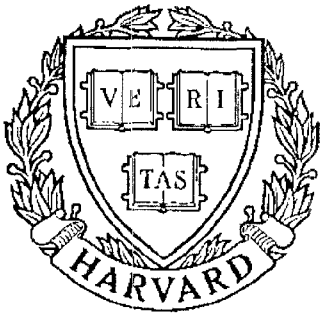
TECHNICAL RESEARCH REPORT

SYSTEMS RESEARCH CENTER



Supported by the National Science Foundation Engineering Research Center Program (NSFD CD 8803012), Industry and the University

On the Effects of the Initial Condition in State Estimation for Discrete-Time Linear Systems

by R.B. Sowers and A.M. Makowski

TR 92-114

ON THE EFFECTS OF THE INITIAL CONDITION IN STATE ESTIMATION FOR DISCRETE–TIME LINEAR SYSTEMS

Richard B. Sowers
University of Southern California

Armand M. Makowski
University of Maryland

Contributed chapter in *Control and Dynamical Systems*, Ed. C.T. Leondes, Academic Press, Inc. Scheduled to appear, 1993.

ON THE EFFECTS OF THE INITIAL CONDITION IN STATE ESTIMATION FOR DISCRETE–TIME LINEAR SYSTEMS

Richard B. Sowers
Center for Applied
Mathematical Sciences
University of Southern California
Los Angeles, CA 90089–1113

Armand M. Makowski
Electrical Engineering Department
and Systems Research Center
University of Maryland
College Park, MD 20742

ABSTRACT

We consider the one–step prediction problem for discrete–time linear systems in correlated Gaussian white plant and observation noises, and non–Gaussian initial conditions. Explicit representations are obtained for the MMSE and LLSE (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 explicitly display the effects of the distribution of the initial condition. With the help of these formulae, we investigate the large–time asymptotics of ε_t, the expected squared difference between the MMSE and LLSE estimates at time t. We characterize the *limit* of the error sequence $\{\varepsilon_t,\ t = 1, 2, \ldots\}$ and obtain some related *rates* of convergence. A complete large–time analysis is provided for the scalar case.

Key Words: correlated noise, Girsanov transformation, Kalman filtering, large–time asymptotics, linear systems, non–Gaussian initial conditions, nonlinear filtering theory.

Contributed chapter in *Control and Dynamical Systems*, Ed. C.T. Leondes, Academic Press, Inc. Scheduled to appear, 1993.

I. INTRODUCTION

In his seminal paper of 1960, Kalman [13] developed a method of estimating the state of a noisy linear dynamical system based upon linear observations corrupted by additive Gaussian white noise. The import of Kalman's state–space approach was that it provided a *dynamical and recursive, and hence computable* description of the estimator, thereby overcoming, for many practical problems, the restrictive stationarity assumptions of the Wiener–Hopf–Kolmogorov theory of linear filtering. Kalman's state–space approach renewed intense interest in filtering theory, eventually leading to a clearer understanding of the general problem of filtering a nonlinear dynamical plant given nonlinear observations. Advances in nonlinear filtering theory have, in turn, motivated fundamental and far–reaching breakthroughs in a wide range of other probabilistic questions, from stochastic control theory [8] to martingale theory [8, 12] to stochastic partial differential equations [24–25].

We shall revisit, in this chapter, Kalman's problem. His model, as we shall understand it here, is that of an $\mathbb{R}^n$–valued *plant* process evolving according to the stochastic discrete–time linear dynamical equations

$$X_0^\circ = \xi, \quad X_{t+1}^\circ = A_t X_t^\circ + W_{t+1}^\circ. \qquad t = 0, 1, \ldots \quad (1.1a)$$

This plant process describes the evolution of some quantity of interest—the so–called system *state*, e.g., the amount of a quantity in a chemical reaction or the position and velocity of an orbiting satellite. Unfortunately, full state information is often not available and we can only measure a sequence of $\mathbb{R}^k$–valued *observations* which are given by the linear equations

$$Y_t = H_t X_t^\circ + V_{t+1}^\circ. \qquad t = 0, 1, \ldots \quad (1.1b)$$

To be rigorous, we denote by $(\Omega, \mathcal{F}, \mathbf{P}^\circ)$ an underlying probability triple on which all random variables (rvs) are defined. Of course, the matrices A_t and H_t are respectively of size $n \times n$ and $k \times n$ for each $t = 0, 1, \ldots$. The statistics of the random noise processes $\mathbf{W}^\circ = \{W_{t+1}^\circ;\, t = 0, 1, \ldots\}$ and $\mathbf{V}^\circ = \{V_{t+1}^\circ;\, t = 0, 1, \ldots\}$ are governed by the following assumptions:

(A.1): The process $(\mathbf{W}^\circ, \mathbf{V}^\circ)$ is a zero–mean Gaussian white noise (GWN) sequence with covariance structure $\mathbf{\Sigma} = \{\Sigma_{t+1};\ t =$

$0, 1, \ldots\}$ given by

$$\Sigma_{t+1} := \mathrm{Cov}\begin{pmatrix} W^\circ_{t+1} \\ V^\circ_{t+1} \end{pmatrix} = \begin{pmatrix} \Sigma^w_{t+1} & \Sigma^{wv}_{t+1} \\ \Sigma^{vw}_{t+1} & \Sigma^v_{t+1} \end{pmatrix}, \quad t = 0, 1, \ldots \tag{1.2}$$

i.e., the $\mathbb{R}^{n+k}$–valued rvs $\{(W^\circ_{t+1}, V^\circ_{t+1});\ t = 0, 1, \ldots\}$ are mutually independent zero–mean Gaussian rvs with covariances (1.2); and

(A.2): For all $t = 1, 2, \ldots$, the covariance matrix Σ^v_t is positive definite (and thus invertible).

The reader is referred to [7, 27] for background material on GWN sequences. In the classical Kalman filtering model, the statistics of the initial condition ξ are assumed to be governed by

(K): The initial condition ξ is a Gaussian rv with mean μ and covariance Δ, and is independent of the process $(\mathbf{W}^\circ, \mathbf{V}^\circ)$.

We note that there is an analogous continuous–time formulation of (1.1a)–(1.1b) using Ito equations [8, 12, 14, 17]. However, we have elected here to study the discrete–time model in order to minimize technicalities and since, in applications, at most a finite number of observations are usually recorded. We also note, in passing, that the superscript '$\circ$' on the plant $\mathbf{X}^\circ = \{X^\circ_t;\ t = 0, 1, \ldots\}$, noises $(\mathbf{W}^\circ, \mathbf{V}^\circ) = \{(W^\circ_{t+1}, V^\circ_{t+1});\ t = 0, 1, \ldots\}$ and measure $\mathbf{P}^\circ$ indicates that these are 'original' components of the model, to be distinguished from auxiliary plant and noise processes and probability measures which we define in the course of the analysis.

In [13], Kalman then posed the problem of estimating in the minimum–mean–square–error sense the state X°_{t+1} of the plant given the observations $Y_0, Y_1, \ldots, Y_t$, for each $t = 0, 1, \ldots$. In particular, he set out to compute the *conditional means*

$$\hat{X}_{t+1} := \mathbf{E}^\circ[X^\circ_{t+1} | \mathcal{Y}_t] \qquad t = 0, 1, \ldots \tag{1.3}$$

where the σ–field $\mathcal{Y}_t$ is defined by

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

In the process of doing so, he also solved the *generalized one–step prediction problem*, which is defined as the finding of the conditional law of X°_{t+1} given

the σ–field $\mathcal{Y}_t$. We may alternately formulate this latter question as the simultaneous evaluation of the conditional expectations

$$\mathbf{E}^\circ[\varphi(X^\circ_{t+1})|\mathcal{Y}_t] \qquad t = 0, 1, \ldots \quad (1.4)$$

for all bounded Borel mappings $\varphi : \mathbb{R}^n \to \mathbb{C}$, with $\mathbb{C}$ denoting the set of complex numbers. Under assumptions **(A.1)**, **(A.2)** and **(K)**, the linearity of (1.1a)–(1.1b) implies that for each $t = 0, 1, \ldots$, the rvs $\{X^\circ_{t+1}, Y_0, Y_1, \ldots, Y_t\}$ are jointly Gaussian and therefore X°_{t+1} is conditionally Gaussian given $\mathcal{Y}_t$ [2, Sec. 2.2]. The generalized one–step prediction problem is then solved once two sequences of finite–dimensional sufficient statistics are known, namely the conditional means of (1.3) and the *conditional covariances*

$$\hat{P}_{t+1} := \mathbf{E}^\circ[(X^\circ_{t+1} - \hat{X}_{t+1})'(X^\circ_{t+1} - \hat{X}_{t+1})|\mathcal{Y}_t], \qquad t = 0, 1, \ldots \quad (1.5)$$

with $'$ denoting transpose. Kalman's breakthrough lies in showing that the processes $\hat{\mathbf{X}} = \{\hat{X}_{t+1};\, t = 0, 1, \ldots\}$ and $\hat{\mathbf{P}} = \{\hat{P}_{t+1};\, t = 0, 1, \ldots\}$ can be described by dynamical recursions [2, pp. 38–39]. These recursions are given by the following coupled system

$$\begin{aligned} X^K_0 &= \mu \\ X^K_{t+1} &= A_t X^K_{t+1} - [A_t P^K_t H'_t + \Sigma^{wv}_{t+1}][H_t P^K_t H'_t + \Sigma^v_{t+1}]^{-1}[Y_t - H X^K_t] \\ & \qquad\qquad t = 0, 1, \ldots \quad (1.6) \end{aligned}$$

and

$$\begin{aligned} P^K_0 &= \Delta, \\ P^K_{t+1} &= A_t P^K_t A'_t + \Sigma^w_{t+1} \qquad\qquad t = 0, 1, \ldots \quad (1.7) \\ &\quad - [A_t P^K_t H'_t + \Sigma^{wv}_{t+1}][H_t P^K_t H'_t + \Sigma^v_{t+1}]^{-1}[A_t P^K_t H'_t + \Sigma^{wv}_{t+1}]'. \end{aligned}$$

Under assumption **(K)**, we have the following identities

$$\hat{X}_{t+1} = X^K_{t+1} \quad \text{and} \quad \hat{P}_{t+1} = P^K_{t+1}. \qquad t = 0, 1, \ldots \quad (1.8)$$

It will shortly become apparent why we have separately stated the *definitions* (1.3), (1.5) and the *recursions* (1.6)–(1.7).

The goal of this paper is a modest one—to relax assumption **(K)**, replacing it with the more realistic assumption **(A.3)** given by

(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 $(\mathbf{W}^\circ, \mathbf{V}^\circ)$. No other *a priori* assumptions, save these on the first two moments, are enforced on F.

We shall investigate in various ways how replacing assumption **(K)** by assumption **(A.3)** affects the solution of the prediction problem. The discussion below constitutes a synthesis of the material which has appeared in the three papers [28–30]. We hope that this account will provide a valuable complement to classical Kalman filtering, as the initial condition is in practice a rather vaguely–defined object about which only first and second moments are known.

The effect of replacing the classical assumption by **(A.3)** is dramatic; the rvs $\{X_{t+1}^\circ, Y_0, Y_1, \ldots, Y_t\}$ are *no longer* jointly Gaussian for each $t = 0, 1, \ldots$, and thus we cannot *a priori* expect that the conditional law of the the state given the observations can be described by any finite collection of sufficient statistics, e.g., $\hat{\mathbf{X}}$ and $\hat{\mathbf{P}}$ as in (1.3) and (1.5). We also note that in this more general case *(1.8) no longer holds; the conditional means and covariances of (1.3) and (1.5) no longer propagate according to (1.6) and (1.7).* Faced with this state of affairs, we might naturally seek to directly describe the *evolution* of the conditional law of the state given the observations, say by a straightforward use of Bayes' rule or via the discrete–time analogue of the celebrated Zakai equation of nonlinear filtering [7, 32]. This is not an easy task, however, for studying the evolution of this conditional law is tantamount to studying the evolution of an *infinite–dimensional* sufficient statistic. Moreover, under the suggested approaches, it seems quite difficult, if not impossible, to clearly follow, over time, the precise influence of the initial distribution. The key to overcoming these difficulties lies in the techniques of [22], where the filter is factored into a collection of *finite–dimensional and computable* sufficient statistics and a functional—the information of the observations is contained *solely* in these statistics, while the initial distribution appears *only* in the structure of the functional. These sufficient statistics obey recursions derived from filtering an auxiliary system of the type (1.1a)–(1.1b) under Kalman's original assumptions **(A.1)**–**(A.2)** and **(K)**. This provides us with a pleasing reaffirmation of the centrality of Kalman's results.

Aware of the demands of real applications, we shall not content ourselves only with a solution of the generalized one–step prediction problem, but shall address also the more practical question of *state estimation* (as in (1.3)). This will be of more direct interest to the engineer, who in general is not concerned with the entire conditional law, but rather with some appropriate estimate of the true state of the plant. For each $t = 0, 1, \ldots$, the estimation of the state X°_{t+1} given the observation σ–field $\mathcal{Y}_t$ is often defined as the problem of finding a Borel measurable $\varphi : (\mathbb{R}^k)^{t+1} \to \mathbb{R}^n$ which minimizes the mean–square error

$$\mathbf{E}^{\circ}[\|X^{\circ}_{t+1} - \varphi(Y_0, Y_1, \ldots, Y_t)\|^2 | \mathcal{Y}_t] \tag{1.9}$$

over some allowable class of Borel measurable functions. If we minimize (1.9) over all Borel measurable mappings φ, we get the *minimum mean square error* (or MMSE) estimate, while if we minimize (1.9) only over all *affine* mappings φ, we get the *linear least square error* (or LLSE) estimate. In fact, the MMSE and LLSE estimators are objects which have already been introduced: It is well known indeed that the MMSE estimators coincide with the sequence (1.3) of conditional means [2, Thm. 2.3.1], and that the LLSE estimator propagates according to (1.6)–(1.7), with $\mathbf{P}^{\mathbf{K}}$ being the sequence of corresponding error covariances [2, Sec. 5.4]. As we remarked in (1.8), under the Gaussian assumption **(K)**, $\hat{X}_{t+1} = X^K_{t+1}$ for all $t = 0, 1, \ldots$ and the MMSE and LLSE estimators coincide. But as soon as we pass to assumption **(A.3)**, the minimization of (1.9) over all Borel measurable mappings φ is *not* the same as the minimization of (1.9) over all affine mappings φ, so that in general the MMSE and LLSE estimators will *not* agree. The difference between the MMSE and LLSE estimators is a *direct consequence* of having a non–Gaussian initial condition.

We shall in this paper not only provide *computable* expressions for the MMSE and LLSE estimators, but also study properties of their difference. We are directed to this study for two reasons. Firstly, this difference, as we mentioned above, is a direct consequence of relaxing **(K)** to **(A.3)**. An understanding of this difference might be useful to the engineer, who, due to computing restrictions, often constructs the LLSE estimator as an approximation of the more accurate MMSE estimator. We shall study the MMSE–LLSE difference by considering the *mean–square error*

$$\varepsilon_t := \mathbf{E}^{\circ}[\|\hat{X}_t - X^K_t\|^2]. \qquad t = 1, 2, \ldots \tag{1.10}$$

After deriving a formula for (1.10), we shall then proceed to an *asymptotic* analysis of the sequence $\boldsymbol{\varepsilon} = \{\varepsilon_t;\, t = 1, 2, \ldots\}$ under the classical assumption of time–homogeneity of the plant and observation dynamics, and noise correlation structure. We shall in particular be interested in situations in which the LLSE estimates are *asymptotically the same* as the MMSE estimates, i.e., when $\lim_t \varepsilon_t = 0$. In these cases, we shall also give a more refined analysis of the *rate* of this convergence. These cases provide a formal justification of the idea widely held by practitioners that short of first and second moment information, precise distributional assumptions on the initial condition can be dispensed with when estimating the plant process on the basis of the observations. Under the assumption of time–homogeneity enforced in this asymptotic analysis, we may write

$$\varepsilon_t = \varepsilon_t\left((A, H, \Sigma), F\right), \qquad t = 1, 2, \ldots \quad (1.11)$$

where A, H and Σ are respectively the time–invariant state and observation gain matrices and noise correlation structure. For each $t = 0, 1, \ldots$, this representation displays the dependence of ε_t on the *system triple* (A, H, Σ) and on the initial distribution F, thus providing a natural organization of our study of the asymptotics of $\boldsymbol{\varepsilon}$. We hope that as a consequence of this analysis, there will emerge a much more precise understanding of the effects of the initial condition on the filtering of (1.1a)–(1.1b).

The reader familiar with more recent developments in filtering theory will already be acquainted with some of the tools used here, namely the Girsanov measure transformation and the Kallianpur–Striebel formula: The former will be used to define a new probability measure under which *explicit* calculations can be performed, while the latter will be invoked to relate the conditional expectations of (1.4) to corresponding conditional expectations under the new measure. In essence, the arguments in this paper amount to *pushing the nonlinear effects of the non–Gaussian initial condition into the probability measure.* To do this, we take as a pattern the techniques of [22] which solve the filtering problem in the corresponding continuous–time case when the plant and observation noises are uncorrelated. It will be of some interest, in fact, to see how the calculations of [22] can be generalized to handle *correlated* plant and observation noises. A pleasing discovery awaits us in that the *structure* of the solution of the prediction problem with correlated noise is essentially the *same* as that for

uncorrelated noise. The only difference lies in the *dynamics* of a collection of finite-dimensional sufficient statistics, while the *functional dependence* of the predictor upon these statistics remains the same as in the uncorrelated case.

The complete organization of this paper is as follows. Some notational conventions are collected in Section II for easy reference. Section III provides a review of the discrete-time Girsanov transform. In Section IV, we carry over to the discrete–time context the arguments developed in [22] for handling the continuous–time problem when the plant and observation noises are *uncorrelated* and the observation noise sequence V° is *standard.* We then show in Section V how to modify these ideas in order to solve the prediction problem in the case of correlated noises. In Section VI we use these results to obtain computable expressions for both the MMSE and LLSE estimators $\hat{\mathbf{X}}$ and $\mathbf{X}^{\mathbf{K}}$. We also apply the machinery developed thus far to give a formula for the error process $\boldsymbol{\varepsilon}$ of (1.10)—this formula is presented in Theorem 6.4. Section VII is devoted to a careful study of the asymptotics of the expression of Theorem 6.4, yielding our most general results about the asymptotics of $\boldsymbol{\varepsilon}$ in the multivariable case. Section VIII, a relatively short section, discusses a key technical result dealing with a partial converse for the asymptotics of $\boldsymbol{\varepsilon}$. We close with Section IX, which contains an even more complete asymptotic analysis in the scalar case (i.e., when $n = k = 1$), when many of the expressions of Sections V–VII can be simplified.

