SRC TR 89-38

Discrete-Time Filtering For Linear Systems In Correlated Noise With Non-Gaussian Initial Conditions: Asymptotic Behavior Of The Difference Between The MMSE & LMSE Estimates

By

R. B. Sowers
&
A. M. Makowski

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
1989		**00-00-1989 to 00-00-1989**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Discrete-Time Filtering for Linear Systems in Correlated noise with Non-Gaussian Initial Conditions: Asymptotic Behavior ot the Difference Between the MMSE and LMSE Estimates	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of Maryland,Systems Research Center,College Park,MD,20742**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**17**	

DISCRETE-TIME FILTERING FOR LINEAR SYSTEMS IN CORRELATED NOISE WITH NON-GAUSSIAN INITIAL CONDITIONS: ASYMPTOTIC BEHAVIOR OF THE DIFFERENCE BETWEEN THE MMSE AND LMSE ESTIMATES

by

Richard B. Sowers[1] and Armand M. Makowski[2]
University of Maryland

28th CDC and T-AC

ABSTRACT

We consider the one-step prediction problem for discrete-time linear systems in correlated plant and observation noises, and non-Gaussian initial conditions. We investigate the asymptotic behavior of the expected square ϵ_t of the difference between the MMSE and LMMSE (or Kalman) estimates of the state given past observations. We characterize the *limit* of the error sequence $\{\epsilon_t,\ t = 0, 1, \ldots\}$ and obtain some related *rates* of convergence, with complete analysis being provided for the scalar case. The discussion is based on the explicit representations which were obtained by the authors in [,] for the MMSE and LMMSE estimates, and which explicitly display the dependence of these quantities on the initial distribution.

Key Words: Filtering, Linear systems, non-Gaussian initial conditions, correlated noises, Girsanov transformation.

---

[1] Department of Electrical Engineering and Systems Research Center, University of Maryland, College Park, MD 20742. The work of this author was supported through an ONR Graduate Fellowship while completing the requirements towards the M.S. degree. Currently with the Department of Mathematics, University of Maryland, College Park, MD 20742.

[2] Electrical Engineering Department and Systems Research Center, University of Maryland, College Park, Maryland 20742. The work of this author was supported partially through the Engineering Research Centers Programs NSF CDR 88-03012 and partially through NSF grant ECS 83-51836.

I. INTRODUCTION

Consider the discrete-time linear time-invariant stochastic system

$$\begin{aligned} X^\circ_{t+1} &= AX^\circ_t + W^\circ_{t+1} \\ X^\circ_0 &= \xi \qquad\qquad t = 0,1,\dots \\ Y_t &= HX^\circ_t + V^\circ_{t+1}. \end{aligned} \tag{1.1}$$

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 the $I\!R^n$-valued plant process $\{X^\circ_t,\ t = 0,1,\dots\}$ and the $I\!R^k$-valued observation process $\{Y_t,\ t = 0,1,\dots\}$. Throughout we make the following assumptions **(A.1)**-**(A.3)**, where

(A.1): The process $\{(W^\circ_{t+1}, V^\circ_{t+1}),\ t = 0,1,\dots\}$ is a stationary zero-mean Gaussian White Noise (GWN) sequence with covariance structure Γ given by

$$\Gamma := \mathrm{Cov}\begin{pmatrix} W^\circ_{t+1} \\ V^\circ_{t+1} \end{pmatrix} = \begin{pmatrix} \Sigma^w & \Sigma^{wv} \\ \Sigma^{vw} & \Sigma^v \end{pmatrix}, \qquad t = 0,1,\dots \tag{1.2}$$

(A.2): The initial condition ξ has distribution F with finite first and second moments μ and Δ, respectively, and is independent of the process $\{(W^\circ_{t+1}, V^\circ_{t+1}),\ t = 0,1,\dots\}$, and

(A.3): The covariance matrices Σ^v and Δ are positive definite.

For each $t = 0,1,\dots$, we form the conditional mean $\hat{X}_{t+1} := E[X^\circ_{t+1}|Y_0, Y_1, \dots, Y_t]$ or MMSE estimate of X°_{t+1} on the basis of $\{Y_0, Y_1, \dots, Y_t\}$. In general, $\hat{X}_{t+1}$ is a *non*-linear function of $\{Y_0, Y_1, \dots, Y_t\}$, in contrast to the LMSE or Kalman estimate of X°_{t+1} on the basis of $\{Y_0, Y_1, \dots, Y_t\}$, which is by definition linear, and which we denote by $\hat{X}^K_{t+1}$. For each $t = 0,1,\dots$, we can then calculate $\epsilon_{t+1} := E[\|\hat{X}_{t+1} - \hat{X}^K_{t+1}\|^2]$ which is an L^2 measure of the agreement between the MMSE and LMSE estimates of X°_t on the basis of $\{Y_0, Y_1, \dots, Y_t\}$.

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

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

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

We are interested in characterizing the *limit* of the error sequence $\{\epsilon_t,\ t = 0,1,\ldots\}$ and in obtaining the corresponding *rate* of convergence. 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 $I > 0$.

To the authors' knowledge, no results have been reported in the literature to study the 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 of Theorem 1 has been derived only relatively recently (although, see [*]). In any case, the work reported here provides a formal justification to 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^o_{t+1} on the basis of $\{Y_0, Y_1, \ldots, Y_t\}$.

