DISCRETE-TIME FILTERING FOR LINEAR SYSTEMS IN CORRELATED NOISE WITH NON-GAUSSIAN INITIAL CONDITIONS: FORMULAS AND ASYMPTOTICS

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 noises, and non-Gaussian initial conditions. Explicit representations are obtained for the MMSE and LMMSE (or Kalman) estimates of the state given past observations, as well as for the expected square of their difference. These formulae are obtained with the help of the Girsanov transformation for Gaussian white noise sequences, and display explicitly the dependence of the quantities of interest on the initial distribution. With the help of these formulae, we completely characterize the asymptotic behavior of the error sequence in the scalar time-invariant case.

I. INTRODUCTION

We consider the one-step prediction problem associated with the stochastic discrete-time linear dynamical system

$$\begin{aligned} X^\circ_{t+1} &= A_t X^\circ_t + W^\circ_{t+1} \\ X^\circ_0 &= \xi \qquad\qquad\qquad\qquad t = 0,1,\ldots \qquad (1.1) \\ Y_t &= H_t X^\circ_t + V^\circ_{t+1} \end{aligned}$$

defined on some probability triple $(\Omega, \mathcal{F}, P)$ which carries the $I\!R^n$-valued plant process $\{X^\circ_t,\ t = 0,1,\ldots\}$ and the $I\!R^k$-valued observation process $\{Y_t,\ t = 0,1,\ldots\}$. Here, for all $t = 0,1,\ldots$, the matrices A_t and H_t are of dimension $n \times n$ and $k \times n$, respectively. Throughout we make the following assumptions (A.1)-(A.3), where

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

$$\Gamma_{t+1} := \mathrm{Cov}\begin{pmatrix} W^\circ_{t+1} \\ V^\circ_{t+1} \end{pmatrix} = \begin{pmatrix} \Sigma^w_{t+1} & \Sigma^{wv}_{t+1} \\ \Sigma^{vw}_{t+1} & \Sigma^v_{t+1} \end{pmatrix}; \qquad t = 0,1,\ldots \qquad (1.2)$$

(A.2): For all $t = 0,1,\ldots$, the covariance matrix Σ^v_{t+1} is positive definite; and

(A.3): The initial condition ξ has distribution F with finite first and second moments μ and Δ, respectively, and is independent of the process $\{(W^\circ_{t+1}, V^\circ_{t+1}),\ t = 0,1,\ldots\}$. No other *a priori* assumptions are enforced on F.

The (one-step) prediction problem associated with (1.1) is defined as the problem of evaluating the conditional expectation

$$E[\phi(X^\circ_{t+1})|\ Y_0,\ldots,Y_t] \qquad t = 0,1,\ldots (1.3)$$

in *Robust Control of Linear Systems and Nonlinear Control*, Eds. E.M.A. Kaashoek, J.H. van Schuppen, and A.C.M. Ran, Birkhäuser Boston, Inc., Cambridge (MA), Proceedings of the International Symposium on the Mathematical Theory of Networks and Systems, Amsterdam (The Netherlands), June 1989, pp. 407—419.

for all bounded Borel mappings $\phi : I\!R^n \to \mathcal{C}$, with $\mathcal{C}$ denoting set of the complex numbers. In this paper, we solve the prediction problem (1.3) associated with (1.1)-(1.2). For each $t = 0, 1, \ldots$, once the conditional distribution of X°_{t+1} given $\{Y_0, \ldots, Y_t\}$ is available, it is possible to construct the MMSE estimate $\hat{X}_{t+1} := E[X^\circ_{t+1} | \, Y_0, \ldots, Y_t]$ of X°_{t+1} on the basis of $\{Y_0, \ldots, Y_t\}$. In general, $\hat{X}_{t+1}$ is a *non*-linear function of $\{Y_0, \ldots, Y_t\}$, in contrast with the LLMSE (or Kalman) estimate $\hat{X}^K_{t+1}$ of X°_{t+1} computed on the basis of $\{Y_0, \ldots, Y_t\}$, which is by definition linear in these quantities. We shall find representations for both $\{\hat{X}_t, \ t = 0, 1, \ldots\}$ and $\{\hat{X}^K_t, \ t = 0, 1, \ldots\}$, and then form the mean square error $\epsilon_t := E[\|\hat{X}_t - \hat{X}^K_t\|^2]$ for $t = 1, 2, \ldots$.

When the plant and observation noises are *uncorrelated*, and the observation noise sequence $\{V_t, \ t = 0, 1, \ldots\}$ is standard (i.e., $\Sigma^{wv}_{t+1} = 0$ and $\Sigma^v_{t+1} = I_n$ for all $t = 0, 1, \ldots$), the prediction problem posed above is the discrete-time counterpart of the situation investigated in [3]. In Section II we state the main results for the nonlinear prediction problem, and outline the proofs in Section III, thus indicating how the technique of [3] extends to the correlated noise situation without major difficulties. We then use these results in Section IV to obtain representations for $\{\hat{X}_t, \ t = 1, 2, \ldots\}$, $\{\hat{X}^K_t, \ t = 1, 2, \ldots\}$, and $\{\epsilon_t, \ t = 1, 2, \ldots\}$. These expressions explicitly display the dependence of the initial distribution F, and form the basis for the large time asymptotic analysis carried out in [6] on the error terms $\{\epsilon_t, \ t = 1, 2, \ldots\}$. In Section V we consider this asymptotic behavior in the scalar time-invariant case, and give a complete characterization of these asymptotics in terms of the plant gain a, the observation gain h, and the noise covariance matrix Γ. Many details have been omitted for the sake of brevity; additional information and material can be found in the thesis [4] and in [5].