Several authors have considered various prediction, estimation and filtering problems for (1.1a)–(1.1b) under assumptions **(A.1)**–**(A.3)**; the continuous–time filtering problem has been studied in [3, 11, 19, 22]. Variations to the basic discrete–time model with a class of non–Gaussian white noises have been discussed in [20–21] with applications to failure detection. Related studies of the evolution of the conditional law of the plant given the observations as a measure–valued Markov process are given in [15–16].

We would like to point out that in this chapter we have studied *only* the one–step prediction problem. Of course, other estimation problems could have been considered, namely, the the so–called filtering and interpolation problems, which are respectively the problems of estimating the states X_t° and X_s°, $s = 0, 1, \ldots, t-1$, on the basis of $\mathcal{Y}_t$ for each $t = 0, 1, \ldots$. These

problems can be addressed with methods similar to those presented here. We have restricted our attention to the one–step prediction problem mainly for calculational convenience and since it is the natural analogue of the continuous–time filtering calculations [19, 22].

To the best of the authors' knowledge, few results have been reported in the literature on the large–time asymptotics of $\boldsymbol{\varepsilon}$ for a general non–Gaussian initial distribution. This may be due to the fact that the key representation result (Theorem 6.4) has been derived only relatively recently [27, 29].

II. NOTATION AND CONVENTIONS

For the sake of easy reference, we have collected here the various notation and conventions used throughout the paper:

The set of real numbers is denoted by $\mathbb{R}$, and $\mathbb{C}$ stands for the set of complex numbers. 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 positive integers n and m, we denote the space of $n \times m$ real matrices by $\mathcal{M}_{n\times m}$; let $O_{n\times m}$ denote the zero element in $\mathcal{M}_{n\times m}$. When $m = n$, we write $\mathcal{M}_n$ for the space $\mathcal{M}_{n\times m}$ of $n \times n$ real matrices, and we denote by $\mathcal{Q}_n$ the cone of $n \times n$ symmetric positive semi–definite matrices. We let I_n and O_n be the unit and zero elements in $\mathcal{M}_n$, respectively.

Elements of random or deterministic sequences will be set in regular type; the corresponding boldface character will denote the sequence itself. Examples which we have already introduced are the plant process $\mathbf{X}^\circ = \{X_t^\circ;\, t = 0, 1, \ldots\}$ and the covariance structure $\boldsymbol{\Sigma} = \{\Sigma_t;\, t = 1, 2, \ldots\}$.

For any matrix K in $\mathcal{M}_n$, with $\mathrm{sp}(K)$ denoting the set of all eigenvalues of K, we set $\lambda_{\min}(K) := \min\{|\lambda| : \lambda \in \mathrm{sp}(K)\}$ and $\lambda_{\max}(K) := \max\{|\lambda| : \lambda \in \mathrm{sp}(K)\}$; it is customary to call $\lambda_{\max}(K)$ the spectral radius of K and to denote it by $\rho(K)$. The mapping $\mathcal{M}_n \to \mathbb{R}_+$ given by

$$\|K\|_{op} := \sup_{v\neq 0} \frac{\|Kv\|}{\|v\|}, \quad K \in \mathcal{M}_n, \tag{2.1}$$

defines the operator norm on $\mathcal{M}_n$ induced by the Euclidean norm on $\mathbb{R}^n$. However, since $\mathcal{M}_n$ is a finite–dimensional Banach space, all norms on $\mathcal{M}_n$ are equivalent [10, Thm. IX.2.1]. This will be valuable in some of our limiting operations in the latter parts of this chapter, as we may safely take

entrywise limits. The following well–known facts about $\|\cdot\|_{op}$ will come in handy:

$$\|K\|_{op}^2 = \lambda_{\max}(K'K), \quad K \in \mathcal{M}_n, \tag{2.2}$$

and

$$\lambda_{\min}(K'K)\|v\|^2 \le \|Kv\|^2 \le \lambda_{\max}(K'K)\|v\|^2, \quad K \in \mathcal{M}_n,\ v \in \mathbb{R}^n. \tag{2.3}$$

The constant mapping $\mathbb{R}^n \to \mathbb{R} : x \to 1$ is denoted by $\mathbf{1}$.

For each matrix R in $\mathcal{Q}_n$, let G_R denote the distribution of a zero–mean $\mathbb{R}^n$–valued Gaussian rv with covariance R.

The following notation will be useful in our representation result for the conditional expectations of (1.4). For every S in $\mathcal{Q}_{2n}$, let X_S and B_S denote generic $\mathbb{R}^n$–valued rvs such that (X_S, B_S) is a $\mathbb{R}^{2n}$–valued zero–mean Gaussian rv with covariance matrix S. For every bounded Borel mapping $\varphi : \mathbb{R}^n \to \mathbb{C}$, we define the mappings $\mathcal{T}\varphi : \mathbb{R}^n \times \mathbb{R}^n \times \mathcal{Q}_{2n} \to \mathbb{C}$ and $\mathcal{U}\varphi : \mathbb{R}^n \times \mathbb{R}^n \times \mathcal{Q}_n \times \mathcal{M}_n \times \mathcal{Q}_{2n} \to \mathbb{C}$ by

$$\mathcal{T}\varphi[x, b; S] := \mathcal{E}\left[\varphi(x + X_S)\exp[b'B_S]\right], \quad x, b \in \mathbb{R}^n,\ S \in \mathcal{Q}_{2n}, \tag{2.4}$$

and

$$\mathcal{U}\varphi[x, b; \Lambda, \Psi; S] := \int_{\mathbb{R}^n} \mathcal{T}\varphi[x + \Psi z, z; S] \exp\left[b'z - \frac{1}{2}z'\Lambda z\right] dF(z), \quad x, b \in \mathbb{R}^n,\ \Lambda \in \mathcal{Q}_n,\ \Psi \in \mathcal{M}_n,\ S \in \mathcal{Q}_{2n}, \tag{2.5}$$

with the understanding that $\mathcal{E}$ denotes integration with respect to the Gaussian distribution of the rv (X_S, B_S).

We denote by $\mathcal{D}(\mathbb{R}^n)$ the set of all square–integrable probability distributions functions on $\mathbb{R}^n$ with positive definite (and thus invertible) covariance matrix, and by $\mathcal{D}^\circ(\mathbb{R}^n)$ the set of those distributions in $\mathcal{D}(\mathbb{R}^n)$ which have zero mean.

III. THE GIRSANOV TRANSFORMATION

Our efforts of Sections IV and V, where we derive expressions for the conditional expectations (1.4), will rely crucially on the celebrated Girsanov

measure transformations for GWN sequences [7, 9]. To streamline the arguments in Sections IV and V, we here summarize the properties of the Girsanov transform.

The essence of the Girsanov transformation is the *translation* of a Gaussian process. As evidence of the simple ideas at the heart of the Girsanov transformation, we begin with the fact that the probability law of a Gaussian rv with mean vector $m \neq 0$ and invertible covariance matrix R is absolutely continuous with respect to the law of a Gaussian rv with *zero* mean and covariance matrix R. It is easy to see that the corresponding Radon–Nikodym derivative is given by

$$\frac{dG_R(x-m)}{dG_R(x)} = \frac{\exp\left(-\frac{1}{2}(x-m)'R^{-1}(x-m)\right)}{\exp\left(-\frac{1}{2}x'R^{-1}x\right)}$$
$$= \exp\left(x'R^{-1}m - \frac{1}{2}m'R^{-1}m\right), \quad x \in \mathbb{R}^n;$$

we are in this simple calculation translating a Gaussian rv by a constant m. The discrete–time Girsanov measure transformation is conceptually very similar, but since we are dealing with *processes*, the class of allowable translates turns out to be much richer.

The basic framework for the discrete–time Girsanov transformation is as follows: The underlying probability space $(\Omega, \mathcal{F}, \mathbf{P}^\circ)$ is equipped with the filtration $\{\mathcal{F}_t;\, t = 0,1,\ldots\}$ of $\mathcal{F}$, i.e., $\{\mathcal{F}_t;\, t = 0,1,\ldots\}$ is an increasing family of sub–σ–fields of $\mathcal{F}$. Let $\mathbf{U} = \{U_t;\, t = 1,2,\ldots\}$ be an $\mathbb{R}^d$–valued zero–mean $(\mathcal{F}_t, \mathbf{P}^\circ)$–GWN sequence with correlation structure $\mathbf{\Lambda} = \{\Lambda_t;\, t = 1,2,\ldots\}$, i.e., for all $t = 0,1,\ldots$, the rv U_{t+1} is $\mathcal{F}_{t+1}$–measurable and

$$\mathbf{E}^\circ\left[\exp\left[i\theta' U_{t+1}\right]\middle|\mathcal{F}_t\right] = \exp\left[-\frac{1}{2}\theta'\Lambda_{t+1}\theta\right], \quad \theta \in \mathbb{R}^n.\ t = 0,1,\ldots \qquad (3.1)$$

For future reference, we note the well–known fact that (3.1) more generally holds for θ in $\mathbb{C}^n$.

A case of special interest arises when in this definition, the filtration is taken to be the natural filtration $\{\mathcal{F}_t^U;\, t = 0,1,\ldots\}$ induced by the sequence $\mathbf{U}$, i.e.,

$$\mathcal{F}_{t+1}^U := \sigma\{U_{s+1};\ s = 0,1,\ldots,t\} \qquad t = 0,1,\ldots \qquad (3.2)$$

with $\mathcal{F}_0^U$ chosen so that $\mathcal{F}_0^U \subset \mathcal{F}_1^U$; in fact, $\mathcal{F}_0^U$ is often selected to be the trivial σ–field on Ω. With the choice (3.2), we simply refer to $\mathbf{U}$ as a zero–mean $\mathbf{P}^\circ$–GWN sequence with correlation structure $\mathbf{\Lambda}$; moreover, when the reference probability measure $\mathbf{P}^\circ$ is clear from the context, we further omit it from the terminology. This is in agreement with usage in earlier sections, since (3.1) always implies that the rvs $\{U_t;\, t = 1, 2, \ldots\}$ are mutually independent. We also say that for any $T \geq 0$, a finite collection $\{U_t;\, t = 1, 2, \ldots, T+1\}$ of rvs is an $(\mathcal{F}_t, \mathbf{P}^\circ)$–GWN sequence if (3.1) holds for all $t = 0, 1, \ldots, T$.

Now, for a given $\mathcal{F}_t$–adapted $\mathbb{R}^d$–valued sequence $\boldsymbol{\chi} = \{\chi_t;\, t = 0, 1, \ldots\}$, we define the sequences $\bar{\mathbf{U}} = \{\bar{U}_t;\, t = 1, 2, \ldots\}$ and $\mathbf{L} = \{L_t;\, t = 0, 1, \ldots\}$ taking values in $\mathbb{R}^d$ and $\mathbb{R}$, respectively, by

$$\bar{U}_{t+1} := U_{t+1} - \Lambda_{t+1}\chi_t \qquad t = 0, 1, \ldots \quad (3.3)$$

and

$$L_0 = 1, \quad L_{t+1} := \prod_{s=0}^{t} \exp\left[\chi_s' U_{s+1} - \frac{1}{2}\chi_s' \Lambda_{s+1}\chi_s\right]. \quad t = 0, 1, \ldots \quad (3.4)$$

The first key fact that underlies the Girsanov transformation is given in the following Lemma:

Lemma 3.1. *The sequence* $\mathbf{L}$ *of positive rvs constitutes an* $(\mathcal{F}_t, \mathbf{P}^\circ)$*–martingale, with*

$$\mathbf{E}^\circ[L_t] = 1. \qquad t = 0, 1, \ldots \quad (3.5)$$

Proof. Fix $t = 0, 1, \ldots$. The rv L_{t+1} being positive, its (conditional) expectations are well defined, though not *a priori* finite. From the relation

$$L_{t+1} = L_t \cdot \exp\left[\chi_t' U_{t+1} - \frac{1}{2}\chi_t' \Lambda_{t+1}\chi_t\right], \quad (3.6)$$

we obtain

$$\mathbf{E}^\circ[L_{t+1}|\mathcal{F}_t] = L_t \cdot \exp\left[-\frac{1}{2}\chi_t' \Lambda_{t+1}\chi_t\right] \mathbf{E}^\circ\left[\exp\left[\chi_t' U_{t+1}\right]|\mathcal{F}_t\right] \quad (3.7)$$

as both the rvs L_t and χ_t are $\mathcal{F}_t$–measurable. Since the process $\mathbf{U}$ is a zero–mean $(\mathcal{F}_t, \mathbf{P}^\circ)$–GWN sequence with correlation structure $\mathbf{\Lambda}$, the

conditional distribution of U_{t+1} given $\mathcal{F}_t$ is that of a zero–mean Gaussian rv with covariance matrix Λ_{t+1}. Therefore, we have

$$\mathbf{E}^\circ\left[\exp\left[\chi_t' U_{t+1}\right]|\mathcal{F}_t\right] = \exp\left[\frac{1}{2}\chi_t'\Lambda_{t+1}\chi_t\right], \tag{3.8}$$

using again the $\mathcal{F}_t$–measurability of the rv χ_t. Substituting (3.8) into (3.7) we obtain the martingale property in the form

$$\mathbf{E}^\circ[L_{t+1}|\mathcal{F}_t] = L_t \quad \mathbf{P}^\circ - a.s.$$

whence

$$\mathbf{E}^\circ[L_{t+1}] = \mathbf{E}^\circ[L_t]. \tag{3.9}$$

The equality (3.5) is now a simple consequence of (3.9) and of the fact that $L_0 = 1$. ∎

For each fixed integer $T = 0, 1, \ldots$, we now define a measure $\bar{\mathbf{P}}_{T+1}$ on $(\Omega, \mathcal{F})$ by

$$\bar{\mathbf{P}}_{T+1}(A) := \int_A L_{T+1} d\mathbf{P}^\circ, \quad A \in \mathcal{F}. \tag{3.10}$$

It follows from (3.5) in Lemma 3.1 that the measure $\bar{\mathbf{P}}_{T+1}$ is indeed a probability measure on $\mathcal{F}$; we denote by $\bar{\mathbf{E}}_{T+1}$ the expectation operator associated with $\bar{\mathbf{P}}_{T+1}$. A simple martingale argument in conjunction with the fact that $L_0 = 1$ shows that $\bar{\mathbf{P}}_{T+1}$ agrees with $\mathbf{P}^\circ$ on $\mathcal{F}_0$. Also, since L_{T+1} is nonzero, we see that $\bar{\mathbf{P}}_{T+1}$ is *mutually absolutely continuous* with $\mathbf{P}^\circ$ on $\mathcal{F}$; the Radon–Nikodym derivatives are given by

$$\frac{d\bar{\mathbf{P}}_{T+1}}{d\mathbf{P}^\circ} = L_{T+1} \quad \text{and} \quad \frac{d\mathbf{P}^\circ}{d\bar{\mathbf{P}}_{T+1}} = L_{T+1}^{-1}.$$

As a consequence of this last fact, the statements $\mathbf{P}^\circ$–a.s. and $\bar{\mathbf{P}}_{T+1}$–a.s. are equivalent and reference to the underlying probability measure is usually dropped. Moreover, for notational convenience we omit a.s. equivalencies in the ensuing discussion, as such omissions have no effect upon the final results.

This change of measure which replaces the base measure $\mathbf{P}^\circ$ by $\bar{\mathbf{P}}_{T+1}$ is what is referred to in the literature as the Girsanov measure transformation. Its most important properties are summarized below.

Theorem 3.2. *With the notation and definitions given above, we have the following facts:*

(a) *The sequence* $\{\bar{U}_t;\ t = 1, 2, \ldots, T+1\}$ *is a zero–mean* $(\mathcal{F}_t, \bar{\mathbf{P}}_{T+1})$*–GWN process with covariance structure given by*

$$\bar{\mathbf{E}}_{T+1}[\bar{U}_{t+1}\bar{U}'_{t+1}|\mathcal{F}_t] = \Lambda_{t+1}; \qquad t = 0, 1, \ldots, T \quad (3.11)$$

and

(b) *The rvs* $\{L_t^{-1}, t = 0, 1, \ldots, T+1\}$ *form an* $(\mathcal{F}_t, \bar{\mathbf{P}}_{T+1})$*–martingale.*

Proof. The rvs $\{\bar{U}_t;\ t = 1, 2, \ldots, T+1\}$ are clearly $\mathcal{F}_t$–adapted. Therefore Claim (a) will be proved if we can show that

$$\bar{\mathbf{E}}_{T+1}\left[\exp\left[i\theta'\bar{U}_{t+1}\right]|\mathcal{F}_t\right] = \exp\left[-\frac{1}{2}\theta'\Lambda_{t+1}\theta\right], \quad \theta \in \mathbb{R}^n. \qquad t = 0, 1, \ldots, T \quad (3.12)$$

Fix $t = 0, 1, \ldots, T$ and θ in $\mathbb{R}^n$. Invoking a standard result on the evaluation of conditional expectations under an absolutely continuous change of measures [18, Sec. 27.4]—the so–called Kallianpur–Striebel formula of nonlinear filtering—we have

$$\bar{\mathbf{E}}_{T+1}\left[\exp\left[i\theta'\bar{U}_{t+1}\right]|\mathcal{F}_t\right] = \frac{\mathbf{E}^\circ\left[\exp\left[i\theta'\bar{U}_{t+1}\right]L_{T+1}|\mathcal{F}_t\right]}{\mathbf{E}^\circ\left[L_{T+1}|\mathcal{F}_t\right]} \quad (3.13)$$

and the remainder of the proof consists in evaluating the numerator and denominator of (3.13).