The organization of this paper is as follows. In Section II we summarize a representation result for $\{\epsilon_t,\ t = 0,1,\ldots\}$ which constitutes the basis for the analysis presented here. In Section III, we investigate the asymptotic behavior of $\{\epsilon_t,\ t = 0,1,\ldots\}$ for a general multivariable system; this is followed in Section IV by a more complete analysis of the scalar case when $n = k = 1$.

The following notation is used throughout. Elements of IR^n are viewed as column vectors and transposition is denoted by $'$. For any positive integers m and n, we denote by $\mathcal{M}_{n\times m}$ the space of $n \times m$ real matrices and by $\mathcal{Q}_n$ the cone of $n \times n$ nonnegative definite matrices. For each positive integer n, let I_n and O_n be the unit and zero elements in $\mathcal{M}_{n\times n}$. Also, for any matrix K in $\mathcal{M}_{n\times n}$, we define $\mathrm{sp}(K)$ as the set of all eigenvalues of K, and set

$$\lambda_{\min}(K) := \min\{|\lambda| : \lambda \in \mathrm{sp}(K)\} \tag{1.4a}$$

and

$$\lambda_{\max}(K) := \max\{|\lambda| : \lambda \in \mathrm{sp}(K)\}. \tag{1.4b}$$

We let $\mathcal{E}_n$ be the convex set of square-integrable probability distributions functions on $(IR^n, \mathcal{B}(IR^n))$ and we define $\mathcal{D}_n$ as the collection of those distributions in $\mathcal{E}_n$ with zero-mean. Finally, for each matrix R in $\mathcal{Q}_n$, G_R denotes the distribution of an IR^n-valued Gaussian RV with zero mean and covariance R.

II. A REPRESENTATION RESULT

The basis for our analysis is a representation result for the sequence $\{\epsilon_t,\ t = 0,1,\ldots\}$ obtained in [ ]. However, before stating this result, we find it useful to observe that there is no loss in

generality in assuming $E[\xi] = 0$. Indeed, with the notation

$$\check{X}_t^\circ := X_t^\circ - \Phi(t,0)\mu \quad \text{and} \quad \check{Y}_t := Y_t - H\Phi(t,0)\mu, \qquad t = 0,1,\ldots \tag{2.1}$$

we see that the RV's $\{\check{X}_t^\circ,\ t = 0,1,\ldots\}$ and $\{\check{Y}_t,\ t = 0,1,\ldots\}$ obey the dynamics

$$\begin{aligned} \check{X}_{t+1}^\circ &= A\check{X}_t^\circ + W_{t+1}^\circ \\ \check{X}_0^\circ &= \check{\xi} \qquad\qquad t = 0,1,\ldots \\ \check{Y}_t &= H\check{X}_t + V_{t+1}^\circ \end{aligned} \tag{2.2}$$

where the RV $\check{\xi} := \xi - \mu$ satisfies the zero-mean condition $E[\check{\xi}] = 0$. If $\hat{E}[A|B]$ denotes the LMSE estimate of A on the basis of B for any square-integrable random vectors A and B, we conclude from basic principles that for each $t = 0,1,\ldots,$

$$\begin{aligned} \hat{E}[\check{X}_{t+1}^\circ|\check{Y}_0,\check{Y}_1,\ldots,\check{Y}_t] &= \hat{E}[\check{X}_{t+1}^\circ|Y_0,Y_1,\ldots,Y_t] \\ &= \hat{E}[X_{t+1}^\circ|Y_0,Y_1,\ldots,Y_t] - A^t\mu \end{aligned} \tag{2.3}$$

and

$$\begin{aligned} E[\check{X}_{t+1}^\circ|\check{Y}_0,\check{Y}_1,\ldots,\check{Y}_t] &= E[\check{X}_{t+1}^\circ|Y_0,Y_1,\ldots,Y_t] \\ &= \hat{X}_{t+1}^K - A^t\mu \end{aligned} \tag{2.4}$$

so that

$$\hat{X}_{t+1} - \hat{X}_{t+1}^K = E[\check{X}_{t+1}^\circ|\check{Y}_0,\check{Y}_1,\ldots,\check{Y}_t] - \hat{E}[\check{X}_{t+1}^\circ|\check{Y}_0,\check{Y}_1,\ldots,\check{Y}_t]. \tag{2.5}$$

Consequently, for any distribution F in $\mathcal{E}_n$ and any triple (A,H,Γ), the relation

$$\epsilon_t\left((A,H,\Gamma),F\right) = \epsilon_t\left((A,H,\Gamma),\bar{F}\right) \tag{2.6}$$

holds where $\bar{F}$ is the element of $\mathcal{D}_n$ given by

$$\bar{F}(x) := F(x-\mu), \quad x \in I\!R^n \tag{2.7}$$

and we may thus restrict our attention to those distributions F in $\mathcal{D}_n$.

We now can state the needed representation result, the proof of which is found in [4].

