \tag{3.2}$$

In effect, (3.2) shows how to bound ϵ_t in such a way as to *separately* consider the asymptotic behavior of $\{Q_t^*, t = 0, 1, \cdots\}$ and the asymptotic behavior of I_F^* as $\{R_t^*, t = 0, 1, \cdots\}$ tends to its limit. We focus our attention first on the asymptotics of $\{Q_t^*, t = 0, 1, \cdots\}$ and $\{R_t^*, t = 0, 1, \cdots\}$, and then study the behavior of I_F^* as R_t^* asymptotically behaves in a well-defined way.

To simplify the notation, we define the matrices $\{K_t, t = 0, 1, \cdots\}$ as the elements of $\mathscr{M}_n$ given by

$$K_t := A - [A P_t H' + \Gamma_{wv}][H P_t H' + \Gamma_v]^{-1} H, \qquad t = 0, 1, \cdots, \tag{3.3}$$

so that the recursion (2.3) may now be rewritten in the form

$$Q_0^* = I_n, \; Q_{t+1}^* = K_t Q_t^* \qquad t = 0, 1, \cdots. \tag{3.4}$$

We observe [3, Thm. 5.2(a), p. 171] that $0 \le P_t \le P_{t+1}$ for all $t = 0, 1, \cdots$, whence for each v in $\mathbb{R}^n$, $0 \le v' P_t v \le v' P_{t+1} v$ and $\lim_t V' P_t V$ always exists, although possibly infinite. As this may not imply the convergence of the iterates $\{P_t, t = 0, 1, \cdots\}$ [1, Prop. D.1.4., p. 370], we find it convenient to introduce the following assumption C.1.

C.1: The sequence $\{P_t, t = 0, 1, \cdots\}$ has a well-defined limit P_∞.

In that case, the sequence $\{K_t,\ t=0,1,\cdots\}$ also has a well-defined limit K_∞ which is given by

$$K_\infty := A - [AP_\infty H' + \Gamma_{wv}][HP_\infty H' + \Gamma_v]^{-1}H. \quad (3.5)$$

Conditions for C.1 to hold are given in Proposition 3. We are now ready to present the basic estimates.

Theorem 2: Under C.1, we have the following estimates. The upper bound

$$\overline{\lim}_t \frac{1}{t}\log \lambda_{\max}(Q_t^{*\prime}Q_t^*) \le 2\log\rho(K_\infty) \quad (3.6a)$$

always holds. The lower bound

$$2\log\lambda_{\min}(K_\infty) \le \underline{\lim}_t \frac{1}{t}\log\lambda_{\min}(Q_t^{*\prime}Q_t^*) \quad (3.6b)$$

holds provided either K_∞ is noninvertible or the matrices $\{K_t,\ t=0,1,\cdots\}$ are all invertible.

It is worth pointing out that if K_t is not invertible for some $t=T$, then Q_t^* is not invertible for all $t>T$ by virtue of (3.4) and the lower bound (3.6b) cannot hold with K_∞ invertible. Before giving a proof of Theorem 2, we recall that $\lambda_{\max}(K'K)=\|K\|_{op}^2$ for any matrix K in $\mathcal{M}_n$.

Proof: For every $N=1,2,\cdots$, we define $Z_j^{(N)} := K_{jN-1}\cdots K_{(j-1)N}$ for $j=1,2,\cdots$. As the recursion (3.4) implies $Q_{jN}^* = Z_j^{(N)}\cdots Z_1^{(N)}$ for all $j=1,2,\cdots$, we readily see that

$$\frac{1}{j}\log\|Q_{jN}^*\|_{op} \le \frac{1}{j}\sum_{i=1}^{j}\log\|Z_i^{(N)}\|_{op} \qquad j=1,2,\cdots. \quad (3.7)$$

Since matrix multiplication is continuous in the matrix norm (1.7), it is thus plain from C.1 that $\lim_j Z_j^{(N)} = K_\infty^N$, whence $\lim_j \|Z_j^{(N)}\|_{op} = \|K_\infty^N\|_{op}$, and the estimate

$$\overline{\lim}_j \frac{1}{j}\log\|Q_{jN}^*\|_{op} \le \log\|K_\infty^N\|_{op} \qquad N=1,2,\cdots \quad (3.8)$$

follows from (3.7) by Cesaro convergence.