By the martingale property of Lemma 3.1, we first see that

$$\mathbf{E}^\circ[L_{T+1}|\mathcal{F}_t] = L_t. \quad (3.14)$$

Next, using the smoothing property of conditional expectations, we readily get

$$\begin{aligned}
&\mathbf{E}^\circ\left[\exp\left[i\theta'\bar{U}_{t+1}\right]L_{T+1}|\mathcal{F}_t\right] \\
&\quad = \mathbf{E}^\circ\left[\mathbf{E}^\circ\left[\exp\left[i\theta'\bar{U}_{t+1}\right]L_{T+1}|\mathcal{F}_{t+1}\right]|\mathcal{F}_t\right] \\
&\quad = \mathbf{E}^\circ\left[\exp\left[i\theta'\bar{U}_{t+1}\right]\mathbf{E}^\circ\left[L_{T+1}|\mathcal{F}_{t+1}\right]|\mathcal{F}_t\right] \quad (3.15) \\
&\quad = \mathbf{E}^\circ\left[\exp\left[i\theta'\bar{U}_{t+1}\right]L_{t+1}|\mathcal{F}_t\right] \quad (3.16) \\
&\quad = L_t \cdot \exp\left[-\frac{1}{2}\chi_t'\Lambda_{t+1}\chi_t\right]\mathbf{E}^\circ\left[\exp\left[i\theta'\bar{U}_{t+1} + \chi_t'U_{t+1}\right]|\mathcal{F}_t\right] \quad (3.17) \\
&\quad = L_t \cdot \exp\left[-\frac{1}{2}\chi_t'\Lambda_{t+1}\chi_t - i\theta'\Lambda_{t+1}\chi_t\right]\mathbf{E}^\circ\left[\exp\left[(i\theta + \chi_t)'U_{t+1}\right]|\mathcal{F}_t\right]. \quad (3.18)
\end{aligned}$$

The passage from (3.15) to (3.16) is validated by the martingale property of $\mathbf{L}$, while (3.17) follows from (3.16) upon using (3.6) and the $\mathcal{F}_t$–measurability of the rv χ_t; substituting (3.3) into (3.17) yields (3.18). Finally, using the first comment that followed (3.1), we see that

$$\mathbf{E}^\circ \left[\exp\left[(i\theta + \chi_t)' U_{t+1}\right] | \mathcal{F}_t\right] = \exp\left[-\frac{1}{2}(\theta - i\chi_t)'\Lambda_{t+1}(\theta - i\chi_t)\right]$$

and substitution of this last fact into (3.18) yields

$$\mathbf{E}^\circ \left[\exp\left[i\theta'\bar{U}_{t+1}\right] L_{T+1} | \mathcal{F}_t\right] = L_t \cdot \exp\left[-\frac{1}{2}\theta'\Lambda_{t+1}\theta\right]. \tag{3.19}$$

We now obtain (3.12) by inserting (3.14) and (3.19) into (3.13).

To establish Claim (b), we observe that the base measure $\mathbf{P}^\circ$ itself can be obtained from the transformed measure $\bar{\mathbf{P}}_{T+1}$ by a Girsanov measure transformation: Indeed, we rewrite (3.3) as

$$U_{t+1} = \bar{U}_{t+1} + \Lambda_{t+1}\chi_t \qquad\qquad t = 0, 1, \ldots$$

where it is now known by the first part of the proof that the sequence $\{\bar{U}_t;\, t = 1, 2, \ldots, T+1\}$ is a zero–mean $(\mathcal{F}_t, \bar{\mathbf{P}}_{T+1})$–GWN process with covariance structure (3.11). Therefore, in analogy with (3.4), with $\bar{U}_{t+1}$ and $-\Lambda_{t+1}\chi_t$ playing the role of U_{t+1} and $\Lambda_{t+1}\chi_t$, respectively, for all $t = 0, 1, \ldots, T$, we define the sequence $\bar{\mathbf{L}} = \{\bar{L}_t;\, t = 0, 1, \ldots\}$ of $\mathbb{R}$–valued rvs by

$$\bar{L}_0 = 1, \quad \bar{L}_{t+1} := \prod_{s=0}^{t} \exp\left[-\chi_s'\bar{U}_{s+1} - \frac{1}{2}\chi_s'\Lambda_{s+1}\chi_s\right]. \quad t = 0, 1, \ldots \tag{3.20}$$

By Lemma 3.1, the rvs $\{\bar{L}_t;\, t = 0, 1, \ldots T+1\}$ form an $(\mathcal{F}_t, \bar{\mathbf{P}}_{T+1})$–martingale, and the proof of Claim (b) is now completed upon observing that

$$\bar{L}_t = L_t^{-1}. \qquad\qquad t = 0, 1, \ldots \tag{3.21}$$

■

We conclude this section with an easy observation: An alternate expression for (3.4) is simply

$$L_{t+1} = \exp\left[\sum_{s=0}^{t}\left\{\chi_s'U_{s+1} - \frac{1}{2}\chi_s'\Lambda_{s+1}\chi_s\right\}\right] \qquad t = 0, 1, \ldots \tag{3.22}$$

and similarly from (3.20)–(3.21), we get

$$L_{t+1}^{-1} = \exp\left[\sum_{s=0}^{t}\left\{-\chi_s'\bar{U}_{s+1} - \frac{1}{2}\chi_s'\Lambda_{s+1}\chi_s\right\}\right]. \qquad t = 0, 1, \ldots$$

Theorem 3.2 implies that the probability measures $\{\bar{\mathbf{P}}_T;\ T = 1, 2, \ldots\}$ obey a *consistency property*—that if T and T' are finite times with $0 < T \leq T'$, then $\{\bar{U}_t;\ t = 1, 2, \ldots, T+1\}$ has the same statistics under both $\bar{\mathbf{P}}_{T+1}$ and $\bar{\mathbf{P}}_{T'+1}$—more completely, $\bar{\mathbf{P}}_{T+1}$ and $\bar{\mathbf{P}}_{T'+1}$ agree on $\mathcal{F}_0 \vee \sigma\{\bar{U}_t;\ t = 1, 2, \ldots, T+1\}$. The reader may then ask if, by setting $T = \infty$ in (3.10), we may find a single probability measure $\bar{\mathbf{P}}$ under which Claim (a) of Theorem 3.2 is true for *all* T. Unfortunately, the existence of such a 'projective limit' of the measures $\{\bar{\mathbf{P}}_T;\ T = 1, 2, \ldots\}$ is rare. In fact, such a probability measure will exist if and only if $\mathbf{L}$ is a uniformly integrable martingale; the reader is referred to [27, Thm. 2.1 and 23, Props. III-1-1 and IV-2-3] for a more detailed analysis of this question. The absence of such a limiting probability measure $\bar{\mathbf{P}}$ will not, however, cause any difficulties in our efforts here. We shall be considering the one–step prediction problem for (1.1a)–(1.1b) on *finite horizons*, which involve only finite subsets of the rvs $\{\xi, (W_t^\circ, V_t^\circ);\ t = 0, 1, \ldots\}$. We will touch upon this matter again in Section V.

IV. THE UNCORRELATED CASE: A REVIEW

We first review the solution of the one–step prediction problem (i.e., evaluating the conditional expectations (1.4)) in the simpler case when the plant and observation noise sequences are *uncorrelated* and the observation noise sequence is *standard*. In other words, we temporarily replace assumption **(A.1)** by assumption **(A.1)***, where

(A.1)*: The process $(\mathbf{W}^\circ, \mathbf{V}^\circ)$ is a zero–mean GWN sequence with covariance structure $\mathbf{\Sigma}$ given by

$$\Sigma_t := \mathrm{Cov}\begin{pmatrix} W_t^\circ \\ V_t^\circ \end{pmatrix} = \begin{pmatrix} \Sigma_t^w & O_{n\times k} \\ O_{k\times n} & I_k \end{pmatrix}. \qquad t = 1, 2, \ldots$$

The situation defined by **(A.1)*** is essentially the discrete–time analogue of the one discussed in [19, 22], and as was done there, the arguments will rely crucially on the Girsanov measure transformation; this time of course,

the discrete–time version, which was presented in Section III, will be used. As our purpose here is to provide some motivation and background for the more complicated arguments of Section V, we review below the various steps leading to the relevant discrete–time Girsanov transformation. In doing so, we are careful to explicitly point out the essential features of our reasoning.

We wish to emphasize once again that the *only* source of non–Gaussian randomness in the model comes from the initial condition ξ. If ξ were a Gaussian rv, then (1.8) would hold and the generalized one–step prediction problem would be fully described by (1.6)–(1.7). Furthermore, the MMSE and LLSE estimation processes $\hat{\mathbf{X}}$ and $\mathbf{X}^{\mathbf{K}}$ would coincide and the error process $\boldsymbol{\varepsilon}$ would be identically zero, whence the asymptotic analysis of Sections VI–IX would be trivial. Given this, our guiding principle is to effect a *decomposition* of (1.1a)–(1.1b) so that, as much as possible, we may separate the effects of the Gaussian white noise sequence $(\mathbf{W}^\circ, \mathbf{V}^\circ)$ from the troublesome effects of the non–Gaussian initial condition ξ.

We begin by noting that the closed–form solution of the recursive equation (1.1a) is simply

$$X_t^\circ = \Phi(t,0)\xi + \sum_{s=0}^{t-1} \Phi(t,s) W_{s+1}^\circ \qquad t = 1,2,\ldots$$

where Φ is the state transition matrix defined by

$$\Phi(s,s) = I_n, \quad \Phi(t+1,s) = A_t \Phi(t,s), \quad s \le t. \qquad s,t = 0,1,\ldots \tag{4.1}$$

This suggests the decomposition

$$X_t^\circ = Z_t + X_t \qquad t = 1,2,\ldots \tag{4.2}$$

where the processes $\mathbf{Z}$ and $\mathbf{X}$ are given by

$$Z_0 = \xi, \quad Z_t = \Phi(t,0)\xi \qquad t = 1,2,\ldots \tag{4.3}$$

and

$$X_0 = 0, \quad X_t = \sum_{s=0}^{t-1} \Phi(t,s) W_{s+1}^\circ. \qquad t = 1,2,\ldots \tag{4.4}$$

The effects of the non–Gaussian initial condition ξ are encoded in the process $\mathbf{Z}$, while $\mathbf{X}$ is a Gaussian sequence. For future reference, we note that

the evolution of the processes $\mathbf{Z}$ and $\mathbf{X}$ are also described by the dynamical equations

$$Z_0 = \xi, \quad Z_{t+1} = A_t Z_t \qquad t = 0, 1, \ldots \quad (4.5)$$

and

$$X_0 = 0, \quad X_{t+1} = A_t X_t + W^{\circ}_{t+1}. \qquad t = 0, 1, \ldots \quad (4.6)$$

Using the decomposition (4.2), we can write (1.1b) as

$$\begin{aligned} Y_t &= H_t X_t + H_t Z_t + V^{\circ}_{t+1} \\ &= H_t X_t + V_{t+1} \qquad t = 0, 1, \ldots \end{aligned} \quad (4.7)$$

where we have set

$$V_{t+1} = V^{\circ}_{t+1} + H Z_t. \qquad t = 0, 1, \ldots$$

Therefore the observation process $\mathbf{Y}$ is the sum of the $\mathbf{P}^{\circ}$–*Gaussian* process $\{H_t X_t;\, t = 0, 1, \ldots\}$ and of the sequence $\mathbf{V}$ which can be interpreted as the *translation* of the $\mathbf{P}^{\circ}$–GWN sequence $\mathbf{V}^{\circ}$ by the process $\{H_t Z_t;\, t = 0, 1, \ldots\}$. This simple observation suggests that after an appropriate Girsanov transformation to be determined, the noise sequence $\mathbf{V}$ can be made to look like a GWN sequence under the transformed measure.

Our next step consists of using the Girsanov measure transformation to see that the law of the *translated* Gaussian process $\mathbf{V}$ is absolutely continuous with respect to the law of a *centered* Gaussian process; this will be made more precise in a moment. As the end result of this measure transformation, we can consider a *new* probability measure under which $\mathbf{V}$ (as opposed to $\mathbf{V}^{\circ}$) is now a zero–mean standard GWN. In short, *the effects of the non–Gaussian initial condition have been pushed into a Radon–Nikodym derivative.*

For the situation at hand, the processes $\mathbf{V}^{\circ}$ and $\{H_t Z_t;\, t = 0, 1, \ldots\}$ play the role of the processes $\mathbf{U}$ and $\boldsymbol{\chi}$ of Section III, respectively. To complete the preparations for the Girsanov transformation, we introduce the filtration $\{\mathcal{F}_t;\, t = 0, 1, \ldots\}$ by setting

$$\mathcal{F}_0 := \sigma\{\xi, W^{\circ}_s;\, s = 1, 2, \ldots\}$$

and

$$\mathcal{F}_t := \mathcal{F}_0 \vee \sigma\{V^{\circ}_s;\, s = 1, 2, \ldots, t\}. \qquad t = 1, 2, \ldots$$

Since the processes $\mathbf{W}^\circ$ and $\mathbf{V}^\circ$ are uncorrelated (and thus independent since Gaussian), $\mathbf{V}^\circ$ will indeed be a zero–mean $(\mathcal{F}_t, \mathbf{P}^\circ)$–GWN sequence. (When considering the correlated case in Section V, $\mathbf{V}^\circ$ will not be a $(\mathcal{F}_t, \mathbf{P}^\circ)$–GWN sequence for this definition of $\mathcal{F}_t$. This will be the main hurdle we shall confront.)

Using (3.22), we see that the sequence $\mathbf{L}$ of Radon–Nikodym derivatives of (3.4) is here given by $L_0 := 1$ and

$$L_{t+1} \qquad\qquad t = 0, 1, \ldots \quad (4.8)$$

$$:= \exp\left[-\xi' \sum_{s=0}^{t} [H_s\Phi(s,0)]' V_{s+1}^\circ - \frac{1}{2}\xi' \sum_{s=0}^{t} [H_s\Phi(s,0)]'[H_s\Phi(s,0)]\xi\right].$$

For any nonnegative integer T, we then define the probability measure $\bar{\mathbf{P}}_{T+1}$ by

$$\bar{\mathbf{P}}_{T+1}(A) := \int_A L_{T+1} d\mathbf{P}^\circ, \quad A \in \mathcal{F}. \qquad (4.9)$$

Applying the results of Section III, we readily conclude the following facts:

(G) *The $\bar{\mathbf{P}}_{T+1}$–statistics of the rvs $\{\xi, W_{r+1};\ r = 0, 1, \ldots, V_{s+1};\ s = 0, 1, \ldots, T\}$ are the same as the $\mathbf{P}^\circ$–statistics of the rvs $\{\xi, W_{r+1}^\circ;\ r = 0, 1, \ldots, V_{s+1}^\circ;\ s = 0, 1, \ldots, T\}$. In particular, under the transformed measure $\bar{\mathbf{P}}_{T+1}$, the rv ξ has distribution F and is independent of the rvs $\{W_{r+1}^\circ;\ r = 0, 1, \ldots;\ V_{s+1};\ s = 0, 1, \ldots, T\}$ which are zero–mean Gaussian rvs with known covariance structure.*

We can summarize thus far the effects of the decomposition (4.2) and of the Girsanov transformation (4.8)–(4.9): The observation process $\mathbf{Y}$ can be viewed as the sum of a *Gaussian* process and of a *translated* Gaussian process, this translate being amenable to a Girsanov transformation which results in property (G).

We now turn to the evaluation of the conditional expectation (1.4) for some fixed time index $t = 0, 1, \ldots$, and for some given bounded Borel mappings $\varphi : \mathbb{R}^n \to \mathbb{C}$. Fixing the time horizon T so that $t \leq T$, we consider the change of measure defined by (4.9) and seek to evaluate (1.4) by performing the calculations under the transformed measure $\bar{\mathbf{P}}_{T+1}$ rather than under $\mathbf{P}^\circ$. To do this, as in Section III, we resort to the Kallianpur–Striebel formula [18, Sec. 27.4], which here takes the form

$$\mathbf{E}^\circ[\varphi(X_{t+1}^\circ)|\mathcal{Y}_t] = \frac{\bar{\mathbf{E}}_{T+1}[\varphi(X_{t+1}^\circ)L_{T+1}^{-1}|\mathcal{Y}_t]}{\bar{\mathbf{E}}_{T+1}[L_{T+1}^{-1}|\mathcal{Y}_t]} \qquad (4.10)$$

since

$$\frac{d\mathbf{P}^\circ}{d\bar{\mathbf{P}}_{T+1}} = L_{T+1}^{-1}.$$

Our problem thus reduces to the evaluation of expressions of the form

$$\bar{\mathbf{E}}_{T+1}[\varphi(X_{t+1}^\circ)L_{T+1}^{-1}|\mathcal{Y}_t]$$

for all bounded Borel mappings $\varphi : \mathbb{R}^n \to \mathbb{C}$.

First, an easy preliminary simplification: The rv X_{t+1}° is $\mathcal{F}_{t+1}$–measurable and $\mathcal{Y}_t$ is also contained in $\mathcal{F}_{t+1}$, whence iterated conditioning and the martingale property of Claim (b) in Theorem 3.2 readily imply the equality

$$\begin{aligned}\bar{\mathbf{E}}_{T+1}[\varphi(X_{t+1}^\circ)L_{T+1}^{-1}|\mathcal{Y}_t] &= \bar{\mathbf{E}}_{T+1}[\bar{\mathbf{E}}_{T+1}[\varphi(X_{t+1}^\circ)L_{t+1}^{-1}|\mathcal{F}_t]|\mathcal{Y}_t] \\ &= \bar{\mathbf{E}}_{T+1}[\varphi(X_{t+1}^\circ)L_{t+1}^{-1}|\mathcal{Y}_t]. \qquad (4.11)\end{aligned}$$

By another argument based on iterated conditioning, we also observe that

$$\begin{aligned}&\bar{\mathbf{E}}_{T+1}[\varphi(X_{t+1}^\circ)L_{t+1}^{-1}|\mathcal{Y}_t] \\ &\quad = \bar{\mathbf{E}}_{T+1}[\bar{\mathbf{E}}_{T+1}[\varphi(X_{t+1}^\circ)L_{t+1}^{-1}|\mathcal{Y}_t \vee \sigma\{\xi\}]|\mathcal{Y}_t]. \qquad (4.12)\end{aligned}$$

The importance of this formula stems from the fact that *under* $\bar{\mathbf{P}}_{T+1}$, *the process* ξ *has known statistics and is independent of* $\mathcal{Y}_t$. (Reviewing our arguments, we see that one essential feature of the decomposition (4.2) and of the Girsanov transformation was that it allowed the non–Gaussian effects of ξ to be put into the oservation noise. Thus the randomness in the observations comes from the Gaussian plant noise and the observation noise, which is also Gaussian under the new measure. Making ξ measurable with respect to $\mathcal{F}_0$ makes ξ independent of these noise sequences and thus of the process $\mathbf{Y}$.)