II. THE FILTERING PROBLEM

II.1.The notation

A word on the notation: For any positive integers n and m, we denote the space of $n \times m$ real matrices by $\mathcal{M}_{n\times m}$ and the cone of $n \times n$ symmetric positive-definite matrices by $\mathcal{Q}_n$. As in [3], for every Σ in $\mathcal{Q}_{2n}$, let X_Σ and B_Σ denote generic $I\!R^n$-valued random variables (RVs) such that (X_Σ, B_Σ) is a $I\!R^{2n}$-valued zero-mean Gaussian RV with covariance matrix Σ. For every bounded Borel mapping $\phi : I\!R^n \to \mathcal{C}$, we define the mappings $\mathcal{T}\phi : I\!R^n \times I\!R^n \times \mathcal{Q}_{2n} \to \mathcal{C}$ and $\mathcal{U}\phi : I\!R^n \times I\!R^n \times \mathcal{Q}_n \times \mathcal{M}_{n\times n} \times \mathcal{Q}_{2n} \to \mathcal{C}$ by

$$\mathcal{T}\phi[x, b; \Sigma] := \mathcal{E}\left[\phi(x + X_\Sigma) \exp[b' B_\Sigma]\right] \tag{1.4}$$

and

$$\mathcal{U}\phi[x, b; \Lambda, \Psi; \Sigma] := \mathcal{E}\left[\mathcal{T}\phi[x + \Psi\xi, \xi; \Sigma] \exp[b'\xi - \frac{1}{2}\xi'\Lambda\xi]\right] \tag{1.5}$$

with $\mathcal{E}$ denoting integration with respect to the Gaussian distribution of the RV (X_Σ, B_Σ).

Throughout, let I_n denote the unit matrix in $\mathcal{M}_{n\times n}$, and let O_n denote the zero element in $\mathcal{M}_{n\times n}$, i.e., the $n \times n$ matrix whose elements are all zero. Elements of $I\!R^n$ are always interpreted as column vectors, and transposition is denoted by $'$. Finally, let $\Psi(\cdot, \cdot)$ be the state transition matrix given by

$$\begin{aligned} \Psi(t,t) &= I_n \\ \Psi(s+1, t) &= [A_s - \Sigma^{wv}_{s+1}(\Sigma^v_{s+1})^{-1} H_s]\Psi(s,t). \qquad s = t, t+1, \ldots \end{aligned} \qquad t = 0, 1, \ldots \tag{1.6}$$

II.2. The main results

We define the $\mathcal{Q}_n$-valued sequence $\{P_t, \ t = 0, 1, \ldots\}$ by the recursions

$$\begin{aligned} P_{t+1} &= A_t P_t A_t' - [A_t P_t H_t' + \Sigma^{wv}_{t+1}][H_t P_t H_t' + \Sigma^v_{t+1}]^{-1}[A_t P_t H_t' + \Sigma^{wv}_{t+1}]' + \Sigma^w_{t+1} \\ P_0 &= O_n \end{aligned} \qquad t = 0, 1, \ldots \tag{2.1}$$

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

$$J_t := H_t P_t H_t' + \Sigma^v_{t+1}. \qquad t = 0, 1, \ldots \tag{2.2}$$

The deterministic sequences $\{Q_t,\ t=0,1,\ldots\}$ and $\{R_t,\ t=0,1,\ldots\}$ take values in $\mathcal{M}_{n\times n}$ and $\mathcal{Q}_n$, respectively, and are defined recursively by

$$\begin{aligned} Q_{t+1} &= A_tQ_t - [A_tP_tH_t' + \Sigma_{t+1}^{wv}]J_t^{-1}H_t\big(Q_t+\Psi(t,0)\big) + \Sigma_{t+1}^{wv}(\Sigma_{t+1}^{v})^{-1}H_t\Psi(t,0) \\ R_{t+1} &= R_t - \big(Q_t+\Psi(t,0)\big)'H_t'J_t^{-1}H_t\big(Q_t+\Psi(t,0)\big) + \Psi'(t,0)H_t'H_t\Psi(t,0) \end{aligned} \qquad t=0,1,\ldots \quad (2.3)$$

with initial conditions $Q_0=R_0=O_n$. From these sequences, we form the $\mathcal{Q}_{2n}$-valued sequence $\{\Sigma_t,\ t=0,1,\ldots\}$ by setting

$$\Sigma_t = \begin{pmatrix} P_t & Q_t \\ Q_t' & R_t \end{pmatrix}. \qquad t=0,1,\ldots \quad (2.4)$$

We also generate the $I\!R^n$-valued processes $\{\bar{X}_t, t=0,1,\ldots\}$ and $\{\bar{B}_t, t=0,1,\ldots\}$ via the recursive relations

$$\begin{aligned} \bar{X}_{t+1} &= \big[A_t - [A_tP_tH_t'+\Sigma_{t+1}^{wv}]J_t^{-1}H_t\big]\bar{X}_t + [A_tP_tH_t'+\Sigma_{t+1}^{wv}]J_t^{-1}Y_t \\ \bar{B}_{t+1} &= \bar{B}_t - \big(Q_t+\Psi(t,0)\big)'H_t'J_t^{-1}H_t\bar{X}_t + \big(Q_t+\Psi(t,0)\big)'H_t'J_t^{-1}Y_t \end{aligned} \qquad t=0,1,\ldots \quad (2.5)$$

with initial values $\bar{X}_0=\bar{B}_0=0$.