Fix $N=1,2\cdots$. For each $t=1,2,\cdots$, there exists a unique nonnegative integer $j_N(t)$ such that $j_N(t)N < t \le (j_N(t)+1)N$. Upon iterating (3.4), we get $Q_t^* = K_{t-1}\cdots K_{j_N(t)N}$ for all $t=1,2,\cdots$, from which $\|Q_t^*\|_{op} \le \|K_{t-1}\|_{op}\cdots\|K_{j_N(t)N}\|_{op}\|Q_{j_N(t)N}^*\|_{op}$, and the inequality

$$\frac{1}{t}\log\|Q_t^*\|_{op} \le \frac{1}{t}\sum_{s=j_N(t)N}^{(j_N(t)+1)N-1}|\log\|K_s\|_{op}| + \frac{j_N(t)}{t}\frac{1}{j_N(t)}\log\|Q_{n_N(t)N}^*\|_{op} \qquad t=1,2,\cdots \quad (3.9)$$

readily follows. Now, since $\lim_t j_{N(t)} = \infty$ monotonically and $\lim_j\|K_j\|_{op} = \|K_\infty\|_{op}$, we obtain

$$\lim_t \frac{1}{t}\sum_{s=j_N(t)N}^{(j_N(t)+1)N-1}|\log\|K_s\|_{op}| \le \lim_t\frac{1}{t}N|\log\|K_\infty\|_{op}| = 0. \quad (3.10)$$

On the other hand, the fact $\lim_t \frac{j_N(t)}{t} = \frac{1}{N}$ and the estimate (3.8) lead to

$$\overline{\lim}_t \frac{j_N(t)}{t}\frac{1}{j_N(t)}\log\|Q_{j_N(t)N}^*\|_{op} = \frac{1}{N}\cdot\overline{\lim}_j\frac{1}{j}\log\|Q_{jN}^*\|_{op} \le \frac{1}{N}\cdot\log\|K_\infty^N\|_{op}. \quad (3.11)$$

Collecting (3.9)–(3.11), we get

$$\overline{\lim}_t \frac{1}{t}\log\|Q_t^*\|_{op} \le \frac{1}{N}\cdot\log\|K_\infty^N\|_{op} \qquad N=1,2,\cdots, \quad (3.12)$$

and (3.6a) then follows by letting N become large in (3.12) since $\lim_N(\|K_\infty^N\|_{op})^{\frac{1}{N}} = \rho(K_\infty)$ [8, p. 271 and Thm. 3.8, p. 284].

As we now turn to the proof of (3.6b), we notice that only the case K_∞ invertible needs to be considered for otherwise the result is trivially true. If we assume that K_t is invertible for all $t=0,1,\cdots$, then Q_t^* is also invertible for all $t=0,1,\cdots$. Upon setting $\hat{K}_t = (K_t')^{-1}$ and $\hat{Q}_t = (Q_t^*)^{-1}$ for all $t=0,1,\cdots$, we observe that (3.4) is equivalent to the recursion $\hat{Q}_{t+1}' = \hat{K}_t\hat{Q}_t'$, $t=0,1,\cdots$, whence

$$\overline{\lim}_t \frac{1}{t}\log\|\hat{Q}_t'\|_{op}^2 \le 2\log\rho(\hat{K}_\infty) \quad (3.13)$$

by the arguments leading to (3.6a). From basic arguments, we see that $\rho(\hat{K}_\infty) = \rho(K_\infty^{-1}) = \lambda_{\min}(K_\infty)^{-1}$ and that $\|\hat{Q}_t'\|_{op}^2 = \lambda_{\max}(\hat{Q}_t\hat{Q}_t') = \lambda_{\max}((Q_t^{*\prime}Q_t^*)^{-1}) = \lambda_{\min}(Q_t^{*\prime}Q_t^*)^{-1}$ for each $t=0,1,\cdots$. These facts, when used in (3.13), immediately imply (3.6b). ■

To see the implications of Theorem 2 on the asymptotics of $\{\epsilon_t,\ t=0,1,\cdots\}$, we shall need the following fact.

Proposition 2: For every distribution F in $\mathcal{G}_n$, we have $\sup_t I_F^*(R_t^*) < \infty$.

Proof: Since $0 \le R_t^* \le R_{t+1}^*$ for all $t=0,1,\cdots$, we conclude from (2.14) that $\sup_t J_F(R_t^*) < \infty$ and the result follows from (2.13). ■

Note that in Proposition 2 we did not impose the requirement that the sequence $\{R_t^*,\ t=0,1,\cdots\}$ be convergent. In analogy with C.1, we introduce assumption C.2.

C.2: The sequence $\{R_t^*,\ t=0,1,\cdots\}$ has a well-defined limit R_∞^* which is positive definite.