To proceed with the evaluation of the inner conditional expectation on the right–hand side of (4.12), we shall write (4.8) more compactly. We define the $\mathbb{R}^n$–valued process $\mathbf{B} = \{B_t;\ t = 0, 1, \ldots\}$ by

$$B_0 := 0, \quad B_{t+1} := \sum_{s=0}^{t} [H_s\Phi(s,0)]'V_{s+1} \qquad t = 0, 1, \ldots \quad (4.13)$$

and the $\mathcal{Q}_n$–valued sequence $\mathbf{M} = \{M_t;\ t = 0, 1, \ldots\}$ by

$$M_0 := O_n, \quad M_{t+1} := \sum_{s=0}^{t} [H_s\Phi(s,0)]'[H_s\Phi(s,0)]. \qquad t = 0, 1, \ldots \quad (4.14)$$

With this notation, (4.8) now becomes

$$L_{t+1}^{-1} = \exp\left[\xi' B_{t+1} - \frac{1}{2}\xi' M_{t+1}\xi\right] \qquad t = 0, 1, \ldots \quad (4.15)$$

so that, combining (4.2)–(4.3), we get

$$\begin{aligned} &\bar{\mathbf{E}}_{T+1}\left[\varphi(X_{t+1}^{\circ})L_{t+1}^{-1}\middle|\mathcal{Y}_t \vee \sigma\{\xi\}\right] \\ &\quad = \exp\left[-\frac{1}{2}\xi' M_{t+1}\xi\right] \qquad (4.16) \\ &\qquad \cdot \bar{\mathbf{E}}_{T+1}\left[\varphi(X_{t+1} + \Phi(t+1,0)\xi)\exp[\xi' B_{t+1}]\middle|\mathcal{Y}_t \vee \sigma\{\xi\}\right]. \end{aligned}$$

Returning to (**G**), which describes the statistics of the relevant rvs under $\bar{\mathbf{P}}_{T+1}$, we argue that the evaluation of (4.16) is conceptually a simple matter: Indeed, from (**G**) we observe that the rvs $\{X_{t+1}, B_{t+1}, Y_0, Y_1, \ldots, Y_t\}$ can be expressed as linear combinations of the rvs $\{W_{s+1}^{\circ};\ s = 0, 1, \ldots, V_{r+1};\ r = 0, 1, \ldots, T\}$, and under $\bar{\mathbf{P}}_{T+1}$ are thus *jointly Gaussian* and *independent* of ξ. Therefore, as a first consequence of these facts, we can write

$$\begin{aligned} &\bar{\mathbf{E}}_{T+1}\left[\varphi(X_{t+1} + \Phi(t+1,0)\xi)\exp[\xi' B_{t+1}]\middle|\mathcal{Y}_t \vee \sigma\{\xi\}\right] \\ &\quad = \bar{\mathbf{E}}_{T+1}\left[\varphi(X_{t+1} + \Phi(t+1,0)z)\exp[z' B_{t+1}]\middle|\mathcal{Y}_t\right]_{z=\xi}. \qquad (4.17) \end{aligned}$$

Next, defining the conditional expectations

$$\bar{X}_{t+1} := \bar{\mathbf{E}}_{T+1}\left[X_{t+1}\middle|\mathcal{Y}_t\right] \quad \text{and} \quad \bar{B}_{t+1} := \bar{\mathbf{E}}_{T+1}\left[B_{t+1}\middle|\mathcal{Y}_t\right] \qquad (4.18)$$

with corresponding errors

$$\tilde{X}_{t+1} := X_{t+1} - \bar{X}_{t+1} \quad \text{and} \quad \tilde{B}_{t+1} := B_{t+1} - \bar{B}_{t+1}, \qquad (4.19)$$

we see that

$$\begin{aligned} &\bar{\mathbf{E}}_{T+1}\left[\varphi(X_{t+1} + \Phi(t+1,0)z)\exp[z' B_{t+1}]\middle|\mathcal{Y}_t\right] \\ &\quad = \exp\left[z'\bar{B}_{t+1}\right] \\ &\qquad \cdot \bar{\mathbf{E}}_{T+1}\left[\varphi(\tilde{X}_{t+1} + x + \Phi(t+1,0)z)\exp\left[z'\tilde{B}_{t+1}\right]\middle|\mathcal{Y}_t\right]_{x=\bar{X}_{t+1}}, \\ &\qquad\qquad z \in \mathbb{R}^n. \quad (4.20) \end{aligned}$$

The evaluation of (4.20) requires only the $\bar{\mathbf{P}}_{T+1}$–conditional distribution of the pair $(\tilde{X}_{t+1}, \tilde{B}_{t+1})$ given the Gaussian rvs $\{Y_0, Y_1, \ldots, Y_t\}$. Since the rvs $\{X_{t+1}, B_{t+1}, Y_0, Y_1, \ldots, Y_t\}$ are jointly $\bar{\mathbf{P}}_{T+1}$–Gaussian, it is well known [6, p. 10] that under $\bar{\mathbf{P}}_{T+1}$, the pair $(\tilde{X}_{t+1}, \tilde{B}_{t+1})$ is *independent* of the rvs $\{Y_0, Y_1, \ldots, Y_t\}$ and has a Gaussian distribution distribution with zero mean and covariance matrix S_{t+1} given by

$$S_{t+1} := \bar{\mathbf{E}}_{T+1}\left[\begin{pmatrix}\tilde{X}_{t+1}\\ \tilde{B}_{t+1}\end{pmatrix}\begin{pmatrix}\tilde{X}_{t+1}\\ \tilde{B}_{t+1}\end{pmatrix}'\right]. \tag{4.21}$$

Therefore, using this fact in (4.20), we find

$$\begin{aligned}&\bar{\mathbf{E}}_{T+1}\left[\varphi(X_{t+1} + \Phi(t+1,0)z)\exp[z'B_{t+1}]\,\middle|\,\mathcal{Y}_t\right]\\ &\quad = \exp\left[z'\bar{B}_{t+1}\right]\mathcal{T}\varphi[\bar{X}_{t+1} + \Phi(t+1,0)z, z; S_{t+1}], \quad z \in \mathbb{R}^n,\end{aligned}$$

where we have used the notation (2.4). It is now plain from (4.17) that

$$\begin{aligned}&\bar{\mathbf{E}}_{T+1}\left[\varphi(X_{t+1} + \Phi(t+1,0)\xi)\exp[z'B_{t+1}]\,\middle|\,\mathcal{Y}_t \vee \sigma\{\xi\}\right]\\ &\quad = \exp\left[\xi'\bar{B}_{t+1}\right]\mathcal{T}\varphi[\bar{X}_{t+1} + \Phi(t+1,0)\xi, \xi; S_{t+1}],\end{aligned}$$

and going back to (4.16), we can now conclude that

$$\begin{aligned}&\bar{\mathbf{E}}_{T+1}\left[\varphi(X^\circ_{t+1})L^{-1}_{t+1}\,\middle|\,\mathcal{Y}_t \vee \sigma\{\xi\}\right]\\ &\quad = \exp\left[-\frac{1}{2}\xi' M_{t+1}\xi + \xi'\bar{B}_{t+1}\right]\\ &\qquad\qquad \cdot \mathcal{T}\varphi[\bar{X}_{t+1} + \Phi(t+1,0)\xi, \xi; S_{t+1}].\end{aligned} \tag{4.22}$$

Finally, by averaging over ξ as indicated by (4.12), we get the relation

$$\bar{\mathbf{E}}_{T+1}[\varphi(X^\circ_{t+1})L^{-1}_{t+1}|\mathcal{Y}_t] = \mathcal{U}\varphi[\bar{X}_{t+1}, \bar{B}_{t+1}; M_{t+1}, \Phi(t+1,0); S_{t+1}] \tag{4.23}$$

where we have used the notation (2.5).

Combining (4.10) and (4.23), we obtain the representation result

$$\mathbf{E}^\circ[\varphi(X^\circ_{t+1})|\mathcal{Y}_t] = \frac{\mathcal{U}\varphi[\bar{X}_{t+1}, \bar{B}_{t+1}; M_{t+1}, \Phi(t+1,0); S_{t+1}]}{\mathcal{U}\mathbf{1}[\bar{X}_{t+1}, \bar{B}_{t+1}; M_{t+1}, \Phi(t+1,0); S_{t+1}]}. \tag{4.24}$$

This formula is the discrete–time analogue of Theorem T5 in [22]. To finish the one–step prediction problem, we only need to calculate $\bar{X}_{t+1}$,

$\bar{B}_{t+1}$, M_{t+1}, $\Phi(t+1,0)$ and S_{t+1}. We shall perform these computations in the more general case of Section V. For the reader eager to carry out in totality the calculations of this section, we note that the $\bar{\mathbf{P}}_{T+1}$–Gaussian rvs $\{(X_{t+1}, B_{t+1});\, t = 0,1,\ldots,T\}$ described by (4.6) and (4.13), obey linear dynamics driven by a $\bar{\mathbf{P}}_{T+1}$–GWN sequence, and that for each $t = 0,1,\ldots,T$, the observation Y_t is a linear combination of the rv (X_{t+1}, B_{t+1}) and of the $\bar{\mathbf{P}}_{T+1}$–Gaussian noise term V_{t+1}. Therefore, classical Kalman filtering theory applies and leads to recursive equations for the rvs $(\bar{X}_{t+1}, \bar{B}_{t+1})$ and the error covariance matrices S_{t+1}, $t = 0,1,\ldots,T$.

Before closing this section, let us continue the calculation of (4.22) in the specific case when $\varphi = \mathbf{1}$; this will be needed in Section VI. Straightforward evaluations yield

$$\bar{\mathbf{E}}_{T+1}\left[L_{t+1}^{-1}\middle|\mathcal{Y}_t \vee \sigma\{\xi\}\right]$$
$$= \exp\left[-\frac{1}{2}\xi'\left(M_{t+1} - \bar{\mathbf{E}}_{T+1}[\tilde{B}'_{t+1}\tilde{B}_{t+1}]\right)\xi + \xi'\bar{B}_{t+1}\right]. \quad (4.25)$$

V. THE CORRELATED CASE

We now turn to the more complicated case where the noise sequences $\mathbf{W}^\circ$ and $\mathbf{V}^\circ$ are allowed to be correlated and $\mathbf{V}^\circ$ to be nonstandard, i.e., we are returning to the general assumption **(A.1)**.

First, a few comments to guide the discussion of this more complex situation: We seek again a decomposition of the form (4.2), with the objective of separating the effects of the non–Gaussian initial condition ξ from those of the GWN sequence $(\mathbf{V}^\circ, \mathbf{W}^\circ)$. However, as we review the arguments of Section IV, we readily see the main difficulty in arguing as was done there on the basis of the decomposition (4.2)–(4.4): *Since* $\mathbf{W}^\circ$ *and* $\mathbf{V}^\circ$ *are correlated, we cannot expect in general that a Girsanov transformation will change the statistics of* $\mathbf{V}$ *without changing the statistics of* $\mathbf{W}^\circ$*!* As a result, under the Girsanov transformation (4.8)–(4.9) based on the decomposition (4.2)–(4.4), the process $\mathbf{X}$ defined by (4.4) will probably not retain its Gaussian character and the arguments of Section IV cannot be carried through.

The remedy to this difficulty is easily seen: We shall perform a Girsanov transformation on the *joint* $\mathbb{R}^{n+k}$–valued process $(\mathbf{W}^\circ, \mathbf{V}^\circ)$, instead

of the GWN process $\mathbf{V}^\circ$ alone (as was done in Section IV). We thus look for a new probability measure under which an appropriate translate of the joint process $(\mathbf{W}^\circ, \mathbf{V}^\circ)$ is Gaussian (up to a finite horizon). Of course, the search for this Girsanov transformation will be initiated via a decomposition of the form (4.2), where this time, the processes $\mathbf{X}$ and $\mathbf{Z}$ are yet to be determined, i.e., we do not *a priori* make the definitions (4.3) and (4.4) in the postulated decomposition

$$X_t^\circ = X_t + Z_t. \qquad t = 0, 1, \ldots \quad (5.1)$$

With this in mind, from (1.1a) and (5.1), we first obtain

$$\begin{aligned} & X_{t+1} + Z_{t+1} \\ & \quad = X_{t+1}^\circ \\ & \quad = A_t X_t^\circ + W_{t+1}^\circ \\ & \quad = A_t X_t + A_t Z_t + W_{t+1}^\circ. \qquad t = 0, 1, \ldots \quad (5.2) \end{aligned}$$

Prompted by the remarks above, we tentatively define the $\mathbb{R}^n$–valued processes $\mathbf{X}$ and $\mathbf{Z}$ via the recursions

$$X_0 = 0, \quad X_{t+1} = A_t X_t - \pi_t + W_{t+1}^\circ \qquad t = 0, 1, \ldots \quad (5.3)$$

and

$$Z_0 = \xi, \quad Z_{t+1} = A_t Z_t + \pi_t \qquad t = 0, 1, \ldots \quad (5.4)$$

for some $\mathbb{R}^n$–valued process $\boldsymbol{\pi} = \{\pi_t;\, t = 0, 1, \ldots\}$; these equations are compatible with (5.2) and generalize (4.5) and (4.6). Equation (4.7), being a direct consequence of the decomposition (4.2), still holds in our present case, i.e.,

$$Y_t = H_t X_t + V_{t+1} \qquad t = 0, 1, \ldots$$

where we have again defined the sequence $\mathbf{V}$ by

$$V_{t+1} = V_{t+1}^\circ + H_t Z_t. \qquad t = 0, 1, \ldots \quad (5.5)$$

Moreover, we observe that (5.3) can also be rewritten as

$$X_0 = 0, \quad X_{t+1} = A_t X_t + W_{t+1} \qquad t = 0, 1, \ldots$$

if the $\mathbb{R}^n$–valued sequence $\mathbf{W}$ is defined by

$$W_{t+1} = W_{t+1}^\circ - \pi_t. \qquad t = 0, 1, \ldots \quad (5.6)$$

Now, going back to the line of reasoning given in Section IV (as amended above), we must perform a Girsanov transformation on the $\mathbb{R}^{n+k}$–valued process $(\mathbf{W}^\circ, \mathbf{V}^\circ)$. The particular form of (5.5) and (5.6) suggests that the relevant Girsanov transformation is the one associated with translating the GWN sequence $(\mathbf{W}^\circ, \mathbf{V}^\circ)$ into the $\mathbb{R}^{n+k}$–valued process $(\mathbf{W}, \mathbf{V})$. We do this as follows: First we define the filtration $\{\mathcal{F}_t;\ t = 0, 1, \ldots\}$ on $(\Omega, \mathcal{F})$ by $\mathcal{F}_0 := \sigma\{\xi\}$ and

$$\mathcal{F}_t := \sigma\{\xi, (W_s^\circ, V_s^\circ);\ s = 1, 2, \ldots, t\}. \qquad t = 1, 2, \ldots$$

Under the enforced independence assumption in **(A.3)**, the sequence $(\mathbf{W}^\circ, \mathbf{V}^\circ)$ is indeed an $(\mathcal{F}_t, \mathbf{P}^\circ)$–GWN sequence. The feasibility of the translation mentioned above will be established if we can find two $\mathcal{F}_t$–adapted sequences $\boldsymbol{\psi}^{\mathbf{w}} = \{\psi_t^w;\ t = 0, 1, \ldots\}$ and $\boldsymbol{\psi}^{\mathbf{v}} = \{\psi_t^v;\ t = 0, 1, \ldots\}$ taking values in $\mathbb{R}^n$ and $\mathbb{R}^k$, respectively, such that

$$\begin{pmatrix} W_{t+1} \\ V_{t+1} \end{pmatrix} = \begin{pmatrix} W_{t+1}^\circ \\ V_{t+1}^\circ \end{pmatrix} - \begin{pmatrix} \Sigma_{t+1}^w & \Sigma_{t+1}^{wv} \\ \Sigma_{t+1}^{vw} & \Sigma_{t+1}^v \end{pmatrix} \begin{pmatrix} \psi_t^w \\ \psi_t^v \end{pmatrix}. \qquad t = 0, 1, \ldots \quad (5.7)$$

Upon comparing this last relation with (5.5) and (5.6), we readily see that the processes $\boldsymbol{\psi}^{\mathbf{w}}$ and $\boldsymbol{\psi}^{\mathbf{v}}$ have to be selected so that

$$-\Sigma_{t+1}^w \psi_t^w - \Sigma_{t+1}^{wv} \psi_t^v = -\pi_t \qquad t = 0, 1, \ldots$$

and

$$-\Sigma_{t+1}^{vw} \psi_t^w - \Sigma_{t+1}^v \psi_t^v = H_t Z_t. \qquad t = 0, 1, \ldots$$

Since Σ_{t+1}^v is invertible, this can be achieved by choosing some as–yet–unspecified $\mathbb{R}^n$–valued sequence $\boldsymbol{\psi} = \{\psi_t;\ t = 0, 1, \ldots\}$ which is $\mathcal{F}_t$–adapted, and by taking the processes $\boldsymbol{\psi}^{\mathbf{w}}$ and $\boldsymbol{\psi}^{\mathbf{v}}$ such that

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

With this choice, the process $\boldsymbol{\pi}$ is given by

$$\begin{aligned} \pi_t &= \Sigma_{t+1}^w \psi_t^w + \Sigma_{t+1}^{wv} \psi_t^v \\ &= [\Sigma_{t+1}^w - \Sigma_{t+1}^{wv} (\Sigma_{t+1}^v)^{-1} \Sigma_{t+1}^{vw}] \psi_t \\ &\quad - \Sigma_{t+1}^{wv} (\Sigma_{t+1}^v)^{-1} H_t Z_t. \qquad t = 0, 1, \ldots \end{aligned}$$

To simplify the Girsanov transformation as much as possible, we take

$$\psi_t = 0, \qquad t = 0, 1, \ldots$$

so that

$$\psi_t^w = 0 \quad \text{and} \quad \psi_t^v = -(\Sigma_{t+1}^v)^{-1} H_t Z_t \qquad t = 0, 1, \ldots \quad (5.8)$$

with

$$\pi_t = -\Sigma_{t+1}^{wv} (\Sigma_{t+1}^v)^{-1} H_t Z_t. \qquad t = 0, 1, \ldots \quad (5.9)$$

Inserting these choices (5.8)–(5.9) into (5.3)–(5.4) and (5.7) we have the following summary of our decomposition:

- **The effect of the initial condition**

$$Z_0 = \xi, \quad Z_{t+1} = [A_t - \Sigma_{t+1}^{wv} (\Sigma_{t+1}^v)^{-1} H_t] Z_t \qquad t = 0, 1, \ldots \quad (5.10)$$

- **The noise processes**

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

- **The auxiliary system**

$$X_0 = 0, \quad X_{t+1} = A_t X_t + W_{t+1} \quad t = 0, 1, \ldots \quad (5.11)$$

and

$$Y_t = H_t X_t + V_{t+1}. \quad t = 0, 1, \ldots \quad (5.12)$$

The Girsanov measure transformation associated with (5.7), under the choice (5.8), will be effected by the sequence $\mathbf{L}$ of Radon–Nykodym derivatives given here by $L_0 := 1$ and

$$L_{t+1} \qquad t = 0, 1, \ldots$$
$$:= \exp\left[-\sum_{s=0}^{t} [H_s Z_s]' (\Sigma_{s+1}^v)^{-1} V_{s+1}^\circ - \frac{1}{2} \sum_{s=0}^{t} [H_s Z_s]' (\Sigma_{s+1}^v)^{-1} [H_s Z_s] \right].$$

This expression follows from Section III with the identification

$$\chi_t = \begin{pmatrix} 0 \\ -(\Sigma_{t+1}^v)^{-1} H_t Z_t \end{pmatrix} \quad \text{and} \quad \Lambda_{t+1} = \Sigma_{t+1}. \qquad t = 0, 1, \ldots$$

Next we repeat our arguments of Section IV to evaluate the conditional expectation (1.4) for some fixed time index $t = 0, 1, \ldots$, and for some given bounded Borel mappings $\varphi : \mathbb{R}^n \to \mathbb{C}$. For each $T = 0, 1, \ldots$, we define the probability measure $\bar{\mathbf{P}}_{T+1}$ through the formula $d\bar{\mathbf{P}}_{T+1}/d\mathbf{P}^\circ = L_{T+1}$—the same formula as (4.9). Fix the time horizon T so that $t \leq T$. We seek to evaluate (1.4) by performing the calculations under the transformed measure $\bar{\mathbf{P}}_{T+1}$ rather than under $\mathbf{P}^\circ$. To do this, we observe that the following property ($\mathbf{G}^*$) holds, where

($\mathbf{G}^*$) *The $\bar{\mathbf{P}}_{T+1}$–statistics of $\{\xi, (W_{s+1}, V_{s+1}); s = 0, 1, \ldots, T\}$ are the same as the $\mathbf{P}^\circ$–statistics of $\{\xi, (W^\circ_{s+1}, V^\circ_{s+1}); s = 0, 1, \ldots, T\}$. As a result, $\{X_t; t = 0, 1, \ldots, T\}$ and $\{V_{t+1}; t = 0, 1, \ldots, T\}$ are jointly $\bar{\mathbf{P}}_{T+1}$–Gaussian with known statistics, and the rv ξ has known $\bar{\mathbf{P}}_{T+1}$–statistics and is $\bar{\mathbf{P}}_{T+1}$–independent of the observations $\{Y_s; s = 0, 1, \ldots, T\}$.*

As a consequence, $\mathbf{X}$ satisfies the same dynamics as $\mathbf{X}^\circ$, except that $\mathbf{W}$ replaces $\mathbf{W}^\circ$, and furthermore the $\bar{\mathbf{P}}_{T+1}$–statistics of $\{W_{t+1}; t = 0, 1, \ldots, T\}$ are the same as the $\mathbf{P}^\circ$–statistics of $\{W^\circ_{t+1}; t = 0, 1, \ldots, T\}$. This very nicely ties our calculations to the dynamics of the original system (1.1a).