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

$$\begin{aligned} P_{t+1} &= AP_tA' - [AP_tH' + \Sigma^{wv}][HP_tH' + \Sigma^v]^{-1}[AP_tH' + \Sigma^{wv}]' + \Sigma^w \\ P_0 &= O_n \end{aligned} \qquad t = 0,1,\ldots \tag{2.8}$$

and, for convenience, introduce the $\mathcal{Q}_k$-valued sequence $\{J_t,\ t = 0,1,\dots\}$, where

$$J_t := HP_tH' + \Sigma^v. \qquad t = 0,1,\dots \quad (2.9)$$

Let the deterministic sequences $\{Q_t^*,\ t = 0,1,\dots\}$ and $\{R_t^*,\ t = 0,1,\dots\}$ in $\mathcal{M}_{n\times n}$ and $\mathcal{Q}_n$, respectively, be defined recursively by

$$\begin{aligned} Q_{t+1}^* &= \left[A - [AP_tH' + \Sigma^{wv}]J_t^{-1}H\right] Q_t^* \\ Q_0^* &= I_n, \end{aligned} \qquad t = 0,1,\dots \quad (2.10)$$

and

$$\begin{aligned} R_{t+1}^* &= R_t^* + Q_t^{*\prime}H'J_t^{-1}HQ_t^* \\ R_0^* &= O_n. \end{aligned} \qquad t = 0,1,\dots \quad (2.11)$$

Then the representation

$$\epsilon_{t+1} = \int_{I\!R^n} \frac{\left\|Q_{t+1}^* \int_{I\!R^n} \left\{z - [R_{t+1}^* + \Delta^{-1}]^{-1}b\right\} \exp[z'b - \frac{1}{2}z'R_{t+1}^*z]dF(z)\right\|^2}{\int_{I\!R^n} \exp[z'b - \frac{1}{2}z'R_{t+1}^*z]dF(z)} dG_{R_{t+1}^*}(b) \quad (2.12)$$

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

In order to simplify the expression (2.12), we define the mapping $I_F : \mathcal{M}_{n\times n} \times \mathcal{Q}_n \to I\!R$ parameterized by the initial distribution F by setting

$$I_F(K,R) := \int_{I\!R^n} \frac{\left\|K \int_{I\!R^n} \left\{z - [R + \Delta^{-1}]^{-1}b\right\} \exp[z'b - \frac{1}{2}z'Rz]dF(z)\right\|^2}{\int_{I\!R^n} \exp[z'b - \frac{1}{2}z'Rz]dF(z)} dG_R(b) \quad (2.13)$$

for all K in $\mathcal{M}_{n\times n}$ and R in $\mathcal{Q}_n$. With this notation, (2.12) may be rewritten as

$$\epsilon_t = I_F(Q_t^*, R_t^*). \qquad t = 1,2,\dots \quad (2.14)$$

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^*.

Although (2.14) provides a simple representation for studying the asymptotic behavior of ϵ_t, we still must study the behavior of I_F under the *joint* asymptotic behavior of $\{Q_t^*,\ t = 0,1,\dots\}$ and $\{R_t^*,\ t = 0,1,\dots\}$. To that end, upon defining the mapping $I_F^* : \mathcal{Q}_n \to I\!R$ by

$$I_F^*(R) := I_F(I_n, R) \quad (2.15)$$

for all R in $\mathcal{Q}_n$, we observe the inequalities

$$\lambda_{\min}(Q_t^{*\prime}Q_t^*)I_F^*(R_t^*) \leq \epsilon_t((A,H,\Gamma),F) \leq \lambda_{\max}(Q_t^{*\prime}Q_t^*)I_F^*(R_t^*). \qquad t=1,2,\ldots \quad (2.16)$$

In effect, (2.16) shows that we may *separately* consider the asymptotic behavior of $\{Q_t^*,\ t=0,1,\ldots\}$ and the asymptotic behavior of I_F^* as $\{R_t^*,\ t=0,1,\ldots\}$ tends to its limit.

III. A STABILITY RESULT

We now commence our analysis of the asymptotic behavior of $\{\epsilon_t,\ t=0,1,\ldots\}$ in the general multivariable case. We focus our attention first on the asymptotics of $\{Q_t^*,\ t=0,1,\ldots\}$ and $\{R_t^*,\ t=0,1,\ldots\}$, and then study the behavior of I_F as Q_t^* and R_t^* asymptotically behave in a well-defined way. As a first step, we provide a stability criterion for the system (A,H,Γ) which is strong enough to ensure that $\lim_t \epsilon_t = 0$ for *any* initial distribution F in $\mathcal{E}_n$. If this stability criterion is satisfied, we may then also make several estimates of the *rate* at which ϵ_t tends to 0. Apart from being interesting from an operational viewpoint, these estimates on the rates of convergence provide an indirect characterization of F as follows: indeed they are *independent* of the initial distribution F when F is not Gaussian, so that if $\lim_t \epsilon_t = 0$ at a fast enough rate, then F *must* necessarily be Gaussian.

We first present some additional notation: We introduce the matrices $\overline{A}$ and $\overline{C}$ in $\mathcal{M}_{n\times n}$ and $\mathcal{Q}_n$ defined by

$$\overline{A} := A - \Sigma^{wv}(\Sigma^v)^{-1}H \quad (3.1)$$

and

$$\overline{C} := \Sigma^w - \Sigma^{wv}(\Sigma^v)^{-1}\Sigma^{vw}. \quad (3.2).$$

The matrices $\{K_t,\ t=0,1,\ldots\}$ in $\mathcal{M}_{n\times n}$ are now defined by

$$K_t := A - [AP_tH + \Sigma^{wv}]J_t^{-1}H \qquad t=0,1,\ldots \quad (3.3)$$

and we set $K_\infty := \lim_t K_t$ whenever this limit is well defined. With this notation, we may rewrite the recursion (2.10) as

$$Q_{t+1}^* = K_tQ_t^*. \qquad t=0,1,\ldots \quad (3.4)$$

The following stability criterion taken from [1] is used in what follows.