Theorem 3: Assume C.1. We have the following estimates for every non-Gaussian distribution F in $\mathcal{G}_n$. The upper bound

$$\overline{\lim}_t \frac{1}{t}\log\epsilon_t \le 2\log\rho(K_\infty) \quad (3.14a)$$

always holds. If in addition C.2 holds, then the lower bound

$$2\log\lambda_{\min}(K_\infty) \le \underline{\lim}_t\frac{1}{t}\log\epsilon_t \quad (3.14b)$$

holds provided either K_∞ is noninvertible or the matrices $\{K_t,\ t=0,1,\cdots\}$ are all invertible.

Proof: The upper bound (3.14a) follows from (3.2) and (3.6a) with the help of Proposition 2. Under C.2, Theorem 6 of Section IV implies $\underline{\lim}_t I_F^*(R_t^*) > 0$ for F non-Gaussian and (3.14b) now follows from (3.2) and (3.6b). ■

We now present some simple implications of Theorems 2 and 3 on the asymptotics considered here. For future reference, we note from (2.4) that

$$R_t^* = \sum_{s=0}^{t-1} Q_s^{*\prime}H'[HP_sH' + \Gamma_v]^{-1}HQ_s^* \qquad t=1,2,\cdots. (3.15)$$

Theorem 4: Assume C.1. If $\rho(K_\infty) <$, then;

1) the sequence $\{Q_t^*,\ t=0,1,\cdots\}$ converges with $\lim_t Q_t^* = 0$;

2) the sequence $\{R_t^*,\ t=0,1,\cdots\}$ has a well-defined limit R_∞^*;

3) for all non-Gaussian distributions F in $\mathcal{G}_n$, the convergence $\lim_t \epsilon_t = 0$ takes place at least exponentially fast according to (3.14a).

Proof: From (3.6a) and the fact that $\lambda_{\max}(Q_t^{*\prime}Q_t^*) = \|Q_t^*\|_{op}^2$

for all $t = 0, 1, \cdots$, we readily obtain $\overline{\lim}_t \frac{1}{t} \log \|Q_t^*\|_{op} < 0$, so that the convergence $\lim_t \|Q_t^*\|_{op} = 0$ takes place at least exponentially fast. Claim 1 now follows from the fact that all norms are equivalent on $\mathcal{M}_n$.

To obtain claim 2, we note that (3.15) that

$$\|R_t^* - R_s^*\|_{op} \le \frac{\lambda_{\max}(H'H)}{\lambda_{\min}(\Gamma_v)} \sum_{r=s}^{t-1} \|Q_r^*\|_{op}^2, \qquad s < t \qquad s, t = 0, 1, \cdots, \quad (3.16)$$

and since $\lim_t \|Q_t^*\|_{op} = 0$ at least exponentially fast, the sequence $\{R_t^*, t = 0, 1, \cdots\}$ is Cauchy, thus convergent, in the matrix norm (1.7). The norm equivalence invoked earlier completes the proof of claim 2. Claim 3 is immediate from (3.14a). ■

We conclude this section with a set of sufficient conditions which ensure C.1 as well as the condition $\rho(K_\infty) < 1$.

Proposition 3: If the pair $(\overline{A}, H)$ is detectable, then C.1 holds. If, in addition, the pair $(\overline{A}, \overline{C}^{1/2})$ is stabilizable, then the matrix K_∞ is asymptotically stable, i.e., $\rho(K_\infty) < 1$.

Proof: The first claim is [3, Thm. 5.2b, p. 172], while the second claim follows from [3, Thm. 5.3, p. 175]. ■

IV. A Partial Converse

In this section, we present the key technical fact required in proving (3.14b). This result provides an indirect characterization of the initial condition as a Gaussian random variable.

Theorem 6: Assume C.2. For any distribution F in $\mathcal{G}_n$, the condition $\underline{\lim}_t I_F^*(R_t^*) = 0$ implies that F is necessarily Gaussian.

Proof: First we introduce the distribution $\hat{F}$ in $\mathcal{G}_n$ which is absolutely continuous with respect to F and whose Radon–Nikodym derivative is given by

$$\frac{d\hat{F}}{dF}(z) = \frac{\exp\left[-\frac{1}{2}z'R_\infty^* z\right]}{\int_{\mathbb{R}^n} \exp\left[-\frac{1}{2}z'R_\infty^* z\right] dF(z)}, \qquad z \in \mathbb{R}^n. \quad (4.1)$$

The moment generating N of $\hat{F}$ is simply

$$N(b) := \int_{\mathbb{R}^n} \exp[z'b]\, d\hat{F}(z), \qquad b \in \mathbb{R}^n. \quad (4.2)$$