From (5.10), we find that

$$Z_t = \Psi(t, 0)\xi \qquad\qquad t = 0, 1, \ldots$$

where here Ψ is the state transition matrix defined by

$$\Psi(s, s) = I_n, \quad \Psi(t+1, s) = [A_t - \Sigma^{wv}_{t+1}(\Sigma^v_{t+1})^{-1} H_t]\Psi(t, s), \qquad s \leq t. \quad s, t = 1, 2, \ldots \quad (5.13)$$

When $\Sigma^{wv}_t = 0$ for all $t = 1, 2, \ldots$, we see that Ψ agrees with Φ. We shall see that essentially the only change we need to make to equations (4.13)–(4.24) will be to replace Φ of (4.1) by Ψ of (5.13), so it will at each step be clear how our current calculations generalize those of Section IV.

The remainder of the arguments leading to the analogue of (4.24) is essentially the same as in Section IV. The Kallianpur–Striebel formula (4.10) still holds, as does the martingale argument of (4.11) and the iterated conditioning argument of (4.12). We can still write $\mathbf{L}$ in the form of (4.15) if we now make the definitions

$$B_0 := 0, \quad B_{t+1} := \sum_{s=0}^{t} [H_s \Psi(s, 0)]' V_{s+1} \qquad t = 0, 1, \ldots \quad (5.14)$$

and

$$M_0 := O_n, \quad M_{t+1} := \sum_{s=0}^{t} [H_s \Psi(s,0)]' (\Sigma_{s+1}^v)^{-1} [H_s \Psi(s,0)]$$

$$t = 0, 1, \ldots \quad (5.15)$$

instead of (4.13) and (4.14). Equations (4.16)–(4.17) still hold if we replace $\Phi(t+1,0)$ by $\Psi(t+1,0)$. We define the rvs $\{\bar{X}_t;\, t = 0,1,\ldots,T\}$, $\{\bar{B}_t;\, t = 0,1,\ldots,T\}$, $\{\tilde{X}_t;\, t = 0,1,\ldots,T\}$ and $\{\tilde{B}_t;\, t = 0,1,\ldots,T\}$ as in (4.18) and (4.19), and then (4.20) holds if we, again, replace $\Phi(t+1,0)$ by $\Psi(t+1,0)$. By virtue of the property $(\mathbf{G}^*)$ and of the equations (5.11)–(5.12) and (5.14), we see that the rvs $\{X_{t+1}, B_{t+1}, Y_0, Y_1, \ldots, Y_t\}$ are all linear combinations of $\{(W_{s+1}, V_{s+1});\, s = 0,1,\ldots,t\}$, whence they are $\bar{\mathbf{P}}_{T+1}$–jointly Gaussian and independent of ξ. Defining S_{t+1} as in (4.21), we find that (4.23) holds with $\Psi(t+1,0)$ replacing $\Phi(t+1,0)$. Finally, we have

$$\mathbf{E}^\circ[\varphi(X_{t+1}^\circ)|\mathcal{Y}_t] = \frac{\mathcal{U}\varphi[\bar{X}_{t+1}, \bar{B}_{t+1}; M_{t+1}, \Psi(t+1,0); S_{t+1}]}{\mathcal{U}\mathbf{1}[\bar{X}_{t+1}, \bar{B}_{t+1}; M_{t+1}, \Psi(t+1,0); S_{t+1}]} \qquad (5.16)$$

which is the generalization of (4.24) to the correlated case. Observe that the dependence of (5.16) upon the statistics $\bar{X}_{t+1}$, $\bar{B}_{t+1}$, M_{t+1}, $\Psi(t+1,0)$ and S_{t+1} is the same as the dependence of (4.24) upon the statistics $\bar{X}_{t+1}$, $\bar{B}_{t+1}$, M_{t+1}, $\Phi(t+1,0)$ and S_{t+1}. It is only the *definitions* of these statistics which is changed and not the *form* of the statistics–bearing functional as we move from the uncorrelated case to the correlated case.

To finish our study of the one–step prediction problem, we should give a method for actually computing $\bar{X}_{t+1}$, $\bar{B}_{t+1}$, and S_{t+1}. By writing down the dynamical equations for $\mathbf{B}$ which correspond to (5.14), we see that the *pair* $(\mathbf{X}, \mathbf{B})$ satisfies the recursive *plant equation*

$$\begin{pmatrix} X_0 \\ B_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

$$\begin{pmatrix} X_{t+1} \\ B_{t+1} \end{pmatrix} = \begin{pmatrix} A_t & 0 \\ 0 & I_n \end{pmatrix} \begin{pmatrix} X_t \\ B_t \end{pmatrix} \qquad t = 0,1,\ldots \quad (5.17a)$$

$$+ \begin{pmatrix} I_n & 0 \\ 0 & \Psi'(t,0) H_t' (\Sigma_{t+1}^v)^{-1} \end{pmatrix} \begin{pmatrix} W_{t+1} \\ V_{t+1} \end{pmatrix}$$

to which we adjoin an *observation* equation

$$Y_t = \begin{pmatrix} H_t & 0 \end{pmatrix} \begin{pmatrix} X_t \\ B_t \end{pmatrix} + \begin{pmatrix} 0 & I_k \end{pmatrix} \begin{pmatrix} W_{t+1} \\ V_{t+1} \end{pmatrix}. \qquad t = 0,1,\ldots \quad (5.17b)$$

Since the rvs $\{(W_{s+1}, V_{s+1});\, s = 0, 1, \ldots, T\}$ constitute a $\bar{\mathbf{P}}_{T+1}$–GWN sequence, the Kalman filtering formulae may be applied to find dynamical equations for the $\bar{\mathbf{P}}_{T+1}$–statistics $\{(\bar{X}_t, \bar{B}_t);\, t = 1, 2, \ldots, T+1\}$ and $\{S_t;\, t = 1, 2, \ldots, T+1\}$.

Prior to applying the Kalman filtering equations to (5.17a)–(5.17b), we make several comments concerning the effect of the horizon length T: The probability measure $\bar{\mathbf{P}}_{T+1}$ was constructed to ensure that the *finite–horizon* sequence $\{(W_{t+1}, V_{t+1});\, t = 0, 1, \ldots, T\}$ is a $\bar{\mathbf{P}}_{T+1}$–GWN sequence, and thus that the system (5.17a)–(5.17b) is amenable to Gaussian linear filtering methods for $t = 0, 1, \ldots, T$. By the dynamic nature of (5.17a)–(5.17b), the rvs $\{X_{t+1}, B_{t+1}, Y_0, Y_1, \ldots, Y_t\}$ are measurable with respect to $\{(W_{s+1}, V_{s+1});\, s = 0, 1, \ldots, t\}$ for each $t = 0, 1, \ldots$. Now, by the *consistency property* mentioned at the end of Section III, the statistics of $\{(W_{s+1}, V_{s+1});\, s = 0, 1, \ldots, t\}$ are the same under any two measures $\bar{\mathbf{P}}_{T+1}$ and $\bar{\mathbf{P}}_{T'+1}$ so long as $t \leq T \leq T'$, and therefore the rvs $\bar{X}_{t+1}$, $\bar{B}_{t+1}$ and S_{t+1} of (4.18) and (4.21) do *not* depend upon the horizon T for $T \geq t$. In particular, we may take $T = t$ in (4.18) and (4.21), setting

$$\bar{X}_{t+1} := \bar{\mathbf{E}}_{t+1}\left[X_{t+1} \middle| \mathcal{Y}_t\right] \quad \text{and} \quad \bar{B}_{t+1} := \bar{\mathbf{E}}_{t+1}\left[B_{t+1} \middle| \mathcal{Y}_t\right] \tag{5.18}$$

and

$$S_{t+1} := \bar{\mathbf{E}}_{t+1}\left[\begin{pmatrix} X_{t+1} - \bar{X}_{t+1} \\ B_{t+1} - \bar{B}_{t+1} \end{pmatrix} \begin{pmatrix} X_{t+1} - \bar{X}_{t+1} \\ B_{t+1} - \bar{B}_{t+1} \end{pmatrix}'\right], \tag{5.19}$$

for all $t = 0, 1, \ldots$, which defines $\bar{\mathbf{X}}$, $\bar{\mathbf{B}}$ and $\mathbf{S}$ as *infinite–horizon* sequences.

Turning now to the structure of the Kalman filter, as in (1.6)–(1.7), we recall that one of the strengths of the Kalman filtering equations is that they are *recursive*, i.e., at any time $t = 0, 1, \ldots$, the one–step predictor of the state at time $t+1$ and the corresponding (conditional) covariance matrix may be obtained by *updating* these quantities at time t using the observation at time t. Combining this with the remarks concerning the consistency of the measures $\{\bar{\mathbf{P}}_{T+1};\, T = 0, 1, \ldots\}$, we conclude that the Kalman filtering equations which yield $\bar{X}_{t+1}$, $\bar{B}_{t+1}$ and S_{t+1} do *not* depend on the choice of the horizon T, with $T \geq t$. A moment of reflection will convince the reader that consequently, *we may get valid recursive equations for* $\bar{\mathbf{X}}$, $\bar{\mathbf{B}}$ *and* $\mathbf{S}$ *by applying the Kalman filtering equations to (5.17a)–(5.17b) as if*

there were a measure $\bar{\mathbf{P}}$ *under which the entire infinite–horizon sequence* $\{(W_{t+1}, V_{t+1}); t = 0, 1, \ldots\}$ *were a* $\bar{\mathbf{P}}$*–GWN sequence.* In line with the closing comments of Section III, we remark that such a measure $\bar{\mathbf{P}}$ will not in general exist—we are merely introducing it as a fictitious construct to aid us in arriving at valid equations for $\bar{\mathbf{X}}$, $\bar{\mathbf{B}}$ and $\mathbf{S}$.

With these thoughts in mind, we apply the Kalman filtering equations (1.6)–(1.7) to the higher–dimensional system (5.17a)–(5.17b). To help us, we first write

$$S_t = \begin{pmatrix} P_t & Q_t \\ Q_t' & R_t \end{pmatrix} \qquad t = 0, 1, \ldots \quad (5.20)$$

as this corresponds to a natural partition of the covariance matrices of (5.19) according to the (conditional) covariances of the processes $\bar{\mathbf{X}}$ and $\bar{\mathbf{B}}$. After some simplification, we get the following recursions for the sequences $\mathbf{P} = \{P_t; t = 0, 1, \ldots\}$, $\mathbf{Q} = \{Q_t; t = 0, 1, \ldots\}$ and $\mathbf{R} = \{R_t; t = 0, 1, \ldots\}$: The $\mathcal{Q}_n$–valued deterministic sequence $\mathbf{P}$ satisfies the recursion

$$\begin{aligned} P_0 &= O_n, \\ P_{t+1} &= A_t P_t A_t' + \Sigma_{t+1}^w \qquad t = 0, 1, \ldots \quad (5.21) \\ &\quad - [A_t P_t H_t' + \Sigma_{t+1}^{wv}][H_t P_t H_t' + \Sigma_{t+1}^v]^{-1}[A_t P_t H_t' + \Sigma_{t+1}^{wv}]'. \end{aligned}$$

For convenience, we introduce the $\mathcal{Q}_k$–valued deterministic sequence $\mathbf{J} = \{J_t; t = 0, 1, \ldots\}$ by

$$J_t := H_t P_t H_t' + \Sigma_{t+1}^v. \qquad t = 0, 1, \ldots$$

The recursion for the $\mathcal{M}_n$–valued sequence $\mathbf{Q}$ is

$$\begin{aligned} Q_0 &= O_n, \\ Q_{t+1} &= A_t Q_t - [A_t P_t H_t' + \Sigma_{t+1}^{wv}] J_t^{-1} H_t (Q_t + \Psi(t, 0)) \\ &\quad + \Sigma_{t+1}^{wv} (\Sigma_{t+1}^v)^{-1} H_t \Psi(t, 0) \qquad t = 0, 1, \ldots \quad (5.22) \end{aligned}$$

and the recursion for the $\mathcal{Q}_n$–valued sequence $\mathbf{R}$ is

$$\begin{aligned} R_0 &= O_n, \\ R_{t+1} &= R_t - (Q_t + \Psi(t, 0))' H_t' J_t^{-1} H_t (Q_t + \Psi(t, 0)) \\ &\quad + \Psi'(t, 0) H_t' H_t \Psi(t, 0). \qquad t = 0, 1, \ldots \quad (5.23) \end{aligned}$$

The processes $\bar{\mathbf{X}}$ and $\bar{\mathbf{B}}$ of (5.18) satisfy the dynamical equations

$$\begin{aligned} \bar{X}_0 &= 0, \\ \bar{X}_{t+1} &= A_t \bar{X}_{t+1} + [A_t P_t H_t' + \Sigma_{t+1}^{wv}] J_t^{-1} [Y_t - H_t \bar{X}_{t+1}] \\ &\qquad\qquad t = 0, 1, \ldots \quad (5.24) \end{aligned}$$

and

$$\bar{B}_0 = 0$$

$$\bar{B}_{t+1} = \bar{B}_t + (Q_t + \Psi(t,0))' H_t' J_t^{-1} [Y_t - H_t \bar{X}_t].$$

$$t = 0, 1, \ldots \quad (5.25)$$

Let us also, if only for the sake of having a dynamical representation for all of the processes, rewrite (5.15) in the form

$$M_0 = O_n, \quad M_{t+1} = M_t + \Psi(t,0)' H_t' (\Sigma_{t+1}^v)^{-1} H_t \Psi(t,0).$$

$$t = 0, 1, \ldots \quad (5.26)$$

Thus we have that for any bounded Borel mapping φ from $\mathbb{R}^n$ to $\mathbb{C}$, (5.16) holds, where $\mathbf{S}$, $\bar{\mathbf{X}}$ and $\bar{\mathbf{B}}$ are defined by (5.20), (5.24) and (5.25). We state this result as a theorem:

Theorem 5.1. *For any bounded Borel mapping $\varphi : \mathbb{R}^n \to \mathbb{C}$ and any $t = 1, 2, \ldots$, we have that*

$$\mathbf{E}^\circ[\varphi(X_{t+1}^\circ)|\mathcal{Y}_t] = \frac{\mathcal{U}\varphi[\bar{X}_{t+1}, \bar{B}_{t+1}; M_{t+1}, \Psi(t+1,0); S_{t+1}]}{\mathcal{U}\mathbf{1}[\bar{X}_{t+1}, \bar{B}_{t+1}; M_{t+1}, \Psi(t+1,0); S_{t+1}]} \quad (5.27)$$

where the processes $\mathbf{S}$, $\bar{\mathbf{X}}$, $\bar{\mathbf{B}}$ and $\mathbf{M}$ are given respectively by (5.21), (5.24), (5.25) and (5.26), and the state transition matrix Ψ is defined by (5.13). The component sequences $\mathbf{P}$, $\mathbf{Q}$ and $\mathbf{R}$ of $\mathbf{S}$ propagate according to (5.21)–(5.23).

Observe the very special structure of (5.27) in that $\mathbf{E}^\circ[\varphi(X_{t+1}^\circ)|\mathcal{Y}_t]$ can be computed by inserting a collection of *finite-dimensional and computable* sufficient statistics which do *not* depend upon the distribution of the initial condition into a functional which is determined *solely* by the initial distribution. Note also that this functional does *not* depend upon time.

The calculation (4.25) at the end of Section V may now be more compactly rewritten as

$$\bar{\mathbf{E}}_{T+1}\left[L_{t+1}^{-1}\middle|\mathcal{Y}_t \vee \sigma\{\xi\}\right]$$

$$= \exp\left[-\frac{1}{2}\xi'(M_{t+1} - R_{t+1})\xi + \xi'\bar{B}_{t+1}\right]. \quad (5.28)$$

As a closing remark, we direct the reader's attention to the similarity of (1.6)–(1.7) and (5.21), (5.24). The reader will note the pleasing fact that the recursions for $\bar{\mathbf{X}}$ and $\mathbf{P}$ are *exactly* those we would get for the conditional means and conditional covariances of (1.1a)–(1.1b) if $\xi = 0$ (i.e., ξ is a degenerate Gaussian rv).

VI. REPRESENTATIONS FOR THE MMSE, LLSE, AND MMSE–LLSE ERROR

Having solved the one–step prediction problem in Theorem 5.1, we now use this result to give representations for the MMSE and LLSE estimates and for the mean square error between these estimates. As these representations are basic to the ensuing asymptotic analysis of ε given in Sections VII and VIII, we devote this entire section to their derivations.