Theorem 2. *If the pair (A,H) is detectable and if the pair $(\overline{A},\overline{C})$ is controllable, then*

1. *The matrix* $P_\infty := \lim_t P_t$ *is well defined and positive definite, and*
2. *The matrix* $K_\infty := A - [AP_\infty H' + \Sigma^{wv}][HP_\infty H' + \Sigma^v]^{-1}H$ *is stable.*

Proof. It is not difficult to verify that

$$\Sigma^w - \Sigma^{wv}(\Sigma^v)^{-1}\Sigma^{vw} = E\Big[\big[W^\circ_{t+1} - E[W^\circ_{t+1}|V^\circ_{t+1}]\big]\big[W^\circ_{t+1} - E[W^\circ_{t+1}|V^\circ_{t+1}]\big]'\Big] \tag{3.5}$$

so that the matrix $\Sigma^w - \Sigma^{wv}(\Sigma^v)^{-1}\Sigma^{vw}$ is symmetric non-negative definite, and its square root is well defined [2, Secs. VIII.6 and VIII.7]. Claim 1 is Appendix 1 in [1] and claim 2 is Theorem 5.1 in [1]. ∎

Because of its importance, we list the assumption of Theorem 2 as the following key condition (C.1), where

(C.1): The pair (A, H) is detectable and the pair $(\overline{A}, \overline{C})$ is controllable.

Theorem 2 implies the following results concerning $\{Q^*_t,\ t = 0, 1, \ldots\}$ and $\{R^*_t,\ t = 0, 1, \ldots\}$. We first observe from (2.11) that $0 \le R^*_t \le R^*_{t+1}$ for all $t = 0, 1, \ldots$. Consequently, $R^*_\infty; = \lim_t R^*_t$ is always well defined and non-negative definite, although possibly infinite, with

$$R^*_t = \sum_{s=0}^{t-1} Q^{*\prime}_s H'[HP_sH' + \Sigma^v]^{-1}HQ^*_s. \qquad t = 1, 2, \ldots \tag{3.7}$$

Theorem 3. *Assume the criterion* **(C.1)** *to be satisfied.*

1. *We have* $\lim_t Q^*_t = 0$ *with*

$$\limsup_t \frac{1}{t} \ln \lambda_{\max}(Q^{*\prime}_t Q^*_t) \le 2 \ln \lambda_{\max}(K_\infty) < 0, \tag{3.8}$$

and if the matrix K_t *is invertible for each* $t = 0, 1, \ldots$, *then*

$$\liminf_t \frac{1}{t} \ln \lambda_{\min}(Q^{*\prime}_t Q^*_t) \ge 2 \ln \lambda_{\min}(K_\infty). \tag{3.9}$$

2. *Moreover,* $R^*_\infty := \lim_t R^*_t$ *is well defined and finite.*

Proof. By Theorem 2, $K_\infty = \lim_t K_t$ exists and is stable if **(C.1)** is satisfied. Claim 1 is now a consequence of the stability of K_∞ and Appendix B of [3]. To obtain the second claim, we note from (3.7) that

$$0 \le R^*_\infty \le \frac{1}{\lambda_{\min}(\Sigma^v)} \sum_{t=0}^{\infty} (HQ^*_t)'(HQ^*_t) \tag{3.10}$$

and the finiteness of R^*_∞ follows from Claim 1. ∎

For ease of exposition, we define the mapping $\phi : I\!R^n \times I\!R^n \times \mathcal{Q}_n \to R_+ \cup \{\infty\}$ by

$$\phi(z,b;R) := \exp[z'b - \frac{1}{2}z'Rz] \tag{3.11}$$

for z and b in $I\!R^n$ and R in $\mathcal{Q}_n$, and we set

$$\Phi(b;R) := \int_{I\!R^n} \phi(z,b;R)dF(z). \tag{3.12}$$

A family of probability measures $\tilde{F}_{b,R}$ on $I\!R^n$, parametrized by b in $I\!R^n$ and R in $\mathcal{Q}_n$, is now introduced. Each probability measure in this family is absolutely continuous with respect to F with Radon-Nikodym derivative given by

$$\frac{d\tilde{F}_{b,R}}{dF}(z) := \begin{cases} \phi(z,b;R)/\Phi(b;R) & \text{if } \Phi(b;R) < \infty \\ 1 & \text{if } \Phi(b;R) = \infty \end{cases} \tag{3.13}$$

for all z in $I\!R^n$. With this notation, the function I_F can be expressed as

$$I_F(K,R) = \int_{I\!R^n} \left\| K \int_{I\!R^n} \left\{ z - [R + \Delta^{-1}]^{-1} b \right\} d\tilde{F}_{b,R}(z) \right\|^2 \Phi(b;R) dG_R(b) \tag{3.14}$$

for all K in $\mathcal{M}_{n\times n}$ and R in $\mathcal{Q}_n$, a form more manageable for our calculations.

Our first observation is contained in

Proposition 1. *For every distribution F in $\mathcal{D}_n$, we have* $\limsup_t I^*_F(R^*_t) < \infty$.