Since the matrix R_∞^* is positive definite, there exists a finite T such that for $t = T, T+1, \cdots$, the matrix R_t^* is also positive definite and thus $G_{R_t^*}$ is absolutely continuous with respect to Lebesgue measure λ on $\mathbb{R}^n$. Applying Fatou's lemma to (2.5), we see from the assumption $\underline{\lim}_t I_F^*(R_t^*) = 0$ that

$$\underline{\lim}_t \frac{\left\| \int_{\mathbb{R}^n} \{z - [R_t^* + \Delta^{-1}]^{-1} b\} \phi(z, b; R_t^*)\, dF(z) \right\|^2}{\Phi(b; R_t^*)} \cdot \frac{dG_{R_t^*}}{d\lambda}(b) = 0 \qquad \lambda - \text{a.e.} \quad (4.3)$$

Under C.2, for each b in $\mathbb{R}^n$, we have $\lim_t \frac{dG_{R_t^*}}{d\lambda}(b) = \frac{dG_{R_\infty^*}}{d\lambda}(b) > 0$ and $\lim_t \Phi(b; R_t^*) = \Phi(b, R_\infty^*) > 0$, with the last following by monotone convergence. We now conclude that

$$\underline{\lim}_t \left\| \int_{\mathbb{R}^n} \{z - [R_t^* + \Delta^{-1}]^{-1} b\} \phi(z, b; R_t^*)\, dF(z) \right\| = 0 \; \lambda - \text{a.e.} \quad (4.4)$$

or equivalently,

$$\int_{\mathbb{R}^n} z\phi(z, b; R_\infty^*)\, dF(z) = [R_\infty^* + \Delta^{-1}]^{-1} b \int_{\mathbb{R}^n} \phi(z, b; R_\infty^*)\, dF(z) \qquad \lambda - \text{a.e.} \quad (4.5)$$

Upon dividing (4.5) by $\int_{\mathbb{R}^n} \exp[-\frac{1}{2}z'R_\infty^* z]\, dF(z)$, we readily see that N must satisfy the conditions

$$\nabla_b N(b) = [R_\infty^* + \Delta^{-1}]^{-1} b N(b), \qquad b \in \mathbb{R}^n \text{ with } N(0) = 1; \quad (4.6)$$

the technical details are found in [5, Sect. VI.2].

The unique solution of (4.6) is

$$N(b) = \exp\left[\tfrac{1}{2} b'[R_\infty^* + \Delta^{-1}]^{-1} b\right], \qquad b \in \mathbb{R}^n \quad (4.7)$$

so that $\hat{F}$ is zero-mean Gaussian with covariance $[R_\infty^* + \Delta^{-1}]^{-1}$. Since $\hat{F}$ has positive definite covariance, we see that $\hat{F}$ is absolutely continuous with respect to λ and, therefore, F must be absolutely continuous with respect to λ by virtue of the mutual absolute continuity of F and $\hat{F}$. After some straightforward calculations, we find

$$\frac{dF}{d\lambda}(z) = \frac{dF}{d\hat{F}}(z) \cdot \frac{d\hat{F}}{d\lambda}(z) = c \exp\left[-\frac{1}{2} z'\Delta^{-1} z\right], \qquad z \in \mathbb{R}^n \quad (4.8)$$

for some positive constant c, i.e., the distribution F is Gaussian. ■

As an immediate consequence of Theorem 6 and of the lower bound in (3.2), we observe that under condition C.2, whenever $\lambda_{\min}(Q_t^{*\prime} Q_t^*) > 0$ for all t sufficiently large, the distribution F is necessarily Gaussian if

$$\lim_t \frac{\epsilon_t((A, H, \Gamma), F)}{\lambda_{\min}(Q_t^{*\prime} Q_t^*)} = 0. \quad (4.9)$$

V. The Scalar Case

In this section, we focus exclusively on the scalar case $n = 1$. We use lower case letters for all deterministic quantities. Moreover, to conform to standard usage, we also set $\gamma_v = \sigma_v^2$, $\gamma_w = \sigma_w^2$, and $\gamma_{vw} = \gamma_{wv} = \gamma$ with $\sigma_v^2 > 0$, so that

$$\bar{a} := a - \frac{\gamma h}{\sigma_v^2} \quad \text{and} \quad \bar{c} := \sigma_w^2 - \frac{\gamma^2}{\sigma_v^2} \ge 0. \quad (5.1)$$

With this notation, we can rewrite the recursions (2.2)–(2.4) as

$$p_0 = 0, \; p_{t+1} = \frac{\bar{a}^2 \sigma_v^2 p_t}{h^2 p_t + \sigma_v^2} + \bar{c} \qquad t = 0, 1, \cdots, \quad (5.2)$$

$$q_0^* = 1, \; q_{t+1}^* = \left(\frac{\bar{a}\sigma_v^2}{h^2 p_t + \sigma_v^2}\right) q_t^* = k_t q_t^* \qquad t = 0, 1, \cdots, \quad (5.3)$$

and

$$r_0^* = 0, \; r_{t+1}^* = r_t^* + \frac{(q_t^*)^2 h^2}{h^2 p_t + \sigma_v^2} \qquad t = 0, 1, \cdots. \quad (5.4)$$