The solution to the prediction problem associated with (1.1) can now be given. On Ω define the filtration $\{\mathcal{Y}_t,\ t=0,1,\ldots\}$ generated by the observations $\{Y_t,\ t=0,1,\ldots\}$, i.e.,

$$\mathcal{Y}_t := \sigma\{Y_0,Y_1,\ldots,Y_t\}. \qquad t=0,1,\ldots \quad (2.6)$$

Theorem 1. *For any bounded Borel mapping $\phi: I\!R^n \to I\!R$ and any $t=0,1,\ldots$, the relationship*

$$E[\phi(X_{t+1}^\circ)|\mathcal{Y}_t] = \frac{U\phi[\bar{X}_{t+1},\bar{B}_{t+1};M_{t+1},\Psi(t+1,0);\Sigma_{t+1}]}{U\mathbb{I}[\bar{X}_{t+1},\bar{B}_{t+1};M_{t+1},\Psi(t+1,0);\Sigma_{t+1}]} \quad P-a.s. \qquad (2.7)$$

holds true, where $\mathbb{I}$ denotes the constant mapping $I\!R^n \to I\!R : x \to 1$ and the $\mathcal{Q}_n$-valued sequence $\{M_t,\ t=0,1,\ldots\}$ is defined recursively by

$$M_{t+1} = M_t + \Psi(t,0)'H_t'(\Sigma_{t+1})^{-1}H_t\Psi(t,0) \qquad t=0,1,\ldots \quad (2.8)$$

with $M_0=O_n$

The probabilistic interpretation as covariance matrices for $\{M_t,\ t=0,1,\ldots\}$ is available in [4, 5]. We readily see that the structure of the predictor in the general situation is not markedly different from that for the uncorrelated case [3]. The noise correlation is encoded in the *universal* sufficient statistics [3] that parametrize the predictor, but does *not* affect the *form* of the statistics bearing functionals.

III. PROOFS

Only the structure of the proof is outlined as details are available in the thesis [4] and in [5]. The approach used here extends the one introduced in [3], and is again based on finding a probability measure $\bar{P}$, absolutely continuous with respect to the original measure P on $\mathcal{F}$, under which the statistical calculations are readily performed. Here, as explained in [4], the *arbitrary* covariance structure of the plant and observation noise sequences leads to the use of a Girsanov transformation on the *joint* $I\!R^{n+k}$-valued sequence $\{(W_{t+1}^\circ, V_{t+1}^\circ),\ t=0,1,\ldots\}$ (and not merely on the observation noise sequence $\{(V_{t+1}^\circ),\ t=0,1,\ldots\}$ as in the uncorrelated case [3, 4]). To that end, define the filtration $\{\mathcal{F}_t,\ t=0,1,\ldots\}$ by

$$\mathcal{F}_{t+1} := \mathcal{F}_0 \vee \sigma\{W_{t+1}^\circ, V_{s+1}^\circ,\ s=0,1,\ldots t\} \qquad t=0,1,\ldots \quad (3.1)$$

with $\mathcal{F}_0 := \sigma\{\xi\}$, and the $I\!R^{n+k}$-valued sequence $\{(W_{t+1},V_{t+1}),\ t=0,1,\ldots\}$ by

$$\begin{pmatrix} W_{t+1} \\ V_{t+1} \end{pmatrix} = \begin{pmatrix} W_{t+1}^\circ \\ V_{t+1}^\circ \end{pmatrix} - \begin{pmatrix} \Sigma_{t+1}^{w} & \Sigma_{t+1}^{wv} \\ \Sigma_{t+1}^{vw} & \Sigma_{t+1}^{v} \end{pmatrix}\begin{pmatrix} \varphi_t^w \\ \varphi_t^v \end{pmatrix} \qquad t=0,1,\ldots \quad (3.2)$$

where $\{\varphi_t^w,\ t=0,1,\ldots\}$ and $\{\varphi_t^v,\ t=0,1,\ldots\}$ are $\mathcal{F}_t$-adapted sequences taking values in $I\!R^n$ and $I\!R^k$, respectively, yet to be specified. Recalling the Girsanov transformation [2], we see that if for *any* two such sequences $\{\varphi_t^w,\ t=0,1,\ldots\}$ and $\{\varphi_t^v,\ t=0,1,\ldots\}$, we define $\{(W_{t+1},V_{t+1}),\ t=0,1,\ldots\}$ by (3.2), then for each $T=0,1,\ldots$ we can find a probability measure $\bar{P}$ on $(\Omega,\mathcal{F})$ satisfying (B) where

(B): *The probability measure $\bar{P}$ is mutually absolutely continuous with P on $\mathcal{F}$ and agrees with P on $\mathcal{F}_0$. Furthermore, $\{(W_{t+1},V_{t+1}),\ t=0,1,\ldots,T\}$ is a zero-mean $(\mathcal{F}_t,\bar{P})$ GWN sequence with the same covariance structure under $\bar{P}$ as the covariance structure under P of the original noise sequence $\{(W_{t+1}^\circ,V_{t+1}^\circ),\ t=0,1,\ldots,T\}$.*

Following [3, 4], we decompose the plant process $\{X_t^\circ,\ t=0,1,\ldots\}$ as

$$X_t^\circ = X_t + Z_t \qquad t=0,1,\ldots \tag{3.3}$$

where the $I\!R^n$-valued $\{X_t,\ t=0,1,\ldots\}$ carries the randomness due to the plant noise process $\{W_{t+1}^\circ,\ t=0,1,\ldots\}$, and where the $I\!R^n$-valued $\{Z_t,\ t=0,1,\ldots\}$ contains only the randomness due to the initial condition ξ. It is argued in [3, 4] that the sequences $\{\varphi_t^w,\ t=0,1,\ldots\}$ and $\{\varphi_t^v,\ t=0,1,\ldots\}$ in (3.2) must necessarily have the form

$$\varphi_t^w = \varphi_t \quad \text{and} \quad \varphi_t^v = -(\Sigma_{t+1}^v)^{-1}[\Sigma_{t+1}^{vw}\varphi_t + H_t Z_t] \qquad t=0,1,\ldots \tag{3.4}$$