Proof. From Jensen's inequality, we conclude that

$$I^*_F(R^*_t) \le J_F(R^*_t) \qquad\qquad t = 1,2,\ldots \tag{3.15}$$

with

$$J_F(R) := \int_{I\!R^n} \int_{I\!R^n} \left\| z - [R + \Delta^{-1}]^{-1} b \right\|^2 d\tilde{F}_{b,R}(z) \Phi(b;R) dG_R(b). \tag{3.16}$$

for all R in $\mathcal{Q}_n$. The definition of $\tilde{F}_{b,R}$ and Tonelli's theorem imply

$$J_F(R) = \int_{I\!R^n} \left[\int_{I\!R^n} \left\| z - [R + \Delta^{-1}]^{-1} b \right\|^2 \exp[z'b] dG_R(b) \right] \exp[-\frac{1}{2}z'Rz] dF(z), \tag{3.17}$$

where the inner integral may be directly evaluated by using standard results on Gaussian RV's. After some tedious calculations, we find that

$$J_F(R) = \mathrm{tr}\left\{[R+\Delta^{-1}]^{-1}R[R+\Delta^{-1}]^{-1}\right\} + \int_{IR^n} z'\Delta^{-1}[R+\Delta^{-1}]^{-1}[R+\Delta^{-1}]^{-1}\Delta^{-1}z\,dF(z) \tag{3.18}$$

for every R in $\mathcal{Q}_n$. Since R_t^* is nonnegative-definite and $R_{t+1}^* \geq R_t^*$ for $t = 0, 1, \ldots$, we have that

$$\limsup_t J_F(R_t^*) < \infty \tag{3.19}$$

which, together with (3.15), concludes the proof. ∎

Note that in Proposition 1, we did not impose the requirement that R_∞^* be finite.

Collecting what we have discovered so far, we obtain the following result.

Theorem 4. *Assume the condition* **(C.1)** *to hold. For any square-integrable distribution F,*

$$\lim_t \epsilon_t((A,H,\Gamma),F) = 0 \tag{3.20a}$$

and

$$\limsup_t \frac{1}{t}\log \epsilon_t((A,H,\Gamma),F) \leq 2\log \lambda_{\max}(K_\infty) < 0. \tag{3.20b}$$

Proof. It suffices only to show (3.20b) since it implies (3.20a). For distributions F in $\mathcal{D}_n$, (3.20b) follows immediately from (2.16), Theorem 3 and Proposition 1. To extend the results to distributions F in $\mathcal{E}_n$, we use the transformation (2.6)-(2.7). ∎

Whereas Theorem 4 establishes an upper bound on the rate at which ϵ_t decays to 0 if **(C.1)** is satisfied, we now show lower bounds for this same rate. These lower estimates require the following condition **(C.2)** on the nonnegative-definite matrix R_∞^*, namely

(C.2): The matrix R_∞^* is positive definite.

We then have the following proposition.

Proposition 2. *If the distribution F is in $\mathcal{D}_n$ and $0 < R_\infty^* < \infty$, then F is necessarily Gaussian if* $\liminf_t I_F^*(R_t^*) = 0$.

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

$$\frac{d\hat{F}}{dF}(z) = \frac{\exp\left[-\frac{1}{2}z'R_\infty^* z\right]}{\int_{IR^n}\exp\left[-\frac{1}{2}z'R_\infty^* z\right]dF(z)}, \quad z \in IR^n. \tag{3.21}$$

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

$$N(b) := \int_{IR^n} \exp[z'b] d\hat{F}(z), \quad b \in IR^n. \tag{3.22}$$

We show that if $\liminf_t I_F^*(R_t^*) = 0$, then N must satisfy the conditions

$$\begin{aligned} \nabla_b N(b) &= \left[R_\infty^* + \Delta^{-1}\right]^{-1} bN(b) \\ N(0) &= 1 \end{aligned} \tag{3.23}$$

on IR^n, from which we conclude that the distribution $\hat{F}$ is Gaussian; we shall use (3.21) to verify that then F must also be Gaussian.

Since the matrix R_∞^* is positive definite, there exists a finite T such that for $t = T, T+1, \ldots$ the matrix R_t^* is also positive definite and thus $G_{R_t^*}$ is absolutely continuous with respect to the Lebesgue measure ν on IR^n. Applying Fatou's Lemma to (2.14), we see from the assumption $\liminf_t I_F^*(R_t^*) = 0$ that

$$\liminf_t \frac{\left\| \int_{IR^n} \left\{ z - [R_t^* + \Delta^{-1}]^{-1} b \right\} \phi(z, b; R_t^*) dF(z) \right\|^2}{\Phi(b; R_t^*)} \frac{dG_{R_t^*}}{d\nu}(b) = 0. \tag{3.24}$$

If $0 < R_\infty^* < \infty$, we see that for all b in IR^n,

$$\lim_t \frac{dG_{R_t^*}}{d\nu}(b) = \frac{dG_{R_\infty^*}}{d\nu}(b) > 0 \tag{3.25}$$

and

$$\lim_t \Phi(b; R_t^*) = \Phi(b, R_\infty^*) > 0 \tag{3.26}$$

with the last following by monotone convergence. Combining (3.25)–(3.26), we now conclude that

$$\liminf_t \left\| \int_{IR^n} \left\{ z - [R_t^* + \Delta^{-1}]^{-1} b \right\} \phi(z, b; R_t^*) dF(z) \right\| = 0 \tag{3.27}$$

for ν-almost every b, or equivalently

$$\int_{IR^n} z\phi(z, b; R_\infty^*) dF(z) = [R_\infty^* + \Delta^{-1}]^{-1} b \int_{IR^n} \phi(z, b; R_\infty^*) dF(z). \tag{3.28}$$

Upon dividing (3.28) by $\int_{IR^n} \exp\left[-\frac{1}{2} z' R_\infty^* z\right] dF(z)$, we obtain (3.23); the technical details are found in [3].