Moreover, the representation (2.7) now takes the form

$$\epsilon_t = (q_t^*)^2 I_F^*(r_t^*), \; F \in \mathcal{G}_1 \qquad t = 1, 2, \cdots. \quad (5.5)$$

As we shall see, the scalar nature of the recursions (5.2)–(5.4) permits simpler arguments which are not available in the multivariable case. Although Theorem 5 might be suggestive of a taxonomy based on the detectability of (a, h) and the stabilizability of $(\bar{a}, \bar{c}^{1/2})$, a more direct classification will emerge from our discussion of the scalar situation. We need only consider four possibilities parametrized by h, $\bar{a}$, and $\bar{c}$, and start with an obvious degeneracy.

Proposition 4: If either $\bar{a} = 0$ or $h = 0$, then $\epsilon_t = 0$ for all $t = 1, 2, \cdots$, and all distributions F in $\mathcal{E}_1$.

Proof: Fix F in $\mathcal{G}_1$. If $\bar{a} = 0$, then $q_t^* = 0$ for all $t = 1, 2, \cdots$, by (5.3), and (5.5) therefore implies $\epsilon_t = 0$ for all $t = 1, 2, \cdots$. On the other hand, if $h = 0$, then (5.4) leads to $r_t^* = 0$ for all $t = 0, 1, \cdots$, so that $\epsilon_t = 0$ for all $t = 1, 2, \cdots$, by direct evaluation of (2.6). We translate these results from $\mathcal{G}_1$ to $\mathcal{E}_1$ by making use of (2.1). ■

We now consider the more interesting situation where both conditions $\bar{a} \neq 0$ and $h \neq 0$ are assumed, in which case $q_t^* \neq 0$, $r_t^* > 0$, and $\epsilon_t > 0$ for all $t = 1, 2, \cdots$. We rewrite (5.2) as $p_{t+1} = T(p_t)$ where the mapping $T:[0, \infty) \to \mathbb{R}$ is given by

$$T(p) := \frac{\bar{a}^2 \sigma_v^2 p}{h^2 p + \sigma_v^2} + \bar{c}, \qquad p \geq 0. \tag{5.6}$$

Since

$$T'(p) = \frac{\bar{a}^2 \sigma_v^4}{\left(h^2 p + \sigma_v^2\right)^2}, \qquad p \geq 0 \tag{5.7}$$

we conclude that T is concave and nondecreasing on $[0, \infty)$. Hence, the iterates $\{p_t, t = 0, 1, \cdots\}$ form a nondecreasing, thus convergent, sequence with limit point p_∞ in $[0, \infty)$. The finiteness of p_∞ is an easy consequence of the relation $p_\infty = T(p_\infty)$ which must necessarily hold.

Consequently, the sequence $\{k_t, t = 0, 1, \cdots\}$ has a limit k_∞ given by

$$k_\infty := \frac{\bar{a}\sigma_v^2}{h^2 p_\infty + \sigma_v^2} \tag{5.8}$$

with $|k_\infty| > 0$ since $\bar{a} \neq 0$ and $p_\infty < \infty$. The convergence of the sequence $\{p_t, t = 0, 1, \cdots\}$ and (5.3) readily imply the Cesaro convergence

$$\lim_t \frac{1}{t} \log \left(q_t^*\right)^2 = \lim_t \frac{1}{t} \sum_{s=0}^{t-1} \log \left(\frac{\bar{a}\sigma_v^2}{h^2 p_s + \sigma_v^2}\right)^2 = 2 \log |k_\infty|. \tag{5.9}$$

It is then easy to see from (3.15) and (5.9) that if $|k_\infty| < 1$, then $r_\infty^* := \lim_t r_t^*$ is well defined and finite, whereas if $|k_\infty| \geq 1$, then $\lim_t r_t^* = \infty$. We now make use of these observations to prove the following result.

Proposition 5: We assume both $h \neq 0$ and $\bar{a} \neq 0$. If either $\bar{c} \neq 0$ or $\bar{c} = 0$ with $|\bar{a}| < 1$, then $|k_\infty| < 1$ and $\lim_t \epsilon_t = 0$ with $\lim_t \frac{1}{t} \log \epsilon_t = 2 \log |k_\infty| < 0$ for all non-Gaussian distributions F in $\mathcal{E}_1$.