for some unspecified $\mathcal{F}_t$-adapted sequence $\{\varphi_t,\ t=0,1,\ldots\}$ taking values in $I\!R^n$. Injecting (3.4) into (3.2), we obtain

$$W_{t+1} = W_{t+1}^\circ + \Sigma_{t+1}^{wv}(\Sigma_{t+1}^v)^{-1}H_t Z_t[\Sigma_{t+1}^w - \Sigma_{t+1}^{wv}(\Sigma_{t+1}^v)^{-1}\Sigma_{t+1}^{vw}]\varphi_t \qquad t=0,1,\ldots \tag{3.5}$$

and the Girsanov theorem gives the appropriate probability measure $\bar{P}$ which satisfies **(B)**, via

$$\begin{aligned}\frac{d\bar{P}}{dP} = \exp\Bigg[&\sum_{s=0}^{T}\Big[\varphi_s'[W_{s+1}^\circ - \Sigma_{s+1}^{wv}(\Sigma_{s+1}^v)^{-1}V_{s+1}^\circ] - Z_s'H_s'(\Sigma_{s+1}^v)^{-1}V_{s+1}^\circ\Big] \\ &+ \frac{1}{2}\sum_{s=0}^{T}\Big[\varphi_s'[\Sigma_{s+1}^w - \Sigma_{s+1}^{wv}(\Sigma_{s+1}^v)^{-1}\Sigma_{s+1}^{vw}]\varphi_s + Z_s'H_s'(\Sigma_{s+1}^v)^{-1}H_s Z_s\Big]\Bigg].\end{aligned} \tag{3.6}$$

In order to complete the description of the probability measure (3.6), we must specify $\{X_t,\ t=0,1,\ldots\}$, $\{Z_t,\ t=0,1,\ldots\}$, and $\{\varphi_t,\ t=0,1,\ldots\}$. To that end we rewrite the evolution of $\{X_t^\circ,\ t=0,1,\ldots\}$ in terms of $\{X_t,\ t=0,1,\ldots\}$, $\{Z_t,\ t=0,1,\ldots\}$ and $\{W_{t+1},\ t=0,1,\ldots\}$. Since we wish to use the properties of $\bar{P}$, it is more natural to write this evolution in terms of $\{W_{t+1},\ t=0,1,\ldots\}$ rather than in terms of $\{W_{t+1}^\circ,\ t=0,1,\ldots\}$, and this leads to

$$\begin{aligned}X_{t+1} + Z_{t+1} &= A_t X_t^\circ + W_{t+1}^\circ \\ &= A_t(X_t + Z_t) + W_{t+1} - \Sigma_{t+1}^{wv}(\Sigma_{t+1}^v)^{-1}H_t Z_t \\ &\quad + [\Sigma_{t+1}^w - \Sigma_{t+1}^{wv}(\Sigma_{t+1}^v)^{-1}\Sigma_{t+1}^{vw}]\varphi_t \\ &= A_t X_t + [A_t - \Sigma_{t+1}^{wv}(\Sigma_{t+1}^v)^{-1}H_t]Z_t + W_{t+1} \\ &\quad + [\Sigma_{t+1}^w - \Sigma_{t+1}^{wv}(\Sigma_{t+1}^v)^{-1}\Sigma_{t+1}^{vw}]\varphi_t.\end{aligned} \qquad t=0,1,\ldots \tag{3.7}$$

This suggests a separation of the dynamics in the form

$$\begin{aligned}X_{t+1} &= A_t X_t + W_{t+1} + [\Sigma_{t+1}^w - \Sigma_{t+1}^{wv}(\Sigma_{t+1}^v)^{-1}\Sigma_{t+1}^{vw}]\varphi_t - \pi_t \\ Z_{t+1} &= [A_t - \Sigma_{t+1}^{wv}(\Sigma_{t+1}^v)^{-1}H_t]Z_t + \pi_t\end{aligned} \qquad t=0,1,\ldots \tag{3.8}$$

with initial conditions $X_0=\zeta$ and $Z_0=\xi-\zeta$ where ζ and $\{\pi_t,\ t=0,1,\ldots\}$ are $I\!R^n$-valued RVs yet to be specified. At this point, we simply assume

$$\varphi_t = 0, \quad \pi_t = 0 \quad \text{and} \quad \zeta = 0 \qquad t=0,1,\ldots \tag{3.9}$$

and summarize the relevant quantities under this constraint (3.9).

• **The effect of the initial condition**

$$\begin{aligned} Z_{t+1} &= [A_t - \Sigma^{wv}_{t+1}(\Sigma^v_{t+1})^{-1}H_t]Z_t \\ Z_0 &= \xi, \end{aligned} \qquad t = 0,1,\ldots \quad (3.10)$$

so that $Z_t = \Psi(t,0)\xi$ for $t = 0,1,\ldots$.