The unique solution of (3.23) is

$$N(b) = \exp\left[\frac{1}{2}b'\left[R_\infty^* + \Delta^{-1}\right]^{-1}b\right], \quad b \in IR^n \tag{3.29}$$

so that $\hat{F}$ is Gaussian with mean 0 and variance $[R_\infty^* + \Delta^{-1}]^{-1}$. Since the variance of $\hat{F}$ is positive definite, we see that $\hat{F}$ is absolutely continuous with respect to ν and therefore F must be absolutely continuous with respect to ν by virtue of the mutual absolute continuity of F and $\hat{F}$. We calculate the density of F with respect to ν by the relation

$$\frac{dF}{d\nu} = \frac{dF}{d\hat{F}} \cdot \frac{d\hat{F}}{\nu} \tag{3.30}$$

and find after some arithmetic that

$$\frac{dF}{d\nu_n}(z) = c\exp\left[-\frac{1}{2}z'\Delta^{-1}z\right], \quad z \in IR^n \tag{3.31}$$

i.e., that the distribution F is Gaussian. ∎

The following is an immediate result of this proposition.

Theorem 5. *If the assumptions* **(C.1)** *and* **(C.2)** *are satisfied, and* $\lambda_{\min}(Q_t^{*\prime}Q_t^*) > 0$ *for all* t *sufficiently large, then the distribution* F *is in fact Gaussian if*

$$\lim_t \frac{\epsilon_t((A,H,\Gamma),F)}{\lambda_{\min}(Q_t^{*\prime}Q_t^*)} = 0. \tag{3.32}$$

Proof. If the distribution F in in $\mathcal{D}_n$ and (3.32) holds, then from the lower bound of (2.16) we see that $\liminf_t I_F^*(R_t^*) = 0$, so necessarily F must be Gaussian in view of Proposition 2. The transformations (2.6) and (2.7) allow us to establish the result for distributions in $\mathcal{E}_n$. ∎

In a similar manner, we may verify a lower bound analogous to the upper bound of Theorem 4.

Theorem 5. *If the assumptions* **(C.1)** *and* **(C.2)** *are satisfied and the matrices* $\{K_t,\ t = 0,1,\ldots\}$ *are invertible, then*

$$\liminf_t \frac{1}{t}\ln\epsilon_t((A,H,\Gamma),F) \geq 2\ln\lambda_{\min}(K_\infty) \tag{3.33}$$

for all non-Gaussian distributions F *in* $\mathcal{E}_n$.

IV. THE SCALAR CASE

We now turn to the case where $n = k = 1$. Recall the following standard definition from control theory [5, Def. 6.5]:

A system (A, H) is said to be *stabilizable* if all unstable modes are in the controllable subspace.

and make the following definition:

A system (A, H) is said to be *marginally stabilizable* if all modes which are not either stable or critically stable are in the controllable subspace.

Our goal in this section is to verify the following claim.

Theorem 7. *Assume $n = k = 1$. We have the following convergence results:*

1. *If the pair $(\bar{A}, \bar{C})$ is marginally stabilizable, $\lim_t \epsilon_t = 0$ for any distribution F in $\mathcal{E}_1$, whereas if the pair $(\bar{A}, \bar{C})$ is not marginally stabilizable, then the asymptotic behavior of ϵ_t depends nontrivially upon F in $\mathcal{E}_1$.*

Moreover we also have the following estimates:

2. *If $(\bar{A}, \bar{C})$ is stabilizable, then $\lim_t \epsilon_t = 0$ at an exponential rate independent of F for F in $\mathcal{E}_1$ non-Gaussian whereas if the pair $(\bar{A}, \bar{C})$ is marginally stabilizable but not stabilizable, then the rate depends non-trivially upon F.*

We shall prove these results by considering a number of cases. Since we are working in IR, we may rewrite (2.8), (2.9) and (2.11) as

$$P_{t+1} = A^2 P_t - \frac{(AP_tH + \Sigma^{wv})^2}{H^2 P_t + \Sigma^v} + \Sigma^w \qquad t = 0, 1, \ldots \quad (4.1)$$
$$P_0 = 0,$$

$$Q^*_{t+1} = \left(\frac{A\Sigma^v - \Sigma^{wv} H}{H^2 P_t + \Sigma^v} \right) Q^*_t \qquad t = 0, 1, \ldots \quad (4.2)$$
$$Q^*_0 = 1,$$

and

$$R^*_{t+1} = R^*_t + \frac{(Q^*_t)^2 H^2}{H^2 P_t + \Sigma^v} \qquad t = 0, 1, \ldots \quad (4.3)$$
$$R^*_0 = 0.$$

Note also that we have

$$\epsilon_t = (Q_t^*)^2 I_F^*(R_t^*) \qquad t = 1,2,\ldots \quad (4.4)$$

for F in $\mathcal{D}_1$.