Proof: Prompted by the remarks made earlier, we begin by showing that $|k_\infty| < 1$ under the stated conditions. If $\bar{c} = 0$, then $p_t = 0$ for all $t = 0, 1, \cdots$, so that $p_\infty = 0$ and the conclusion $|k_\infty| \leq |\bar{a}| < 1$ follows when $|\bar{a}| < 1$. If $\bar{c} \neq 0$, then necessarily $\bar{c} > 0$ and therefore $p_\infty > 0$ (since $\bar{c} = p_1 \leq p_\infty$). Consequently, p_∞ is the only finite solution to the fixed point equation $T(p) = p$ on $(0, \infty)$, and geometric considerations based on the concavity and monotonicity of T readily lead to $T'(p_\infty) < 1$. The conclusion $|k_\infty| < 1$ now follows from the fact that $T'(p_\infty) = k_\infty^2$.

As pointed out earlier, here $q_t^* \neq 0$ and $r_t^* > 0$ for all $t = 0, 1, \cdots$, whence $r_\infty^* > 0$ since $\{r_t^*, t = 0, 1, \cdots\}$ is an increasing sequence. On the other hand, we saw earlier that $|k_\infty| < 1$ implies $r_\infty^* < \infty$. Therefore, from Proposition 1 and Theorem 6, we obtain $0 < \underline{\lim}_t I_F^*(r_t^*) \leq \overline{\lim}_t I_F^*(r_t^*) < \infty$ for every non-Gaussian F in $\mathcal{G}_1$. As a result, $\lim_t \frac{1}{t} \log \epsilon_t = \lim_t \frac{1}{t} \log (q_t^*)^2 = 2 \log |k_\infty| < 0$ for all F non-Gaussian in $\mathcal{G}_1$, and thus in $\mathcal{E}_1$ by translation. ■

Notice that Proposition 5 is almost a direct consequence of Theorem 3 since in the scalar case, we have $\lambda_{\min}(k_\infty) = \rho(k_\infty) = |k_\infty|$, and we need only establish that conditions C.1 and C.2 hold true under the assumptions of Proposition 5. We found it interesting, however, to provide a direct argument tailored to the scalar case.

It now remains to investigate the case $\bar{c} = 0$ and $|\bar{a}| \geq 1$, still with $h \neq 0$. We shall see that the initial state distribution F has a nontrivial effect on the large time asymptotics of $\{\epsilon_t, t = 0, 1, 2, \cdots, \}$. *A priori*, it would seem natural that the initial distribution F should have some effect on the asymptotics of the mean squared error between the MMSE and LMSE filters. However, in both cases considered thus far in Propositions 4 and 5, the effect of the system parameters (a, h, Γ) have dominated these asymptotics. Only when $\bar{c} = 0$ and $|\bar{a}| \geq 1$, does F have a significant effect. We shall establish this dependence by performing a complete analysis for two specific initial distributions F, and by noting the different asymptotics of $\{\epsilon_t, t = 0, 1, \cdots\}$. We first verify a general result which complements Proposition 2.

Proposition 6: For any distribution F in $\mathcal{G}_1$, we have $I_F^*(r) \leq \frac{4}{r}$ for all $r > 0$ so that $\lim_r I_F^*(r) = 0$.

Proof: We note that the functional I_F^* is *independent* of the system dynamics (a, h, Γ). Consequently, for the purpose of argumentation, we can take the system (1.1) to be

$$X_t^o = \xi, \; Y_t = \xi + V_{t+1}^o, \qquad t = 0, 1, \cdots, \tag{5.10}$$

with $a = h = 1$, $\sigma_w^2 = \gamma = 0$, $\sigma_v^2 > 0$. For this system, $q_t^* = 1$, $r_t^* = \frac{t}{\sigma_v^2}$, and $\epsilon_t = I_F^*\left(\frac{t}{\sigma_v^2}\right)$ for all $t = 0, 1, \cdots$. We now set

$$\check{X}_{t+1} := \frac{1}{t+1} \sum_{s=0}^{t} Y_s \qquad t = 0, 1, \cdots, \tag{5.11}$$

and observe that since $\check{X}_{t+1}$ is a linear estimate of X_{t+1}^o on the basis of $\{Y_0, \cdots, Y_t\}$, it has larger mean squared error than both the LMSE estimates $\hat{X}_{t+1}^K$ and the MMSE estimate $\hat{X}_{t+1}$. Therefore, using the triangular inequality, we readily find

$$E\left[|\hat{X}_{t+1} - \hat{X}_{t+1}^K|^2\right] \leq 4E\left[|\check{X}_{t+1} - X_{t+1}^o|^2\right] \qquad t = 0, 1, \cdots, \tag{5.12}$$

so that

$$I_F^*\left(\frac{t}{\sigma_v^2}\right) = \epsilon_t \leq 4E\left[\left|\frac{1}{t} \sum_{s=0}^{t-1} V_{s+1}^o\right|^2\right] = \frac{4\sigma_v^2}{t} \qquad t = 1, 2 \cdots, \tag{5.13}$$

and the result follows since σ_v^2 is arbitrary. ■

We now consider the following two distributions F_1 and F_2 in $\mathcal{G}_1$.