• **The noise processes**

$$\begin{aligned} \begin{pmatrix} W_{t+1} \\ V_{t+1} \end{pmatrix} &= \begin{pmatrix} W^\circ_{t+1} \\ V^\circ_{t+1} \end{pmatrix} - \begin{pmatrix} \Sigma^w_{t+1} & \Sigma^{wv}_{t+1} \\ \Sigma^{vw}_{t+1} & \Sigma^v_{t+1} \end{pmatrix} \begin{pmatrix} 0 \\ -(\Sigma^v_{t+1})^{-1}H_tZ_t \end{pmatrix} \\ &= \begin{pmatrix} W^\circ_{t+1} + \Sigma^{vw}_{t+1}(\Sigma^v_{t+1})^{-1}H_tZ_t \\ V^\circ_{t+1} + H_tZ_t \end{pmatrix}. \end{aligned} \qquad t = 0,1,\ldots \quad (3.11)$$

• **The auxiliary system**

$$\begin{aligned} X_{t+1} &= A_tX_t + W_{t+1} \\ X_0 &= 0 \\ Y_t &= H_tX_t + V_{t+1}. \end{aligned} \qquad t = 0,1,\ldots \quad (3.12)$$

• **The change of measure**

$$\frac{d\bar{P}}{dP} = \exp\Big[-\sum_{s=0}^{T} Z'_sH'_s(\Sigma^v_{s+1})^{-1}V^\circ_{s+1} + \frac{1}{2}\sum_{s=0}^{T} Z'_sH'_s(\Sigma^v_{s+1})^{-1}H_sZ_s\Big]. \quad (3.13)$$

The properties of our decomposition and change of measure are summarized in

Proposition 1. *Let the filtration $\{\mathcal{F}_t,\ t = 0,1,\ldots\}$ be given by (3.1). If the sequences $\{X_t,\ t = 0,1,\ldots\}$, $\{Z_t,\ t = 0,1,\ldots\}$ and $\{(W_{t+1},V_{t+1}),\ t = 0,1,\ldots\}$ are defined by (3.10)-(3.12) and if the probability measure $\bar{P}$ is defined by (3.13), then $\bar{P}$ and P are mutually absolutely continuous on $\mathcal{F}$, agree on $\mathcal{F}_0$, and the process $\{(W_{t+1},V_{t+1}),\ t = 0,1,\ldots,T\}$ is a zero-mean $(\mathcal{F}_t,\bar{P})$ GWN sequence with covariance structure $\{\Gamma_{t+1},\ t = 0,1,\ldots,T\}$.*

Motivated by the form of (3.13), we define the $\mathbb{R}$-valued sequence $\{L_t,\ t = 0,1,\ldots\}$ by

$$L_{t+1} := \exp\Big[-\sum_{s=0}^{t} Z'_sH'_s(\Sigma^v_{s+1})^{-1}V^\circ_{s+1} + \frac{1}{2}\sum_{s=0}^{t} Z'_sH'_s(\Sigma^v_{s+1})^{-1}H_sZ_s\Big] \qquad t = 0,1,\ldots \quad (3.14)$$

with $L_0 = 1$, and observe that $d\bar{P}/dP = L_{T+1}$. The arguments of [3] and [4] can now be applied *in extenso* to yield the results (2.1)-(2.8) over the finite horizon.

The final step now consists in extending these results from the finite horizon $t = 0,1,\ldots,T$ to the infinite horizon $t = 0,1,\ldots$. To that end, note the following: The dynamics of the sequences $\{(\bar{X}_t,\bar{B}_t),\ t = 0,1,\ldots,T+1\}$ and $\{\Sigma_t,\ t = 0,1,\ldots,T+1\}$ are *independent* of T. Moreover, although the transformed measure $\bar{P}$ used in the derivation depends *a priori* on T, the definitions of the mappings $\mathcal{T}\phi$ and $\mathcal{U}\phi$ are independent of T. These remarks are sufficient to yield Theorem 1 from the finite-horizon results of this section. ■

Following on the comments made at the end of the proof, we could have displayed explicitly the dependence of the transformed measure $\bar{P}$ on the parameter T, say through the notation $\bar{P}_{T+1}$. Although $\bar{P}_{T+1} = \bar{P}_T$ on the σ-field $\mathcal{F}_T$ for all $T = 0,1,\ldots$, and the probability measure $\bar{P}_{T+1}$ is mutually absolutely continuous with respect to P, it is *not* true in general [4] that the *projective system* $\{\bar{P}_T,\ T = 0,1,\ldots\}$ has a limit $\bar{P}$ which is absolutely continuous with respect to P on the σ-field $\vee_T\mathcal{F}_T$, i.e., there does not exist necessarily a probability measure $\bar{P}$ on $\vee_T\mathcal{F}_T$ such that $\bar{P}$ is absolutely continuous with respect to P, and $\bar{P}_T = \bar{P}$ on the σ-field $\mathcal{F}_T$ for all $T = 0,1,\ldots$. Although this could *a priori* complicate matters for the infinite-horizon situation, we shall not concern ourselves with this difficulty in what follows. Indeed, in the remainder of this paper, only statements for finite t will be made and the notation $\bar{P}$ (and $\bar{E}$) will be used throughout with the understanding that $\bar{P} = \bar{P}_{T+1}$ for some $t < T$. As should be clear from earlier comments, the exact choice of T is irrelevant.

IV. REPRESENTATIONS FOR $\{\hat{X}_t,\ t = 0, 1, \ldots\}$, $\{\hat{X}^K_t,\ t = 0, 1, \ldots\}$ AND $\{\epsilon_t,\ t = 0, 1, \ldots\}$.

Using Theorem 1, we now develop formulae for $\{\hat{X}_t,\ t = 0, 1, \ldots\}$, $\{\hat{X}^K_t,\ t = 0, 1, \ldots\}$ and $\{\epsilon_t,\ t = 0, 1, \ldots\}$. We do this under the additional assumption (A.4), where

(A.4): The covariance matrix Δ is positive-definite.