We first observe a degeneracy when $H = 0$.

Proposition 3. *If $H = 0$, then $\epsilon_t = 0$ for all $t = 1,2,\ldots$ and all distributions F in $\mathcal{E}_1$.*

Proof. If $H = 0$, then $R_t^* = 0$ for all $t = 0,1,\ldots$, so $\epsilon_t = 0$ for all $t = 1,2,\ldots$ for all F in $\mathcal{D}_1$ by directly evaluating (2.12); by translation the result is true for all F in $\mathcal{E}_1$. ∎

We could prove Proposition 3 more directly in the case where $\Sigma^{wv} = 0$, for then the sequences $\{X_t^\circ,\ t = 0,1,\ldots\}$ and $\{Y_t,\ t = 0,1,\ldots\}$ are independent, so the MMSE and LMSE filters coincide. We now consider the more interesting case when $H \neq 0$. Note from (3.2) that $(\bar{A}, \bar{C})$ is controllable if and only if $\bar{C} \neq 0$, i.e., if and only if $\Sigma^v \Sigma^w \neq (\Sigma^{wv})^2$. We have

Proposition 4. *If $H \neq 0$ and $(\bar{A}, \bar{C})$ is controllable, then $\lim_t \epsilon_t = 0$ for all distributions F in $\mathcal{E}_1$. If $\bar{A} = 0$, then $\epsilon_t = 0$ for all t and all F in $\mathcal{E}_1$, whereas if $\bar{A} \neq 0$, then*

$$\lim_t \frac{1}{t} \ln \epsilon_t = 2 \ln \left| \bar{A} \left(\frac{\Sigma^v}{H^2 P_\infty + \Sigma^v} \right) \right| < 0 \qquad (4.5)$$

for all non-Gaussian distributions F in $\mathcal{E}_1$.

Proof. If $\bar{A} = 0$, then $Q_t^* = 0$ for all $t = 0,1,\ldots$ by (4.2). We see from (4.4) that $\epsilon_t = 0$ for all $t = 1,2,\ldots$ and all F in $\mathcal{D}_1$, whence $\epsilon_t = 0$ for all $t = 1,2,\ldots$ and all F in $\mathcal{E}_1$. If $\bar{A} \neq 0$, then $K_t \neq 0$ and $Q_t^* \neq 0$ for all $t = 0,1,\ldots$, so $R_\infty^* > 0$ from (4.3) and we may apply Theorems 4 and 6 to verify (4.5). ∎

We next consider the case when $(\bar{A}, \bar{C})$ is stabilizable but uncontrollable. We can quickly verify by induction on t that when $\bar{C} = 0$, $P_t = 0$ and $Q_t^* = \bar{A}^t$ for all $t = 0,1,\ldots$.

Proposition 5. *If $H = 0$ and $(\bar{A}, \bar{C})$ is stabilizable but uncontrollable, i.e., $|\bar{A}| < 1$ and $\bar{C} = 0$, then $\lim_t \epsilon_t$ tends to zero with*

$$\lim_t \frac{1}{t} \ln \epsilon = 2 \ln |\bar{A}| < 0. \qquad (4.6)$$

Proof. Since $|\bar{A}| < 1$, we have from (4.3) that $0 < R_\infty^* < \infty$. In view of Propositions 1 and 2, we have that

$$0 < \liminf_t I_F^*(R_t^*) \leq \limsup_t I_F^*(R_t^*) < \infty; \qquad (4.7)$$

we then arrive at (4.6) by means of (4.4). ∎

Turning now to the case where $(\bar{A}, \bar{C})$ is not stabilizable, we shall prove the dependencies given in Theorem 7 by analyzing the asymptotics of ϵ_t for two specific initial distributions. First, however, let us verify a general result.

Proposition 6. *For any distribution F in $\mathcal{D}_1$, $\limsup_t tI_F^*(t) < \infty$ and therefore $\lim_t I_F^*(t) = 0$.*

Proof. Since the functional I_F^* is *independent* of the system dynamics (A, H, Γ), we may assume for the purpose of argumentation that our system is

$$\begin{aligned} X_t^\circ &= \xi \\ Y_t &= \xi + V_{t+1}^\circ. \end{aligned} \qquad t = 0, 1, \ldots \quad (4.8)$$

Here $A = H = \Sigma^v = 1$ and $\Sigma^w = \Sigma^{wv} = 0$. For this system (which we note to be marginally stabilizable), $Q_t^* = 1$ and $R_t^* = t$ for all $t = 1, 2, \ldots$, so $\epsilon_t = I_F^*(t)$ for $t = 0, 1, \ldots$. For all $t = 0, 1, \ldots$, define the linear estimate $\check{X}_t$ of X_t° on the basis of $\{Y_0, \ldots, Y_t\}$ to be

$$\begin{aligned} \check{X}_{t+1} &:= \frac{1}{t+1} \sum_{s=0}^{t} Y_s \\ \check{X}_0^\circ &:= 0. \end{aligned} \qquad t = 0, 1, \ldots \quad (4.9)$$