Distribution F_1: Distribution F_1 admits a density with respect to Lebesgue measure λ on $\mathbb{R}$ given by

$$\frac{dF_1}{d\lambda}(z) = \sum_{i=1}^{m} a_i \frac{1}{\sqrt{2\pi\rho^2}} \exp\left[-\frac{1}{2}\frac{(z-\mu)^2}{\rho^2}\right], \qquad z \in \mathbb{R} \tag{5.14}$$

where $\rho > 0$, $0 < \alpha_i < 1$ for $i = 1, 2, \cdots, m$, $\sum_{i=1}^m \alpha_i = 1$, and $\sum_{i=1}^m \alpha_i \mu_i = 0$. We exclude the case where F_1 is actually Gaussian.

Distribtion F_2: Under F_2, the RV ξ takes on a finite number of values $z_1 < z_2 \cdots < z_m$ with probabilities $p_1, p_2, \cdots, p_m$, respectively, such that $\sum_{i=1}^m p_i z_i = 0$.

The following two facts are proved in [5].

Fact 1: We have

$$I_{F_1}^*(r) = \frac{K + o(1)}{(\rho^2 r + 1)^2}, \qquad r > 0 \tag{5.15}$$

for some $K > 0$.

Fact 2: We also have

$$I_{F_2}^*(r) = \frac{1 + o(1)}{r}, \qquad r > 0. \tag{5.16}$$

We now can prove the following results.

Proposition 7: If $h \neq 0$, $|\bar{a}| = 1$, and $\bar{c} = 0$, then $\lim_t \epsilon_t = 0$ for any distribution F in $\mathscr{E}_1$, with $\overline{\lim}_t \frac{1}{t} \log \epsilon_t \leq 0$. This convergence takes place at a rate which depends nontrivially upon F for non-Gaussian F.

Proof: Under the stated hypothesis, we have $p_t = 0$, $(q_t^*)^2 = 1$, $r_t^* = \frac{h^2}{\sigma_v^2} t$, and $\epsilon_t = I_F^*\left(\frac{h^2}{\sigma_v^2} t\right)$ for all $t = 0, 1, \cdots$, and all F in $\mathscr{D}_1$, the extension to $\mathscr{E}_1$ being as before. The conclusions $\lim_t \epsilon_t = 0$ and $\overline{\lim}_t \frac{1}{t} \log \epsilon_t \leq 0$ are immediate consequences of Proposition 6. However, direct calculations show that if $F = F_1$, then $\lim_t t^2 \epsilon_t = \frac{K}{\rho^2}$, whereas if $F = F_2$, then $\lim_t t\epsilon_t = 1$ (so that $\lim_t \frac{1}{t} \log \epsilon_t = 0$ in both cases). ■

And finally, we have the following.

Proposition 8: If $h \neq 0$, $|\bar{a}| > 1$, and $\bar{c} = 0$, then $\overline{\lim}_t \epsilon_t < \infty$ for all distributions F in $\mathscr{E}_1$, the asymptotic behavior depending nontrivially upon F for non-Gaussian F.

Proof: Under the stated hypotheses on (a, h, Γ), $p_t = 0$, $(q_t^*)^2 = \bar{a}^{2t}$, $r_t^* = \frac{h^2}{\sigma_v^2} \frac{\bar{a}^{2t} - 1}{a^2 - 1}$ for all $t = 0, 1, \cdots$. Thus, $\lim_t r_t^* = \infty$ with $\lim_t (q_t^*)^2 / r_t^* = \sigma_v^2(\bar{a}^2 - 1)/h^2$ and we are lead to write

$$\epsilon_t = \frac{(q_t^*)^2}{r_t^*}\left(r_t^* I_F^*(r_t^*)\right) \leq 4\frac{\sigma_v^2}{h^2}(\bar{a}^2 - 1) \cdot \frac{\bar{a}^{2t}}{\bar{a}^{2t} - 1}, \qquad t = 1, 2, \cdots \tag{5.17}$$

where the inequality follows from Proposition 6. We now see that $\overline{\lim}_t \epsilon_t < \infty$ for all F in $\mathscr{D}_1$, and thus for all distributions F in $\mathscr{E}_1$. However, if $F = F_1$, then $\lim_t \epsilon_t = 0$, whereas if $F = F_2$, then $\lim_t \epsilon_t = 1$. ■