As perhaps the most direct way of representing the MMSE process $\hat{\mathbf{X}}$ using Theorem 5.1, we will first find a representation for the conditional characteristic function

$$\mathbf{E}^\circ\left[\exp[i\theta' X^\circ_{t+1}]\middle|\mathcal{Y}_t\right], \quad \theta \in \mathbb{R}^n, \qquad\qquad t = 0, 1, \ldots$$

and then differentiate it with respect to θ at $\theta = 0$. This line of arguments is readily validated by the fact that the rv X°_{t+1} is an element of $L^2(\Omega, \mathcal{F}, \mathbf{P}^\circ)$. To carry out the calculations, we find it convenient to define for each θ in $\mathbb{R}^n$, the bounded Borel mapping $\varphi_\theta : \mathbb{R}^n \to \mathbb{C}$ by

$$\varphi_\theta(x) := \exp[i\theta' x], \quad x \in \mathbb{R}^n.$$

Fix $t = 0, 1, \ldots$ and θ in $\mathbb{R}^n$, and recall the definitions (2.4) and (2.5). We find by simple calculations that

$$\begin{aligned}&\mathcal{T}\varphi_\theta[x, b; S_{t+1}]\\ &\quad = \exp\left[i\theta' x - \frac{1}{2}\theta' P_{t+1}\theta + i\theta' Q_{t+1} b + \frac{1}{2} b' R_{t+1} b\right], \quad x, b \in \mathbb{R}^n.\end{aligned}$$

Another simple calculation shows that

$$\begin{aligned}&\mathcal{U}\varphi_\theta[x, b; M_{t+1}, \Psi(t+1, 0); S_{t+1}]\\ &\quad = \mathbf{E}^\circ\left[\exp\left[i\theta'(x + Q^\star_{t+1}\xi) - \frac{1}{2}\theta' P_{t+1}\theta + b'\xi - \frac{1}{2} z' R^\star_{t+1} z\right]\right]\\ &\quad = \int_{\mathbb{R}^n} \exp\left[i\theta'(x + Q^\star_{t+1} z) - \frac{1}{2}\theta' P_{t+1}\theta + b' z - \frac{1}{2} z' R^\star_{t+1} z\right] dF(z),\\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad x, b \in \mathbb{R}^n, \quad (6.1)\end{aligned}$$

where we have set

$$Q^\star_t := Q_t + \Psi(t, 0), \quad \text{and} \quad R^\star_t := M_t - R_t.$$

Since $\mathbf{1} = \varphi_\theta$ for $\theta = 0$, we can set $\theta = 0$ in (6.1) in order to get

$$\mathcal{U}\mathbf{1}[x, b; M_{t+1}, \Psi(t+1, 0); S_{t+1}]$$
$$= \int_{\mathbb{R}^n} \exp\left[b'z - \frac{1}{2} z' R^\star_{t+1} z\right] dF(z),$$
$$x, b \in \mathbb{R}^n. \quad (6.2)$$

Combining the expressions (6.1) and (6.2), when evaluated at $x = \bar{X}_{t+1}$ and $b = \bar{B}_{t+1}$, we conclude from Theorem 5.1 that

$$\mathbf{E}^\circ \left[\exp[i\theta' X^\circ_{t+1}] \middle| \mathcal{Y}_t\right] \quad (6.3)$$
$$= \frac{\int_{\mathbb{R}^n} \exp\left[i\theta'(\bar{X}_{t+1} + Q^\star_{t+1} z) - \frac{1}{2}\theta' P_{t+1}\theta + \bar{B}'_{t+1} z - \frac{1}{2} z' R^\star_{t+1} z\right] dF(z)}{\int_{\mathbb{R}^n} \exp\left[\bar{B}'_{t+1} z - \frac{1}{2} z' R^\star_{t+1} z\right] dF(z)}.$$

Finally, upon taking the gradient in (6.3) at $\theta = 0$, we obtain the following result.

Theorem 6.1. *For each $t = 0, 1, \ldots$, the relation*

$$\mathbf{E}^\circ[X^\circ_{t+1} | \mathcal{Y}_t] = \bar{X}_{t+1} + Q^\star_{t+1} \frac{\int_{\mathbb{R}^n} z \exp\left[z' \bar{B}_{t+1} - \frac{1}{2} z' R^\star_{t+1} z\right] dF(z)}{\int_{\mathbb{R}^n} \exp\left[z' \bar{B}_{t+1} - \frac{1}{2} z' R^\star_{t+1} z\right] dF(z)} \quad (6.4)$$

holds.

In the spirit of the recursions (5.22)–(5.23), we can now write down the dynamical equations for the matrices $\mathbf{Q}^\star$ and $\mathbf{R}^\star$; we find that

$$Q^\star_0 = I_n, \quad Q^\star_{t+1} = \left[A_t - [A_t P_t H'_t + \Sigma^{wv}] J_t^{-1} H_t\right] Q^\star_t \quad t = 0, 1, \ldots \quad (6.5)$$

and

$$R^\star_0 = O_n, \quad R^\star_{t+1} = R^\star_t + Q^{\star\prime}_t H'_t J_t^{-1} H_t Q^\star_t. \qquad t = 0, 1, \ldots \quad (6.6)$$

Using the process $\mathbf{Q}^\star$ instead of $\mathbf{Q}$, we can simplify the dynamics of $\bar{\mathbf{B}}$ to read

$$\bar{B}_0 = 0, \quad \bar{B}_{t+1} = \bar{B}_t + Q^{\star\prime}_t H'_t J_t^{-1} [Y_t - H_t \bar{X}_t]. \qquad t = 0, 1, \ldots$$

Proceeding now on to the LLSE process $\mathbf{X^K}$, we can easily get a representation for it by making use of the following two facts: (a) The LLSE or Kalman filter depends solely on the first and second moments of the

involved rvs [2, Sec. 5.2]; (b) The MMSE and LLSE filters coincide if ξ has a Gaussian distribution. Governed by these thoughts, we replace F in (6.4) by a Gaussian distribution of mean μ and covariance Δ to get, after some calculations, the following representation result.

Theorem 6.2. *If the covariance matrix Δ of the initial condition ξ is invertible, then we have the relation*

$$X_{t+1}^K = \bar{X}_{t+1} \qquad t = 1, 2, \ldots$$
$$+ Q_{t+1}^\star \left[R_{t+1}^\star + \Delta^{-1}\right]^{-1} \left[\bar{B}_{t+1} + \Delta^{-1}\mu\right].$$

In view of this last result, we strengthen **(A.3)** to condition **(A.4)**, namely

(A.4): The initial condition ξ has distribution F with finite first and second moments μ and Δ, respectively, and is independent of the process $(\mathbf{W}^\circ, \mathbf{V}^\circ)$. Furthermore, Δ is invertible, so that F belongs to $\mathcal{D}(\mathbb{R}^n)$.

For the remainder of this paper we assume **(A.1)**, **(A.2)** and **(A.4)** to hold.

The final representation we seek is for the error process $\boldsymbol{\varepsilon}$ defined by (1.10) and which measures how well the MMSE and LLSE estimates agree in quadratic mean. We shall naturally start from the representation results of Theorems 6.1 and 6.2. In order to make the notation less cumbersome, we introduce the mapping $\gamma : \mathbb{R}^n \times \mathbb{R}^n \times \mathcal{Q}_n \to \mathbb{R}$ defined by

$$\gamma(z, b; R) := \exp\left[z'b - \frac{1}{2}z'Rz\right], \quad z, b \in \mathbb{R}^n,\ R \in \mathcal{Q}_n. \tag{6.7}$$

Fix $t = 0, 1, \ldots$. Combining Theorems 6.1 and 6.2, we write the difference $\hat{X}_{t+1} - X_{t+1}^K$, with the help of the function γ, in the form

$$\hat{X}_{t+1} - X_{t+1}^K = Q_{t+1}^\star \frac{\int_{\mathbb{R}^n} \{\cdots\}\gamma(z, \bar{B}_{t+1}; R_{t+1}^\star)dF(z)}{\int_{\mathbb{R}^n} \gamma(z, \bar{B}_{t+1}; R_{t+1}^\star)dF(z)} \tag{6.8}$$

where

$$\{\cdots\} = z - \left[R_{t+1}^\star + \Delta^{-1}\right]^{-1} \left[\bar{B}_{t+1} + \Delta^{-1}\mu\right].$$

We clearly see from this formula that $\hat{X}_{t+1} - X_{t+1}^K$—which we *a priori* know depends upon the observations $\{Y_0, Y_1, \ldots, Y_t\}$—in fact depends upon these

observations *only* through the rv $\bar{B}_{t+1}$. To calculate ε_{t+1}, we thus must, after applying the mapping $x \mapsto \|x\|^2$, average the expression (6.8) over $\bar{B}_{t+1}$.

To gain some insight into how to perform these calculations, recall the definition (4.18) of $\bar{B}_{t+1}$ as well as the fact that *the rvs* $\{B_{t+1}, Y_0, Y_1, \ldots, Y_t\}$ *are jointly* $\bar{\mathbf{P}}_{t+1}$*–Gaussian.* In contrast, the $\mathbf{P}^\circ$–statistics of $\bar{B}_{t+1}$ are in general non–Gaussian! This state of affairs suggests that the computations for ε_{t+1} should be performed under $\bar{\mathbf{P}}_{t+1}$ *instead* of under $\mathbf{P}^\circ$. We shall in fact see that many of the properties of $\bar{\mathbf{P}}_{t+1}$ which were used in deriving Theorem 5.1 will also be used here.

We begin by observing that

$$\begin{aligned}
\varepsilon_{t+1} &= \bar{\mathbf{E}}_{t+1}\left[\|\hat{X}_{t+1} - X_{t+1}^K\|^2 L_{t+1}^{-1}\right] \\
&= \bar{\mathbf{E}}_{t+1}\left[\|\hat{X}_{t+1} - X_{t+1}^K\|^2 \bar{\mathbf{E}}_{t+1}\left[L_{t+1}^{-1} | \mathcal{Y}_t \vee \sigma\{\xi\}\right]\right] \qquad (6.9) \\
&= \bar{\mathbf{E}}_{t+1}\left[\|\hat{X}_{t+1} - X_{t+1}^K\|^2 \gamma(\xi, B_{t+1}; M_{t+1})\right]. \qquad (6.10)
\end{aligned}$$

In passing from (6.9) to (6.10) we have used (5.28). We also recall that the rv B_{t+1} and the $\mathcal{Y}_t$–measurable rv $\hat{X}_{t+1} - X_{t+1}^K$ are $\bar{\mathbf{P}}_{t+1}$–independent of ξ. Therefore, a simple conditioning argument in (6.10) followed by an application of Tonelli's theorem yields

$$\begin{aligned}
\varepsilon_{t+1} &= \bar{\mathbf{E}}_{t+1}\left[\|\hat{X}_{t+1} - X_{t+1}^K\|^2 \int_{\mathbb{R}^n} \gamma(z, B_{t+1}; M_{t+1}) dF(z)\right] \\
&= \int_{\mathbb{R}^n} \bar{\mathbf{E}}_{t+1}\left[\|\hat{X}_{t+1} - X_{t+1}^K\|^2 \gamma(z, B_{t+1}; M_{t+1})\right] dF(z). \qquad (6.11)
\end{aligned}$$

We next decompose B_{t+1} as $B_{t+1} = \bar{B}_{t+1} + \tilde{B}_{t+1}$ as in (4.18)–(4.19). Since $\hat{X}_{t+1} - X_{t+1}^K$ is $\sigma\{\bar{B}_{t+1}\}$–measurable, as we remarked above, we find that

$$\begin{aligned}
&\bar{\mathbf{E}}_{t+1}\left[\|\hat{X}_{t+1} - X_{t+1}^K\|^2 \gamma(z, B_{t+1}; M_{t+1})\right] \\
&\quad = \bar{\mathbf{E}}_{t+1}\left[\|\hat{X}_{t+1} - X_{t+1}^K\|^2 \bar{\mathbf{E}}_{t+1}\left[\gamma(z, B_{t+1}; M_{t+1}) \middle| \bar{B}_{t+1}\right]\right], \quad z \in \mathbb{R}^n.
\end{aligned}$$

In Sections IV and V we used the fact that $\bar{B}_{t+1}$ and $\tilde{B}_{t+1}$ are $\bar{\mathbf{P}}_{t+1}$–independent; we again use this fact, together with the definition of the matrix R_{t+1} as the $\bar{\mathbf{P}}_{t+1}$–covariance of the zero–mean $\bar{\mathbf{P}}_{t+1}$–Gaussian rv

$\tilde{B}_{t+1}$, to make the simplification

$$\begin{aligned}\bar{\mathbf{E}}_{t+1}\left[\gamma(z, B_{t+1}; M_{t+1})\middle|\bar{B}_{t+1}\right] &= \bar{\mathbf{E}}_{t+1}\left[\gamma(z, b+\tilde{B}_{t+1}; M_{t+1})\right]_{b=\bar{B}_{t+1}} \\ &= \exp\left[z'\bar{B}_{t+1} - \frac{1}{2}z'R^{\star}_{t+1}z\right] \\ &= \gamma(z, \bar{B}_{t+1}; R^{\star}_{t+1}), \quad z \in \mathbb{R}^n.\end{aligned}$$

Therefore,

$$\begin{aligned}&\bar{\mathbf{E}}_{t+1}\left[\|\hat{X}_{t+1} - X^K_{t+1}\|^2\gamma(z, B_{t+1}; M_{t+1})\right] \\ &\quad = \bar{\mathbf{E}}_{t+1}\left[\|\hat{X}_{t+1} - X^K_{t+1}\|^2\gamma(z, \bar{B}_{t+1}; R^{\star}_{t+1})\right], \quad z \in \mathbb{R}^n. \qquad (6.12)\end{aligned}$$

Inserting (6.12) into (6.11), we readily see from (6.8) that

$$\varepsilon_{t+1} = \int_{\mathbb{R}^n} \bar{\mathbf{E}}_{t+1}\left[\|\hat{X}_{t+1} - X^K_{t+1}\|^2\gamma(z, \bar{B}_{t+1}; R^{\star}_{t+1})\right] dF(z) \qquad (6.13)$$

$$= \bar{\mathbf{E}}_{t+1}\left[\frac{\left\|Q^{\star}_{t+1}\int_{\mathbb{R}^n}\{\cdots\}\gamma(z, \bar{B}_{t+1}; R^{\star}_{t+1})dF(z)\right\|^2}{\int_{\mathbb{R}^n}\gamma(z, \bar{B}_{t+1}; R^{\star}_{t+1})dF(z)}\right] \qquad (6.14)$$

where

$$\{\cdots\} = z - \left[R^{\star}_{t+1} + \Delta^{-1}\right]^{-1}\left[\bar{B}_{t+1} + \Delta^{-1}\mu\right]$$

after a simple cancellation. It now remains only to find the $\bar{\mathbf{P}}_{t+1}$–statistics of the rv $\bar{B}_{t+1}$; these are presented in the following lemma.

Lemma 6.3. *For all $t = 0, 1, \ldots$, under the probability measure $\bar{\mathbf{P}}_{t+1}$, the rv $\bar{B}_{t+1}$ is Gaussian with zero mean and covariance $R^{\star}_{t+1}$.*

Proof. The arguments of Section V already show us that $\bar{B}_{t+1}$ is $\bar{\mathbf{P}}_{t+1}$–Gaussian: Indeed, the rv $\bar{B}_{t+1}$ is the conditional $\bar{\mathbf{P}}_{t+1}$–expectation of B_{t+1} given the rvs $\{Y_0, Y_1, \ldots, Y_t\}$, and moreover the rvs $\{B_{t+1}, Y_0, Y_1, \ldots, Y_t\}$ are jointly $\bar{\mathbf{P}}_{t+1}$–Gaussian. The rv $\bar{B}_{t+1}$ has zero mean under $\bar{\mathbf{P}}_{t+1}$ since B_{t+1} has zero mean under $\bar{\mathbf{P}}_{t+1}$. Finally, since the rvs $\bar{B}_{t+1}$ and $\tilde{B}_{t+1}$ are $\bar{\mathbf{P}}_{t+1}$–independent, it is easy to see that

$$\mathrm{Cov}_{\bar{\mathbf{P}}_{t+1}} B_{t+1} = \mathrm{Cov}_{\bar{\mathbf{P}}_{t+1}} \bar{B}_{t+1} + \mathrm{Cov}_{\bar{\mathbf{P}}_{t+1}} \tilde{B}_{t+1}.$$

From (5.14) and (5.15), we see that $\mathrm{Cov}_{\bar{\mathbf{P}}_{t+1}} B_{t+1} = M_{t+1}$, and by definition, R_{t+1} is the $\bar{\mathbf{P}}_{t+1}$–covariance of the rv $\tilde{B}_{t+1}$. This completes the proof. ■

This last property leads to a simple representation for the error process ε when used on (6.13)–(6.14).

Theorem 6.4. *For each $t = 0, 1, \ldots$, we have the representation*

$$\varepsilon_{t+1} = \int_{\mathbb{R}^n} \frac{\left\| Q^\star_{t+1} \int_{\mathbb{R}^n} \{\cdots\} \gamma(z, b; R^\star_{t+1}) dF(z) \right\|^2}{\int_{\mathbb{R}^n} \gamma(z, b; R^\star_{t+1}) dF(z)} dG_{R^\star_t}(b) \qquad (6.15)$$

where

$$\{\cdots\} = z - \left[R^\star_{t+1} + \Delta^{-1}\right]^{-1} \left[b + \Delta^{-1}\mu\right].$$

VII. THE ASYMPTOTIC BEHAVIOR OF ε: THE MULTIVARIABLE CASE

We now begin our study of the asymptotics of the error process ε, with Theorem 6.4 as our main tool, and based upon the dependencies suggested by (1.11). In order to make such an analysis feasible, we must of course enforce some structure on the asymptotics of the matrix–valued sequences $\mathbf{A}$, $\mathbf{H}$ and $\mathbf{\Sigma}$ which describe the dynamics of the system (1.1a)–(1.1b). We shall from here on restrict ourselves to the classical time–homogeneous situation. This is captured by assumption **(A.5)** enforced hereafter, where

(A.5): The matrix–valued sequences $\mathbf{A}$, $\mathbf{H}$ and $\mathbf{\Sigma}$ are constant; that is, there are matrices A, H and Σ in $\mathcal{M}_n$, $\mathcal{M}_{k\times n}$ and $\mathcal{Q}_{n+k}$, respectively, such that for all $t = 0, 1, \ldots$, $A_t = A$, $H_t = H$ and $\Sigma_{t+1} = \Sigma$.