Using the facts that $\check{X}_t$ is a *linear* estimator, that $\hat{X}_t^K$ is the LMSE estimator, and that $\hat{X}_t$ is the MMSE estimator, we have

$$\begin{aligned} \|\hat{X}_t - \hat{X}_t^K\|_\Omega &\le \|\hat{X}_t - X_t^\circ\|_\Omega + \|\hat{X}_t^K - X_t^\circ\|_\Omega \\ &\le \|\check{X}_t - X_t^\circ\|_\Omega + \|\hat{x}_t^K - x_t^\circ\|_\Omega. \\ &= 2\|\check{X}_t - X_t^\circ\|_\Omega, \end{aligned} \quad (4.10)$$

where $\|X\|_\Omega := \left[E(X^2)\right]^{1/2}$ for any square-integrable RV X. From (4.10),

$$I_F^*(t) = \epsilon_t \le 4E\left[\left|\frac{1}{t+1}\sum_{s=0}^{t} V_{s+1}^\circ\right|^2\right] = \frac{4}{t+1} \qquad t = 0, 1, \ldots \quad (4.11)$$

and the claim is now immediate. ∎

We shall now consider two distributions F_1 and F_2 in $\mathcal{D}_1$.

Distribution F_1. Distribution F_1 admits a density with respect to Lebesgue measure λ on IR given by

$$\frac{dF_1}{d\lambda}(z) = \sum_{i=1}^{n} \alpha_i \frac{1}{\sqrt{2\pi\rho}} \exp\left[-\frac{1}{2}\frac{(z-\mu_i)^2}{\rho^2}\right] \qquad z \in IR \quad (4.12)$$

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

Distribution F_2. Under F_2, the RV ξ takes on a finite number of values $z_1 < z_2 \ldots < z_n$ with probabilities $p_1, p_2, \ldots, p_n$ respectively with $\sum_{i=1}^{n} p_i z_i = 0$.

The following two facts are proved in [3].

Fact 1. *We have*

$$I^*_{F_1}(t) = \frac{K + o_1(\frac{1}{t})}{(\rho^2 t + 1)^2}, \quad t > 0 \qquad (4.13)$$

where $\lim_t o_1(\frac{1}{t}) = 0$ *and* $K > 0$.

and

Fact 2. *We also have*

$$I^*_{F_2}(t) = \frac{1 + o_1(\frac{1}{t})}{t} \quad t > 0 \qquad (4.14)$$

where $\lim_t o_1(\frac{1}{t}) = 0$.

We now can prove the rest of Theorem 7.

Proposition 7. *If* $H \neq 0$ *and* $(\bar{A}, \bar{C})$ *is marginally stabilizable but not stabilizable, i.e.,* $|\bar{A}| = 1$ *and* $\bar{C} = 0$, *then* $\lim_t \epsilon_t = 0$ *for any distribution* F *in* $\mathcal{E}_1$, *but the rate of this convergence depends nontrivially upon* F *for* F *non-Gaussian.*

Proof. We have under the hypothesis that $\epsilon_t = (1)^t I^*_F(t)$ for all $t = 0, 1, \ldots$ and all F in $\mathcal{D}_1$, the extension to $\mathcal{E}_1$ being as before. By Proposition 6, $\lim_t \epsilon_t = 0$; however, if $F = F_1$, $\lim_t \ln(\epsilon_t/(\ln t)) = -2$, whereas if $F = F_2$, $\lim_t \ln(\epsilon_t/(\ln t)) = -1$. ∎

Finally, we conclude with

Proposition 8. $H \neq 0$ *and* $(\bar{A}, \bar{C})$ *is not marginally stabilizable, i.e.,* $|\bar{A}| > 1$ *and* $\bar{C} = 0$, *then* $\limsup_t \epsilon_t < \infty$ *for all distributions* F *in* $\mathcal{E}_1$, *the asymptotic behavior depending nontrivially upon* F *for* F *not Gaussian.*

Proof. It is easy to verify that under the hypotheses on (A, H, Γ), $\lim_t R_t^* = \infty$ but $\lim_t (Q_t^*)^2/R_t^* = \Sigma^v(\bar{A}^2 - 1)/H^2$. For F in $\mathcal{D}_1$, then

$$\epsilon_t = \frac{(Q_t^*)^2}{R_t^*}\,(R_t^* I_F^*(R_t^*)). \qquad t = 1,2,\ldots \quad (4.15)$$

Applying Proposition 6, we get $\limsup_t \epsilon_t < \infty$ for all F in $\mathcal{D}_1$, and thus for all distributions F in $\mathcal{E}_1$. However, if $F = F_1$, $\lim_t \epsilon_t = 0$, whereas if $F = F_2$, then $\lim_t \epsilon_t = 1$. ∎

The proof of Theorem 7 is complete; all the cases have been considered.

REFERENCES

[1] P. Caines and D. Mayne, "On the discrete time Ricatti equation of optimal control," *Inter. J. of Control*, vol. 12, pp. 785–794 (1970) and vol. 14, pp. 205–207 (1971).

[2] F. Gantmacher, *The Theory of Matrices*, Chelsea Publishing Co., New York, 1977.

[3] R. B. Sowers, *New Discrete-Time Filtering Results*, M.S. Thesis, Electrical Engineering Department, University of Maryland, College Park (MD), August 1988.

[4] R. B. Sowers and A.M. Makowski, "Discrete-Time Filtering For Linear Systems in Correlated Noise With Non-Gaussian Initial Conditions," Systems Research Center Technical Report TR 88-85 (1988).

[5] H. Kwakernaak and R. Sivan, *Linear Optimal Control Systems*, John Wiley & Sons, 1972.