To state these representation results, we find it convenient to introduce the auxiliary quantities $\{Q^*_t,\ t = 0, 1, \ldots\}$ and $\{R^*_t,\ t = 0, 1, \ldots\}$ in $\mathcal{M}_{n\times n}$ and $\mathcal{Q}_n$, respectively, by setting

$$Q^*_t := Q_t + \Psi(t, 0) \quad \text{and} \quad R^*_t := M_t - R_t. \qquad t = 0, 1, \ldots \quad (4.1)$$

With this notation, we have

Theorem 2. *For all $t = 0, 1, \ldots$, the representations*

$$\hat{X}_{t+1} = \bar{X}_{t+1} + Q^*_{t+1} \frac{\int_{I\!R^n} z \exp\left[z'\bar{B}_{t+1} - \frac{1}{2} z' R^*_{t+1} z\right] dF(z)}{\int_{I\!R^n} \exp\left[z'\bar{B}_{t+1} - \frac{1}{2} z' R^*_{t+1} z\right] dF(z)} \quad (4.2)$$

and

$$\hat{X}^K_{t+1} = \bar{X}_{t+1} + Q^*_{t+1} \left[R^*_{t+1} + \Delta^{-1}\right]^{-1} \left[\bar{B}_{t+1} + \Delta^{-1}\mu\right] \quad (4.3)$$

hold P-a.s.

Before discussing a proof of this result, several points are in order:

(i): The expression (4.3) provides a non-standard representation for the Kalman filter associated with system (1.1). This representation is notable in that it explicitly displays the effects of the mean μ and covariance Δ of the initial condition ξ; the only dependence of the filtering formulae on μ and Δ is through the affine mapping $x \mapsto [R^*_{t+1} + \Delta^{-1}]^{-1}[x + \Delta^{-1}\mu]$.

(ii): We readily see from (1.6), (2.3) and (2.8) that

$$\begin{aligned} Q^*_{t+1} &= \left[A_t - [A_t P_t H'_t + \Sigma^{wv}_{t+1}] J_t^{-1} H_t\right] Q^*_t \\ R^*_{t+1} &= R^*_t + Q^{*\prime}_t H'_t J_t^{-1} H_t Q^*_t \end{aligned} \qquad t = 0, 1, \ldots \quad (4.5)$$

with initial conditions $Q^*_0 = I_n$ and $R^*_0 = O_n$. Note also that the dynamics (2.5) then simplifies into

$$\begin{aligned} \bar{B}_{t+1} &= \bar{B}_t - Q^{*\prime}_t H'_t J_t^{-1} H_t \bar{X}_t + Q^{*\prime}_t H'_t J_t^{-1} Y_t \\ \bar{B}_0 &= 0. \end{aligned} \qquad t = 0, 1, \ldots \quad (4.6)$$

The following two technical lemmas will be useful in the forthcoming discussion. The proofs are available in [4] and are omitted here in the interest of brevity.

Lemma 1. *For $t = 0, 1, \ldots$, $\bar{B}_{t+1}$ is a zero-mean Gaussian RV with covariance R^*_{t+1} under $\bar{P}$.*

Lemma 2. *For any $t = 0, 1, \ldots$ and any $I\!R$-valued, nonnegative $\mathcal{Y}_t \vee \sigma\{\xi\}$-measurable RV X, the relation*

$$E[X] = \bar{E}\left[X \exp[\xi' \bar{B}_{t+1} - \frac{1}{2} \xi' R^*_{t+1} \xi]\right] \quad (4.7)$$

holds true.

A proof of Theorem 2. The first step consists in finding a representation for the conditional characteristic function $E\left[\exp[i\theta' X^\circ_{t+1}] \middle| \mathcal{Y}_t\right]$. Under the enforced moment assumptions on ξ, an expression for the conditional mean is recovered by differentiating this characteristic function with respect to θ and then setting $\theta = 0$.

Finally, by substituting a Gaussian distribution for F in this representation for $\hat{X}_{t+1}$, we obtain a formula for $\hat{X}^K_{t+1}$. Details are available in [4]. ∎

Theorem 2 now leads us to a simple representation of the errors $\{\epsilon_{t+1},\ t=0,1,\ldots\}$. In what follows, for each Λ in $\mathcal{Q}_n$, G_Λ denotes a normal distribution with zero mean and covariance Λ.

Theorem 3. *The representation*

$$\epsilon_{t+1} = \int_{I\!R^n} \frac{\left\| Q^*_{t+1} \int_{I\!R^n} \{ z - [R^*_{t+1} + \Delta^{-1}]^{-1}[b + \Delta^{-1}\mu] \} \exp[z'b - \frac{1}{2} z' R^*_{t+1} z] dF(z) \right\|^2}{\int_{I\!R^n} \exp[z'b - \frac{1}{2} z' R^*_{t+1} z] dF(z)} dG_{R^*_{t+1}}(b) \qquad t = 0,1,\ldots \quad (4.8)$$

holds true.

Proof. We observe directly from Theorem 2 that

$$\begin{aligned} &\hat{X}_{t+1} - \hat{X}^K_{t+1} \\ &= Q^*_{t+1} \cdot \frac{\int_{I\!R^n} \{ z - [R^*_{t+1} + \Delta^{-1}]^{-1}[\bar{B}_{t+1} + \Delta^{-1}\mu] \} \exp[z'\bar{B}_{t+1} - \frac{1}{2} z' R^*_{t+1} z] dF(z)}{\int_{I\!R^n} \exp[z'\bar{B}_{t+1} - \frac{1}{2} z' R^*_{t+1} z] dF(z)} \end{aligned} \quad (4.9)$$

for all $t = 0,1,\ldots$, whence upon changing to the measure $\bar{P}$, we find that

$$\epsilon_{t+1} = \bar{E}\left[\frac{\left\| Q^*_{t+1} \int_{I\!R^n} \{ z - [R^*_{t+1} + \Delta^{-1}]^{-1}[\bar{B}_{t+1} + \Delta^{-1}\mu] \} \exp[z'\bar{B}_{t+1} - \frac{1}{2} z' R^*_{t+1} z] dF(z) \right\|^2}{\int_{I\!R^n} \exp[z'\bar{B}_{t+1} - \frac{1}{2} z' R^*_{t+1} z] dF(z)} \right]$$

by using Lemma 2. We now obtain (4.8) by a simple application of Lemma 1 on this last relation. ∎

V. ASYMPTOTICS – THE SCALAR CASE

We now use the representation result of Theorem 3 to investigate the asymptotic behavior of the sequence $\{\epsilon_t,\ t = 1,2,\ldots\}$ in the scalar case, i.e., $n = k = 1$, when the system dynamics are time invariant.