It is clear from the form of (6.15) that we may study ε_t as a functional of the initial distribution F and of the matrices $Q^\star_t$ and $R^\star_t$, for each $t = 1, 2, \ldots$. Not surprisingly then, a central component of our efforts will be an analysis of the asymptotics of the sequences $\mathbf{Q}^\star$ and $\mathbf{R}^\star$. An examination of the dynamics of these sequences will indicate several cases which are amenable to straightforward analysis. However, as often the case with general multivariable systems, a complete analysis is not possible.

A considerable simplification occurs in (6.15) if $\mu = E[\xi] = 0$. Of course, since the dynamics (1.1a)–(1.1b) are linear, no generality is lost by assuming that the initial distribution F is centered. Indeed, a simple

translation argument shows that for any distribution F in $\mathcal{D}(\mathbb{R}^n)$ with mean μ, the relation

$$\varepsilon_t((A,H,\Sigma),F) = \varepsilon_t((A,H,\Sigma),\tilde{F}) \qquad t = 1,2,\ldots \quad (7.1)$$

holds, where $\tilde{F}$ is the centered distribution in $\mathcal{D}^\circ(\mathbb{R}^n)$ defined by the translation

$$\tilde{F}(x) = F(x-\mu), \quad x \in \mathbb{R}^n. \quad (7.2)$$

Consequently, we shall henceforth in this section consider only distributions F in $\mathcal{D}^\circ(\mathbb{R}^n)$.

In order to more clearly understand the dependence of the right-hand side of (6.15) upon the distribution F and the matrices $Q_t^\star$ and $R_t^\star$, we define the mapping $I_F : \mathcal{M}_n \times \mathcal{Q}_n \to \mathbb{R}$ as

$$I_F(K,R) = \int_{\mathbb{R}^n} \frac{\left\| K \int_{\mathbb{R}^n} \left\{ z - \left[R + \Delta^{-1}\right]^{-1} b \right\} \gamma(z,b;R) dF(z) \right\|^2}{\int_{\mathbb{R}^n} \gamma(z,b;R) dF(z)} dG_R(b),$$
$$K \in \mathcal{M}_n,\ R \in \mathcal{Q}_n. \quad (7.3)$$

In Lemma 7.1 below we verify that this expression is well defined for all matrices K in $\mathcal{M}_n$ and R in $\mathcal{Q}_n$, so that (6.15) may indeed be rewritten as

$$\varepsilon_t = I_F(Q_t^\star, R_t^\star). \qquad t = 1,2,\ldots \quad (7.4)$$

This clearly separates the dependence of ε_t on the matrices $Q_t^\star$ and $R_t^\star$, which depend only on the system triple (A,H,Σ), 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 the matrices $Q_t^\star$ and $R_t^\star$.*

Before beginning with our calculations, we set

$$\Gamma(b;R) := \int_{\mathbb{R}^n} \gamma(z,b;R) dF(z), \quad b \in \mathbb{R}^n,\ R \in \mathcal{Q}_n.$$

Some of our manipulations of (7.3) will be clearer when using this notation. We first show that (7.3) is well defined for all K in $\mathcal{M}_n$ and all R in $\mathcal{Q}_n$.

Lemma 7.1. *Let F be a distribution in $\mathcal{D}^\circ(\mathbb{R}^n)$. For all K in $\mathcal{M}_n$ and R in $\mathcal{Q}_n$, the quantity $I_F(K,R)$ is well defined and finite, with alternate representation*

$$I_F(K,R) = \int_{\mathbb{R}^n} \{\cdots\}\Gamma(b;R)dG_R(b), \quad K \in \mathcal{M}_n,\ R \in \mathcal{Q}_n, \quad (7.5)$$

where

$$\{\cdots\} = \left\| K \int_{\mathbb{R}^n} \left\{ z - [R + \Delta^{-1}]^{-1} b \right\} \frac{\gamma(z,b;R)}{\Gamma(b;R)} dF(z) \right\|^2 .$$

Proof. In view of (7.4), we already know, via the probabilistic arguments of Section VI, that (7.3) is well defined and finite when $K = Q_t^\star$ and $R = R_t^\star$ for all $t = 1, 2, \ldots$, given any system triple (A, H, Σ). Rather than working to extend these probabilistic calculations to cover all K and R, we shall instead study formula (7.3) by simple techniques from analysis. In the process, we shall obtain a bound—inequality (7.9)—which will prove useful later on.

Fix K in $\mathcal{M}_n$ and R in $\mathcal{Q}_n$. First we show that $I_F(K,R)$ is well defined. Observe that whenever b lies in the range $\mathrm{Im}(R)$ of R, the quadratic form in the exponent of γ in (6.7) is amenable to a completion of squares, namely

$$z'b - \frac{1}{2}z'Rz = \frac{1}{2}b'R^\#b - \frac{1}{2}\left(z - R^\#b\right)' R \left(z - R^\#b\right), \qquad z \in \mathbb{R}^n,\ b \in \mathrm{Im}(R),$$

where $R^\#$ denotes the Moore–Penrose pseudo–inverse of R [2, pp. 329–330]. Consequently,

$$0 < \gamma(z,b;R) \le \exp\left[\frac{1}{2}b'R^\#b\right], \quad z \in \mathbb{R}^n,\ b \in \mathrm{Im}(R), \qquad (7.6)$$

and the bound

$$0 < \Gamma(b;R) \le \exp\left[\frac{1}{2}b'R^\#b\right], \quad b \in \mathrm{Im}(R), \qquad (7.7)$$

holds. The bound (7.6) and the finite second moment assumption **(A.2)** on ξ together imply that the inner integral in the numerator of (7.3) is well defined and finite for each b in $\mathrm{Im}(R)$. Therefore, since the support of the

Gaussian distribution G_R is exactly $\mathrm{Im}(R)$, we conclude, using (7.7), that $I_F(K,R)$ is indeed well defined. As a result of this discussion, we see that the alternate expression (7.5) is indeed valid.

Next, to show that $I_F(K,R)$ is finite, we first observe from Jensen's inequality and from the definition of the operator norm (2.1) that

$$\left\| K \int_{\mathbb{R}^n} \{z - [R+\Delta^{-1}]^{-1}b\} \frac{\gamma(z,b;R)}{\Gamma(b;R)} dF(z) \right\|^2$$
$$\leq \|K\|_{op}^2 \int_{\mathbb{R}^n} \left\| z - [R+\Delta^{-1}]^{-1}b \right\|^2 \frac{\gamma(z,b;R)}{\Gamma(b;R)} dF(z),$$
$$b \in \mathrm{Im}(R). \quad (7.8)$$

Next, we set

$$J_F(R)$$
$$:= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \left\| z - [R+\Delta^{-1}]^{-1}b \right\|^2 \gamma(z,b;R) dF(z) dG_R(b)$$
$$= \int_{\mathbb{R}^n} \left[\int_{\mathbb{R}^n} \left\| z - [R+\Delta^{-1}]^{-1}b \right\|^2 \exp[z'b] dG_R(b) \right] \exp\left[-\frac{1}{2} z'Rz \right] dF(z)$$

where the last equality follows from Tonelli's theorem, and it is now plain from (7.5) and (7.8) that

$$I_F(K,R) \leq \|K\|_{op}^2 J_F(R). \quad (7.9)$$

Finally, after some tedious calculations, we find that

$$J_F(R) = \mathrm{trace}\ \left([R+\Delta^{-1}]^{-1} R [R+\Delta^{-1}]^{-1} \right) \quad (7.10)$$
$$+ \int_{\mathbb{R}^n} z' \Delta^{-1} [R+\Delta^{-1}]^{-1} [R+\Delta^{-1}]^{-1} \Delta^{-1} z dF(z)$$

and therefore since ξ has finite second moments, $J_F(R)$ is finite and so is $I_F(K,R)$ as a result of (7.9). ∎

With Lemma 7.1 in hand, we now commence the study of the asymptotics of $\boldsymbol{\varepsilon}$ through the representation (7.4). This requires that we study the behavior of I_F under the *joint* asymptotic behavior of the sequences $\mathbf{Q}^\star$ and $\mathbf{R}^\star$. By using (2.3) on (7.3), we get

$$\lambda_{\min}(Q_t^{\star\prime} Q_t^\star) I_F(I_n, R_t^\star)$$
$$\leq \varepsilon_t((A,H,\Sigma),F)$$
$$\leq \lambda_{\max}(Q_t^{\star\prime} Q_t^\star) I_F(I_n, R_t^\star), \qquad t = 1,2,\ldots \quad (7.11)$$

thereby separating somewhat the effects of $Q_t^\star$ from those of $R_t^\star$. This directs our study towards an understanding of the behavior of $Q_t^{\star\prime}Q_t^\star$ and $I_F(I_n, R_t^\star)$ for large times t under different assumptions on the initial distribution F and the system triple (A, H, Σ).

Some general comments on the structure of the dynamics of the matrix–valued sequences $\mathbf{P}$, $\mathbf{Q}^\star$ and $\mathbf{R}^\star$ are in order at this point. Firstly, the equation (5.21) for $\mathbf{P}$ is a discrete–time Ricatti equation. Since discrete–time Ricatti equations have been extensively studied, there exists a fairly large body of results concerning the large–time asymptotics of $\mathbf{P}$. Turning next to the sequence $\mathbf{Q}^\star$, we rewrite (6.5) as

$$Q_0^\star = I_n, \quad Q_{t+1}^\star = K_t Q_t^\star \qquad t = 0, 1, \ldots \quad (7.12)$$

where we have set

$$K_t := A - [AP_tH' + \Sigma^{wv}][HP_tH' + \Sigma^v]^{-1}H. \qquad t = 0, 1, \ldots$$

Thus we expect $\mathbf{Q}^\star$ to exhibit some sort of exponential growth or decay which will depend primarily upon the spectrum of the limiting matrix $\lim_t K_t$, if this limit exists. As the existence of this limit defines an important situation, we formally introduce the following condition **(C.1)**, where

(C.1): The limit $K_\infty := \lim_t K_t$ exists.

Of course, if $P_\infty := \lim_t P_t$ exists, then **(C.1)** holds true, and we have

$$K_\infty = A - [AP_\infty H' + \Sigma^{wv}][HP_\infty H' + \Sigma^v]^{-1}H.$$

Finally, we can see more clearly what to expect of $\mathbf{R}^\star$ by rewriting (6.6) as

$$R_t^\star = \sum_{s=0}^{t-1} Q_s^{\star\prime} H'[HP_sH' + \Sigma^v]^{-1} H Q_s^\star. \qquad t = 1, 2, \ldots \quad (7.13)$$

We note here for future reference that $\mathbf{R}^\star$ is nondecreasing in the sense of the partial ordering on $\mathcal{Q}_n$; i.e., $v'R_t^\star v \leq v'R_{t+1}^\star v$ for all v in $\mathbb{R}^n$ and all $t = 0, 1, \ldots$. However, in general this does *not* imply convergence of the sequence $\mathbf{R}^\star$ which would be equivalent to the convergence of the sequence of cross–terms $\{u'R_t^\star v;\ t = 0, 1, \ldots\}$ for all u and v in $\mathbb{R}^n$. That the previous

convergence "on the diagonal" is not sufficient to ensure full convergence can be seen from the parallelogram identity, which here takes the form

$$(u+v)'R_t^\star(u+v) - (u-v)'R_t^\star(u-v) = 4u'R_t^\star v,$$
$$u, v \in R^n. \quad t = 0, 1, \ldots$$

Therefore the cross–terms $u'R_t^\star v$ will not converge unless *both* monotone sequences $(u+v)'R_t^\star(u+v)$ and $(u-v)'R_t^\star(u-v)$ have a *finite* limit. In view of these comments, we shall find the following to be a valuable hypothesis on the asymptotics of $\mathbf{R}^\star$:

(C.2): The sequence $\mathbf{R}^\star$ has a well–defined limit $R_\infty^\star$ which is positive definite and thus invertible.

We observe from (7.13) that a natural situation in which **(C.2)** holds arises when the sequence $\mathbf{Q}^\star$ tends to zero, assumably with some *exponential* rate.

First we study the asymptotic behavior of $\mathbf{Q}^\star$. We make rigorous our earlier comment that $\mathbf{Q}^\star$ exponentially grows or decays with rate depending upon K_∞ whenever **(C.1)** holds.

Proposition 7.2. *Under* **(C.1)**, *we have the following estimates: The upper bound*

$$\overline{\lim}_t \frac{1}{t} \ln \lambda_{\max}(Q_t^{\star\prime} Q_t^\star) \leq 2 \ln \rho(K_\infty) \tag{7.14a}$$

always holds. The lower bound

$$2 \ln \lambda_{\min}(K_\infty) \leq \underline{\lim}_t \frac{1}{t} \ln \lambda_{\min}(Q_t^{\star\prime} Q_t^\star) \tag{7.14b}$$

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

If K_T is singular for some T, then by (7.12) $Q_t^\star$ is singular for all $t \geq T$. In this case (7.14b) cannot hold unless K_∞ is also singular. This is the reason for the two assumptions for (7.14b). In the analogous continuous–time calculations, $\mathbf{Q}^\star$ obeys a linear differential equation in $\mathcal{M}_n$, with the net result that $Q_t^\star$ will be invertible for all t.

Proof of Proposition 7.2. First we develop the recursion (7.12) to obtain

$$Q_{t+1}^\star = K_t K_{t-1} \ldots K_{s+1} Q_s^\star, \quad s \leq t. \qquad s, t = 0, 1, \ldots \tag{7.15}$$

The natural basis of our arguments would be the following sequence of steps:

$$\overline{\lim}_t \frac{1}{t} \ln \lambda_{\max}(Q_t^{\star\prime} Q_t^\star) = 2\,\overline{\lim}_t \frac{1}{t} \ln \|Q_t^\star\|_{op}$$

$$\leq 2\,\overline{\lim}_t \frac{1}{t} \sum_{s=1}^{t} \ln \|K_s\|_{op} \tag{7.16}$$

$$= 2 \ln \|K_\infty\|_{op}. \tag{7.17}$$

The first equality comes from (2.2) while (7.16) follows from (7.15) (with $s = 0$) with the help of standard properties of the operator norm; the passage to (7.17) is validated by invoking **(C.1)**, the continuity of the operator norm and Cesaro convergence. Unfortunately, (7.17) is short of the desired result (7.14a), which gives a tighter bound than (7.17) since in general $\rho(K_\infty) \leq \|K_\infty\|_{op}$ [31].

The arguments leading to (7.17) must therefore be modified. The missing step is to be found in a well–known fact from matrix theory [31, p. 271 and Thm. 3.8, p. 284] stating that

$$\lim_N \left(\|K_\infty^N\|_{op}\right)^{1/N} = \rho(K_\infty). \tag{7.18}$$

This suggests that we consider the evolution of $\mathbf{Q}^\star$ at time instants $\{0, N, 2N, \ldots\}$ for each $N = 1, 2, \ldots$. To do this, we fix an integer $N = 1, 2, \ldots$ and define the matrices $\{K_t^{(N)};\ t = 0, 1, \ldots\}$ by

$$K_t^{(N)} := K_{t+N} K_{t+N-1} \cdots K_{t+1}. \qquad t = 0, 1, \ldots \tag{7.19}$$

Note that under **(C.1)**, we have the convergence $\lim_t K_t^{(N)} = K_\infty^N$. By standard properties of the operator norm, we see from (7.15) and (7.19) that

$$\overline{\lim}_j \frac{1}{j} \ln \|Q_{jN}^\star\|_{op} \leq \overline{\lim}_j \frac{1}{j} \sum_{k=0}^{j-1} \ln \|K_{kN}^{(N)}\|_{op}$$

$$= \ln \|K_\infty^N\|_{op}, \tag{7.20}$$

where the last equality (7.20) is obtained by Cesaro convergence since

$$\lim_k \|K_{kN}^{(N)}\|_{op} = \|K_\infty^N\|_{op}.$$

To pass from the estimate (7.20) on the lattice $\{0, N, 2N, \cdots\}$ to a corresponding estimate on $\{0, 1, \ldots\}$, we define

$$j_N(t) := \lfloor t/n \rfloor, \quad N = 1, 2, \ldots \qquad\qquad t = 0, 1, \ldots$$

where $\lfloor \cdot \rfloor$ is the integer floor function; for each $t = 0, 1, \ldots$, $j_N(t)$ is the unique integer such that

$$j_N(t)N \leq t < (j_N(t) + 1)N, \quad N = 1, 2, \ldots . \tag{7.21}$$

For each fixed integer $N = 1, 2, \ldots$, we readily conclude from (7.15) (with $s = j_N(t)N$) that

$$\begin{aligned} \overline{\lim}_t \frac{1}{t} \ln \|Q_t^\star\|_{op} \leq \overline{\lim}_t \frac{1}{t} \sum_{k=j_N(t)+1}^{t} |\ln \|K_k\|_{op}| \\ + \overline{\lim}_t \frac{j_N(t)}{t} \frac{1}{j_N(t)} \ln \|Q^\star_{j_N(t)N}\|_{op}. \end{aligned} \tag{7.22}$$

Using (7.21), we observe that

$$\frac{1}{t} \sum_{k=j_N(t)+1}^{t} |\ln \|K_k\|_{op}| \leq \frac{N}{j_N(t)N} \max_{t-N \leq k \leq t} |\ln \|K_k\|_{op}| \tag{7.23}$$

and

$$\frac{1}{N} \left(1 - \frac{N}{t}\right) \leq \frac{j_N(t)}{t} \leq \frac{1}{N}. \tag{7.24}$$

Since $\lim_t j_N(t) = \infty$ monotonically, it is now apparent from (7.22)–(7.24) that

$$\begin{aligned} \overline{\lim}_t \frac{1}{t} \ln \|Q_t^\star\|_{op} \leq & \frac{1}{N} \overline{\lim}_j \frac{1}{j} \ln \|Q_{jN}^\star\|_{op} \\ = & \frac{1}{N} \ln \|K_\infty^N\|_{op}. \end{aligned} \tag{7.25}$$

where the last equality follows from (7.20).