We conclude with the following remark which is also valid in the multivariable case and which complements some of the results obtained so far. By an argument similar to the one leading to (5.12) we readily see that for each $\delta > 0$

$$\epsilon_t \leq 4E\left[|X_t^o - \hat{X}_t^K|^2\right] = 4p_t^\delta \qquad t = 1, 2, \cdots \tag{5.18}$$

where the error variance $\{p_t^\delta, t = 0, 1, \cdots\}$ are generated through the recursion (5.2) with initial condition $p_0^\delta = \delta$. The sequence $\{p_t^\delta, t = 0, 1, \cdots\}$ is either monotone nondecreasing or monotone nonincreasing, thus convergent, with limit point p_∞^δ. Therefore, whenever $p_\infty^\delta < \infty$, we conclude by inspection that

$$\epsilon_t \leq 4 \max\{\delta, p_\infty^\delta\} \qquad t = 1, 2, \cdots. \tag{5.19}$$

In particular, under the conditions of Proposition 8, i.e., $h \neq 0$, $|\bar{a}| > 1$, and $\bar{c} = 0$, we have (5.19) with $p_\infty^\delta = \frac{h^2}{\sigma_v^2(\bar{a}^2 - 1)}$ (a fact in agreement with the conclusion of Proposition 8).

As all possible combinations of $\bar{a}$, $\bar{c}$, and h have now been considered, a careful review of our analysis suggests the following classification. For any matrices $\tilde{A}$ and $\tilde{C}$ in $\mathscr{M}_n$, the pair $(\hat{A}, \tilde{C})$ is said to be *marginally stabilizable* if all modes which are neither stable nor critically stable, are in the controllable subspace. Equipped with this notion, we can now rewrite the results of this section in terms which are also meaningful for the multivariable case. As such, this formulation provides a useful starting point for investigating the asymptotics in the nonscalar case.

Theorem 7: We have the following convergence results:

1a) if the pair $(\bar{a}, \bar{c})$, is marginally stabilizable, $\lim_t \epsilon_t = 0$ for any distribution F in $\mathscr{E}_1$;

1b) if the pair $(\bar{a}, \bar{c})$ is not marginally stabilizable, then the asymptotic behavior of ϵ_t depends nontrivially upon F in $\mathscr{E}_1$.

Moreover, we also have the following estimates:

2a) if $(\bar{a}, \bar{c})$ is stabilizable, then $\lim_t \epsilon_t = 0$ at an exponential rate independent of F for non-Gaussian F in $\mathscr{E}_1$;

2b) if $(\bar{a}, \bar{c})$ is marginally stabilizable but not stabilizable, then the rate depends nontrivially upon F.

Acknowledgment

The authors thank the referee for pointing out an embarassing mistake in an earlier version of this note and for making suggestions that led to an improved presentation.

References

[1] B. D. O. Anderson and J. B. Moore, *Optimal Filtering* (Information and System Sciences Series). Englewood Cliffs, NJ: Prentice-Hall, 1979.

[2] M. H. A. Davis and R. B. Vinter, *Stochastic Modeling and Control* (Monographs on Statistics and Applied Probability). London: Chapman and Hall, 1985.

[3] P. E. Caines, *Linear Stochastic Systems.* New York: Wiley, 1988.

[4] A. M. Makowski, "Filtering formulae for partially observed linear systems with non-Gaussian initial conditions," *Stochastics*, vol. 16, pp. 1-24, 1986.

[5] R. B. Sowers, "New discrete-time filtering results," M.S. thesis, Dep. Elec. Eng., Univ. Maryland, College Park, MD, Aug. 1988; also in Systems Research Center, Tech. Rep. TR 88-85, 1988.

[6] R. B. Sowers and A. M. Makowski, "Discrete-time filtering for linear systems in correlated noise with non-Gaussian initial conditions," in *Proc. 23rd Ann. Conf. Inform. Sci. Syst.*, Johns Hopkins University, Baltimore, MD, Mar. 1989, pp. 308-313.

[7] —, "Discrete-time filtering for linear systems in correlated noise with non-Gaussian initial conditions: Formulas and asymptotics," in *Robust Control of Linear Systems and Nonlinear Control*, E. M. A. Kaashoek, J. H. van Schuppen, and A. C. M. Ran, Eds. Cambridge, MA: Birkhäuser, also in *Proc. 1989 Int. Symp. Math. Theory Networks Syst.*, Amsterdam, The Netherlands, June 1989 1990.

[8] G. W. Stewart, *Introduction to Matrix Computations.* New York: Academic, 1973.