Let us first fix some notation. In accordance with common usage, we use lower case letters to denote scalar quantities so that $X^\circ_t = x^\circ_t$, $A_t = a$, $H_t = h$, etc. Let $\mathcal{D}$ denote the collection of square-integrable distribution functions on $I\!R$, and let $\mathcal{D}_0$ denote the zero-mean elements of $\mathcal{D}$. For $r \geq 0$, let G_r denote the zero-mean Gaussian distribution with variance r.

We characterize the asymptotic behavior of $\{\epsilon_t,\ t = 1,2,\ldots\}$ in terms of the auxiliary quantities

$$\bar{a} := a - \frac{\sigma^{wv} h}{\sigma^v} \quad \text{and} \quad \bar{c} := \sigma^w - \frac{(\sigma^{wv})^2}{\sigma^v} \quad (5.1)$$

As a final remark before presenting the asymptotic analysis, observe that there is no loss of generality in considering only initial distributions in $\mathcal{D}_0$. Indeed, since both the true and wide-sense conditional expectation operators are linear and since the state x°_t is affine in the initial condition ξ, we may subtract out $E[\xi]$ when forming the difference $\epsilon_t = \hat{x}_t - \hat{x}^K_t$. For F in $\mathcal{D}_0$, (4.8) reduces to

$$\epsilon_t = (q^*_t)^2 I_F(r^*_t) \qquad t = 1,2,\ldots \quad (5.2)$$

where for $r \geq 0$ we define

$$I_F(r) := \int_{I\!R} \frac{\left| \int_{I\!R} \left\{ z - \frac{b}{r + 1/\Delta} \right\} \exp[zb - \frac{1}{2} z^2 r] dF(z) \right|^2}{\int_{I\!R} \exp[zb - \frac{1}{2} z^2 r] dF(z)} dG_r(b). \quad (5.3)$$

Note that for any non-Gaussian F in $\mathcal{D}_0$, I_F is positive definite in that $I_F(r) = 0$ if and only if $r = 0$ [4].

Moreover, if $h = 0$ or $\bar{a} = 0$, then the system dynamics imply $r_t^* = 0$ for all $t = 0, 1, \ldots$, whence $\epsilon_t = 0$ for all $t = 1, 2, \ldots$ and all distributions F in $\mathcal{D}_0$. Thus only the cases $h \neq 0$ and $\bar{a} \neq 0$ are of interest. The main result of this section is

Theorem 4. *We have the following convergence results when $n = k = 1$ and when $h \neq 0$ and $\bar{a} \neq 0$.*

1. *If $\bar{c} \neq 0$, or if $|\bar{a}| \leq 1$ and $\bar{c} = 0$, then $\lim_t \epsilon_t = 0$ for any distribution F in $\mathcal{D}$, whereas if $\bar{c} = 0$ and $|\bar{a}| > 1$, then the asymptotic behavior of ϵ_t depends nontrivially upon F in $\mathcal{D}$.*

Moreover we also have the following estimates:

2. *If $\bar{c} \neq 0$, or if $\bar{c} = 0$ and $|\bar{a}| < 1$, $\lim_t \epsilon_t = 0$ at an exponential rate independent of F for F in $\mathcal{D}$ non-Gaussian whereas if $\bar{c} = 0$ and $|\bar{a}| = 1$, then the rate depends non-trivially upon F.*

The proof of Theorem 4 is given in Propositions 2–4 below by considering all possible cases.

Proposition 2. *Assume $h \neq 0$ and $\bar{a} \neq 0$.*

1. *If $\bar{c} \neq 0$, $\lim_t \epsilon_t = 0$ for all distributions F in $\mathcal{D}$, with rate*

$$\lim_t \frac{1}{t} \ln \epsilon_t = 2 \ln \left| \bar{a} \left(\frac{\sigma^v}{h^2 p_\infty + \sigma^v} \right) \right| < 0 \tag{5.4}$$

   for all non-Gaussian distributions F in $\mathcal{D}$, where $p_\infty := \lim_t p_t$.

2. *If $\bar{c} = 0$ and $|\bar{a}| < 1$, then $\lim_t \epsilon_t = 0$ for all distributions F in $\mathcal{D}$ with rate*

$$\lim_t \frac{1}{t} \ln \epsilon = 2 \ln |\bar{a}| < 0 \tag{5.5}$$

for all non-Gaussian F in $\mathcal{D}$.

Proof. If $\bar{c} \neq 0$, then the pair $(\bar{a}, \bar{c})$ is controllable and p_∞ is well-defined, finite and positive and by standard results [1, Theorem 5.1 and Appendix 1] we conclude that

$$\left| a - \frac{a p_\infty h + \sigma^v}{h^2 p_\infty + \sigma^v} h \right| = \left| \frac{\bar{a} \sigma^v}{h^2 p_\infty + \sigma^v} \right| < 1. \tag{5.6}$$

It is not now difficult to see from (5.3) that

$$\lim_t \frac{1}{t} \ln (q_t^*)^2 = 2 \ln \left| \frac{\bar{a} \sigma^v}{h^2 p_\infty + \sigma^v} \right| < 0, \tag{5.7}$$

and that r_∞^* thus must be finite and positive. It then follows from standard arguments that

$$0 < \liminf_t I_F(r_t^*) \leq \limsup_t I_F(r_t^*) < \infty, \tag{5.8}$$

and this, together with (5.7), is sufficient to prove claim 1 for F in $\mathcal{D}_0$. The proof of claim 2 is similar. When $0 < |\bar{a}| < 1$ and $\bar{c} = 0$, the dynamics of q_t^* yield $\lim_t t^{-1} \ln (q_t^*)^2 = 2 \ln |\bar{a}| < 0$. The dynamics of r_t^* then require that $0 < r_\infty^* < \infty$, and that again (5.8) hold. The combination of (5.2) and (5.8) prove claim 2 for F in $\mathcal{D}_0$. ■

The dependencies given in Theorem 4 when $\bar{c} = 0$ and $|\bar{a}| \geq 1$ are now illustrated through some simple examples. First, however, we verify a general result.

Proposition 3. *For any distribution F in $\mathcal{D}_0$, $\limsup_t t I_F(t) < \infty$ 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

$$\begin{aligned} x_t^\circ &= \xi \\ y_t &= \xi + v_{t+1}^\circ. \end{aligned} \qquad t = 0, 1, \ldots \tag{5.9}$$

Here $a = h = \sigma^v = 1$ and $\sigma^w = 0$ so that $\bar{c} = 0$ and $|\bar{a}| = 1$. Consequently $q_t^* = 1$ and $r_t^* = t$ for all $t = 1, 2, \ldots$, whence $\epsilon_t = I_F(t)$ 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

$$\check{x}_{t+1} := \frac{1}{t+1} \sum_{s=0}^{t} y_s \qquad t = 0, 1, \ldots \quad (5.10)$$
$$\check{x}_0 := 0.$$

Since $\check{x}_t$ is a *linear* estimator, $\hat{x}_t^K$ the LLMSE estimator and $\hat{x}_t$ the MMSE estimator, we conclude that $E[|\hat{x}_t - \hat{x}_t^K|^2] \leq 4E[|\check{x}_t - x_t^\circ|^2]$ by a straightforward application of the triangle inequality. From (5.10), we verify that $I_F(t) = \epsilon_t \leq 4/(t+1)$, and the claim is immediate. ∎

We now consider two distributions F_1 and F_2 in $\mathcal{D}_0$.

Distribution F_1. Distribution F_1 admits a density with respect to Lebesgue measure λ on $I\!R$ given by

$$\frac{dF_1}{d\lambda}(z) = \sum_{i=1}^{n} \alpha_i \frac{1}{\sqrt{2\pi\rho^2}} \exp\left[-\frac{1}{2}\frac{(z-\mu_i)^2}{\rho^2}\right] \qquad z \in I\!R \quad (5.11)$$

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 and $\sum_{i=1}^n p_i z_i = 0$.

The following two facts are proved in [4].

Fact 1. *We have* $\lim_t (\rho^2 t + 1)^2 I_{F_1}(t) = K$ *for some* $K > 0$.

Fact 2. *We also have* $\lim_t t I_{F_2}(t) = 1$.

We now can prove the rest of Theorem 4.

Proposition 4. *Assume* $h \neq 0$ *and* $\bar{c} \neq 0$.

1. *If* $|\bar{a}| = 1$, *then* $\lim_t \epsilon_t = 0$ *for any distribution* F *in* $\mathcal{D}$, *the rate of convergence depending nontrivially upon* F *for* F *non-Gaussian,*

2. *If* $|\bar{a}| > 1$, *then* $\limsup_t \epsilon_t < \infty$ *for all distributions* F *in* $\mathcal{D}$, *with the asymptotic behavior depending nontrivially upon* F *for* F *not Gaussian.*

Proof. *Claim 1.* Under the enforced assumptions, we have $\epsilon_t = (1)^t I_F(t)$ for all $t = 0, 1, \ldots$ and all F in $\mathcal{D}_0$. By Proposition 2, $\lim_t \epsilon_t = 0$; however, if $F = F_1$, $\lim_t (\ln \epsilon_t / \ln t) = -2$, whereas if $F = F_2$, $\lim_t (\ln \epsilon_t / \ln t) = -1$. *Claim 2.* 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 $\mathcal{D}_0$, then

$$\epsilon_t = \frac{(q_t^*)^2}{r_t^*} \left(r_t^* I_F(r_t^*)\right). \qquad t = 1, 2, \ldots \quad (5.12)$$

Again applying Proposition 2, we get $\limsup_t \epsilon_t < \infty$ for all F in $\mathcal{D}_0$. However, if $F = F_1$, $\lim_t \epsilon_t = 0$, whereas if $F = F_2$, then $\lim_t \epsilon_t = 1$. ∎

AFFILIATIONS

The work of R. B. Sowers was supported through an ONR Graduate Fellowship while completing the requirements towards the M.S. degree. The work of A. M. Makowski was supported partially through the Engineering Research Centers Programs NSF CDR 88-03012 and partially through NSF grant ECS 83-51836.

REFERENCES

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

[2] G. DiMasi and W. Runggaldier, "On measure transformations for combined filtering and parameter estimation in discrete time," *Systems & Control Letters* **2**, pp. 57-62 (1982).

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

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

[5] R.B. Sowers and A.M. Makowski, "Discrete-time filtering for linear systems in correlated noise with non-Gaussian initial conditions," *Proceedings of the Conference on Information Sciences and Systems*, The Johns Hopkins University, Baltimore (MD), March 1989.

[6] R. B. Sowers and A.M. Makowski, "Filtering for discrete-time linear systems with non-Gaussian initial conditions: Asymptotic behavior of the difference between the MMSE and LMSE estimates," submitted to *IEEE Transactions on Automatic Control* (1989).