Now letting N go to infinity in (7.25) and using (7.18), we finally get the desired estimate (7.14a) via (2.2)

The proof of (7.14b) is similar to (7.14a) if all the matrices $\{K_t;\ t = 0, 1, \ldots\}$ are invertible: Indeed, under such circumstance, all the matrices in the sequence $\mathbf{Q}^\star$ are invertible and we can write the following recursion

$$(Q_0^{\star -1})' = I_n, \quad (Q_{t+1}^{\star -1})' = (K_t^{-1})'(Q_t^{\star -1})' \qquad t = 0, 1, \ldots \tag{7.26}$$

for the inverse transposed matrices; this recursion has essentially the same form as the recursion (7.12) for the sequence $\mathbf{Q}^\star$. Therefore, if in addition K_∞ is invertible, then the sequence of matrices $\{(K_t^{-1})';\ t = 0, 1, \ldots\}$ (which plays the role of the sequence $\mathbf{K}$ for (7.26)) satisfies condition **(C.1)**, i.e., it has a limit with $\lim_t (K_t^{-1})' = (K_\infty^{-1})'$. Consequently, the arguments which validated (7.14a) in the first part of the proof apply to give

$$\begin{aligned} \overline{\lim}_t \frac{1}{t} \ln \|(Q_t^{\star -1})'\|_{op}^2 &\leq 2 \ln \rho((K_\infty^{-1})') \\ &= 2 \ln \rho(K_\infty^{-1}) \\ &= 2 \ln \lambda_{\min}(K_\infty)^{-1} \end{aligned} \tag{7.27}$$

where the last two equalities readily follow from well–known facts from matrix theory [31]. Next, invoking (2.2) again, we also have

$$\begin{aligned} \|(Q_t^{\star -1})'\|_{op}^2 &= \lambda_{\max}((Q_t^{\star -1})(Q_t^{\star -1})') \\ &= \lambda_{\max}((Q_t^{\star\prime} Q_t^\star)^{-1}) \\ &= (\lambda_{\min}(Q_t^{\star\prime} Q_t^\star))^{-1}. \qquad t = 0, 1, \ldots \end{aligned} \tag{7.28}$$

Combining (7.27) and (7.28), we obtain that (7.14b) holds when the matrices $\{K_t;\ t = 0, 1, \ldots\}$ and K_∞ are all invertible. In the other case, when K_∞ is singular, (7.14b) holds trivially. This covers all the cases in the hypothesis for the lower bound (7.14b). ∎

Proposition 7.2 gives part of the asymptotic analysis suggested by (7.11). The following result helps complete the picture.

Proposition 7.3. *For every distribution F in $\mathcal{D}^\circ(\mathbb{R}^n)$, we have* $\sup_t\ I_F(I_n, R_t^\star) < \infty$.

Proof. Note that we have not required condition **(C.2)** to hold. In light of the bound (7.9), it is sufficient to show that

$$\overline{\lim}_t J_F(R^\star) < \infty \tag{7.29}$$

for all distributions F in $\mathcal{D}^\circ(\mathbb{R}^n)$. The functional J_F being continuous on $\mathcal{Q}_n$, we clearly have (7.29) under the assumption **(C.2)**. When **(C.2)** does not hold in the scalar case (i.e., $n = 1$), $\mathbf{R}^\star$ is a scalar sequence, and we may carry out some simple calculations on (7.10) to show (7.29). Some analogous but more complicated manipulations of (7.10) also yield (7.29)

in the multivariable case when **(C.2)** does not hold; details are omitted for the sake of brevity. ∎

Combining Proposition 7.3 with (7.11) and the estimate (7.14a) yields the following upper bound.

Theorem 7.4. *If assumption* **(C.1)** *holds, then*

$$\overline{\lim}_t \frac{1}{t} \ln \varepsilon_t \leq 2 \ln \rho(K_\infty).$$

The analogous asymptotic lower bound is made a bit more complicated by the possibility that we may have $\underline{\lim}_t I_F(I_n, R_t^\star) = 0$. In the next section, we prove the following "converse".

Proposition 7.5. *If assumption* **(C.2)** *holds, then* $\underline{\lim}_t I_F(I_n, R_t^\star) = 0$ *implies that the distribution* F *of the initial condition is Gaussian.*

Collecting this result and the lower bound of (7.14b) gives the following lower bound.

Theorem 7.6. *If assumptions* **(C.1)** *and* **(C.2)** *hold, then for every non–Gaussian initial distribution* F *in* $\mathcal{D}^\circ(\mathbb{R}^n)$, *we have*

$$\underline{\lim}_t \frac{1}{t} \ln \varepsilon_t \geq 2 \ln \lambda_{\min}(K_\infty)$$

if either K_∞ *is singular or all the matrices* $\{K_t;\ t = 0, 1, \ldots\}$ *are invertible.*

Recall that in the case where F is actually Gaussian, the MMSE and LLSE estimators agree so that $\varepsilon_t = 0$ for all $t = 1, 2, \ldots$.

We now present some straightforward implications of Theorems 7.4 and 7.6.

Theorem 7.7. *Assume* **(C.1)**. *If* $\rho(K_\infty) < 1$, *then*

- **(a)** *The sequence* $\mathbf{Q}^\star$ *converges to zero;*
- **(b)** *The sequence* $\mathbf{R}^\star$ *has a well–defined limit* $R_\infty^\star$; *and*
- **(c)** *For all distributions* F *in* $\mathcal{D}(\mathbb{R}^n)$, *the convergence* $\lim_t \varepsilon_t = 0$ *takes place at least exponentially fast according to Theorem 7.4.*

Proof. Claim (a) is almost a direct consequence of (7.14a), (2.3) and of the equivalence of norms on $\mathcal{M}_n$. To prove Claim (b), we should note via (7.13) that

$$\|R_t^\star - R_s^\star\|_{op} \leq \frac{\lambda_{\max}(H'H)}{\lambda_{\min}(\Sigma^v)} \sum_{r=s}^{t-1} \|Q_r^\star\|_{op}^2, \quad s < t. \quad s, t = 0, 1, \ldots$$

Since under assumption **(C.1)**, the convergence $\lim_t \|Q_t^\star\|_{op} = 0$ is exponentially fast when $\rho(K_\infty) < 1$, the sequence $\mathbf{R}^\star$ is thus Cauchy and hence convergent in $\mathcal{M}_n$ under the conditions of the theorem. Claim (c) is simply a qualititative interpretation of Theorem 7.4. ∎

Given the enormous literature on the discrete–time Ricatti equation, we should expect that there are some well–known conditions under which the hypotheses of Theorem 7.7 hold. We conclude this section with these results:

Theorem 7.8. *If the pair (A, H) is detectable, then* **(C.1)** *holds. If in addition, the pair $(\overline{A}, \overline{C}^{1/2})$ is stabilizable, where*

$$\overline{A} := A - \Sigma^{wv}(\Sigma^v)^{-1}H \quad \textit{and} \quad \overline{C} := \Sigma^w - \Sigma^{wv}(\Sigma^v)^{-1}\Sigma^{vw},$$

then condition **(C.1)** *holds and the matrix K_∞ is asymptotically stable, i.e., $\rho(K_\infty) < 1$.*

Proof. The first claim is Theorem 5.2(b) of [4, p. 172], while the second claim follows from Theorem 5.3 of [4, p. 175]. ∎

It is worth pointing out that the matrix $\bar{C}$ may be interpreted as the conditional covariance of W_{t+1}° given V_{t+1}° for each $t = 0, 1, \ldots$, and hence is positive semi–definite. Essentially, Theorem 7.8 gives a set of conditions under which the mean–square MMSE–LLSE error tends to zero. Note however that the hypotheses do *not* imply that either the MMSE or LLSE estimators provide good estimates of the plant process $\mathbf{X}^\circ$. Indeed, to elaborate somewhat upon this point, assume the stronger hypotheses that the pair (A, H) is detectable and the pair $(\bar{A}, \bar{C}^{1/2})$ is controllable. Then it is a classical result [5] that the recursion (1.7) for the error covariances associated with the LLSE estimates has a well–defined and positive definite limit P_∞^K which is the unique solution of the steady–state Ricatti equation

$$\begin{aligned} P_\infty^K &= AP_\infty^K A' + \Sigma^w \\ &\quad - [AP_\infty^K H' + \Sigma^{wv}][HP_\infty^K H' + \Sigma^v]^{-1}[AP_\infty^K H' + \Sigma^{wv}]'. \end{aligned}$$

Hence, under these stronger hypotheses, we conclude that

$$\lim_t \mathbf{E}^\circ[\|X_{t+1}^K - X_{t+1}^\circ\|^2] = \text{trace}\,(P_\infty^K) > 0. \tag{7.30}$$

Moreover, by the 'orthogonality principle' for MMSE estimators [2, Sec. 2.3], we readily get that

$$\mathbf{E}^\circ[\|X_{t+1}^K - X_{t+1}^\circ\|^2] = \mathbf{E}^\circ[\|\hat{X}_{t+1} - X_{t+1}^\circ\|^2] + \varepsilon_{t+1}. \quad t = 0, 1, \ldots \tag{7.31}$$

Now by combining (7.30) and the results of Theorems 7.7 and 7.8, we see from (7.31) that the relation

$$\lim_t \mathbf{E}^\circ[\|\hat{X}_{t+1} - X^\circ_{t+1}\|^2] = \text{trace}\,(P^K_\infty) > 0$$

also holds. Thus, under some appropriate assumptions, the performance of the LLSE and MMSE estimators are (essentially) equivalent for large times, i.e., nothing is (asymptotically) lost by using the simpler LLSE estimator rather than the more computationally demanding MMSE estimator.

VIII. A PARTIAL CONVERSE

This rather short section is devoted solely to the proof of Proposition 7.5. Before starting the discussion, we observe from (7.11) that a sufficient condition for $\underline{\lim}_t I_F(I_n, R^\star_t) = 0$ is that $\lambda_{\min}(Q^{\star\prime}Q^\star_t) > 0$ for all t large enough and

$$\lim_t \frac{\varepsilon_t((A,H,\Sigma),F)}{\lambda_{\min}(Q^{\star\prime}_t Q^\star_t)} = 0.$$

We suggest the following probabilistic interpretation for this: Whenever $\lambda_{\min}(Q^{\star\prime}Q^\star_t) > 0$ for all t large enough, *if* $\boldsymbol{\varepsilon}$ *decays quickly enough, then* ξ *must in fact be Gaussian, in which case* ε_t *is identically zero for all* $t = 1, 2, \ldots$. This not only provides an indirect characterization of the initial condition as a Gaussian or non–Gaussian rv, but also may be viewed as some form of *phase transition* as we vary the initial condition between Gaussian and non–Gaussian rvs.

Proof of Proposition 7.5. We base our study of the implications of $\lim_t I_F(I_n, R^\star_t) = 0$ on the formula (7.3).

A natural first step is to try to interchange the integration operations of (7.3) with the limiting operation of letting $R^\star_t$ tend to $R^\star_\infty$. To do this, we must study the behavior of both the integrand in (7.3) with $K = I_n$ and $R = R^\star_t$, *and* the measures $\{G_{R^\star_t};\, t = 1, 2, \ldots\}$ as we let t tend to infinity. We consider first the more difficult question of the probability measures $\{G_{R^\star_t};\, t = 1, 2, \ldots\}$. Assumption **(C.2)**, which is that the sequence $\mathbf{R}^\star$ has a well–defined and invertible limit $R^\star_\infty$, allows us to control the behavior of these probability measures by comparing them to Lebesgue measure λ on $\mathbb{R}^n$. Since the set of invertible $n \times n$ real matrices is an open subset of $\mathcal{M}_n$ [10, Thm. X.2.1], assumption **(C.2)** implies the existence of a finite T such

that for $t = T, T+1, \ldots$, the matrix $R_t^\star$ is invertible and the probability measure $G_{R_t^\star}$ is thus absolutely continuous with respect to λ for all such t. Fatou's lemma and the condition $\underline{\lim}_t\, I_F(I_n, R_t^\star) = 0$ then immediately imply

$$\underline{\lim}_t \frac{\left\| \int_{\mathbb{R}^n} \left\{ z - [R_t^\star + \Delta^{-1}]^{-1} b \right\} \gamma(z, b; R_t^\star) dF(z) \right\|^2}{\Gamma(b; R_t^\star)} \cdot \frac{dG_{R_t^\star}}{d\lambda}(b) = 0$$

for λ–almost all b in $\mathbb{R}^n$. Under **(C.2)**, we find

$$\lim_t \frac{dG_{R_t^\star}}{d\lambda}(b) = \frac{dG_{R_\infty^\star}}{d\lambda}(b) > 0, \quad b \in \mathbb{R}^n,$$

while by dominated convergence with the bound (7.6), we also have that

$$\lim_t \Gamma(b; R_t^\star) = \Gamma(b, R_\infty^\star) > 0, \quad b \in \mathbb{R}^n.$$

This allows us to focus on the numerator of the integrand of (7.3) and thus to conclude that if $\underline{\lim}_t\, I_F(I_n, R_t^\star) = 0$, then

$$\underline{\lim}_t \left\| \int_{\mathbb{R}^n} \left\{ z - [R_t^\star + \Delta^{-1}]^{-1} b \right\} \gamma(z, b; R_t^\star) dF(z) \right\| = 0$$

for λ–almost all b in $\mathbb{R}^n$. Appealing once more to dominated convergence and (7.6), we see that

$$\int_{\mathbb{R}^n} z \gamma(z, b; R_\infty^\star) dF(z) = [R_\infty^\star + \Delta^{-1}]^{-1} b \int_{\mathbb{R}^n} \gamma(z, b; R_\infty^\star) dF(z) \qquad (8.1)$$

for λ–almost all b in $\mathbb{R}^n$. This equation clearly enforces some constraints upon the distribution F—to complete the proof of Proposition 7.5 we shall show that in fact they imply that F is Gaussian.

The implications of (8.1) are not directly obvious. Things become a bit clearer, however, upon expanding $\gamma(z, b; R_t^\star)$; we get that

$$\begin{aligned} &\int_{\mathbb{R}^n} z \exp\left[z'b - \frac{1}{2} z' R_t^\star z \right] dF(z) \\ &\qquad = [R_\infty^\star + \Delta^{-1}]^{-1} b \int_{\mathbb{R}^n} \exp\left[z'b - \frac{1}{2} z' R_t^\star z \right] dF(z) \end{aligned} \qquad (8.2)$$

for λ–almost all b in $\mathbb{R}^n$. If we define an auxiliary probability measure $\hat{F}$ on $\mathbb{R}^n$ which is absolutely continuous with respect to F via its Radon–Nikodym derivative

$$\frac{d\hat{F}}{dF}(z) := \frac{\exp\left[-\frac{1}{2} z' R_\infty^\star z \right]}{\int_{\mathbb{R}^n} \exp\left[-\frac{1}{2} z' R_\infty^\star z \right] dF(z)}, \quad z \in \mathbb{R}^n, \qquad (8.3)$$

then (8.2) becomes

$$\int_{\mathbb{R}^n} z \exp[z'b] d\hat{F}(z) = [R_\infty^\star + \Delta^{-1}]^{-1} b \int_{\mathbb{R}^n} \exp[z'b] d\hat{F}(z),$$

for λ–almost all b in $\mathbb{R}^n$. Rewriting (8.1) in this way now suggests an analysis in terms of the moment generating function N of $\hat{F}$, where

$$N(b) := \int_{\mathbb{R}^n} \exp[z'b] d\hat{F}(z) = \int_{\mathbb{R}^n} \frac{\gamma(z, b; R_\infty^\star)}{\Gamma(0, R_\infty^\star)} dF(z), \quad b \in \mathbb{R}^n. \tag{8.4}$$

By (7.6) and some straightforward technical calculations [27, Sec. VI.2], we see that N is well defined and differentiable on all of $\mathbb{R}^n$, with

$$\nabla N(b) = \int_{\mathbb{R}^n} z \frac{\gamma(z, b; R_\infty^\star)}{\Gamma(0, R_\infty^\star)} dF(z), \quad b \in \mathbb{R}^n.$$

Comparing this last fact with (8.1) and (8.4), we find that N satisfies the differential equation

$$N(0) = 1, \quad \nabla N(b) = \left[R_\infty^\star + \Delta^{-1}\right]^{-1} b N(b), \quad b \in \mathbb{R}^n,$$

and by the uniqueness of solutions of ordinary differential equations we get

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

By the uniqueness of moment generating functions, the distribution $\hat{F}$ is now completely characterized—*it is a Gaussian distribution with zero mean and covariance* $R_\infty^\star + \Delta^{-1}$. The matrix $R_\infty^\star + \Delta^{-1}$ being positive definite, this implies that $\hat{F}$ is absolutely continuous with respect to λ, with Radon–Nikodym derivative

$$\frac{d\hat{F}}{d\lambda}(z) = \frac{\exp\left[-\frac{1}{2} z' [R_\infty^\star + \Delta^{-1}] z\right]}{(2\pi)^{n/2} \sqrt{\det\left[(R_\infty^\star + \Delta^{-1})^{-1}\right]}}, \quad z \in \mathbb{R}^n. \tag{8.5}$$

To complete the proof of the proposition, we now must show that since $\hat{F}$ is Gaussian, so is the original distribution F. An obvious implication of (8.3) and of the absolute continuity of $\hat{F}$ with respect to λ is that F is also absolutely continuous with respect to λ. In fact, combining (8.3) and (8.5), we get the formula

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

for some positive constant c. This completes our proof—F is Gaussian. ∎

IX. THE SCALAR CASE

In this section we continue the analysis of Section VII, here focusing exclusively on the scalar case $n = k = 1$. We shall see that a number of simplifications of (5.21), (6.5), (6.6) and (7.3) are now possible, allowing a much more complete study of the asymptotics of ε.

To emphasize the scalar nature of our calculations and to conform to standard notation, we shall use lower case letters for all of our scalar quantities. Thus

$$\Sigma = \begin{pmatrix} \sigma^w & \sigma^{wv} \\ \sigma^{wv} & \sigma^v \end{pmatrix}$$

with $\sigma^v > 0$, the system triple now is (a, h, Σ) and the variance of the initial condition is δ. Proposition 7.8 suggests that the quantities

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

will play a fundamental role in our ensuing analysis. We can more immediately see the utility of these quantities in the scalar case by rewriting (5.21), (6.5) and (6.6) in the forms

$$p_0 = 0, \quad p_{t+1} = \frac{\bar{a}^2 \sigma^v p_t}{h^2 p_t + \sigma^v} + \bar{c}, \qquad t = 0, 1, \ldots \quad (9.1)$$

$$q_0^\star = 1, \quad q_{t+1}^\star = k_t q_t^\star \qquad t = 0, 1, \ldots \quad (9.2)$$

with

$$k_t := \left(\frac{\bar{a} \sigma^v}{h^2 p_t + \sigma^v} \right), \qquad t = 0, 1, \ldots \quad (9.3)$$

and

$$r_0^\star = 0, \quad r_{t+1}^\star = r_t^\star + \frac{(q_t^\star)^2 h^2}{h^2 p_t + \sigma^v}. \qquad t = 0, 1, \ldots \quad (9.4)$$

We also have in the scalar case the advantage that the upper and lower bounds of (7.11) collapse into the equality

$$\varepsilon_t = (q_t^\star)^2 I_F(1, r_t^\star), \qquad t = 1, 2, \ldots \quad (9.5)$$

which holds for all F in $\mathcal{D}^\circ(\mathbb{R})$.

A natural way to organize our efforts here is to taxonomize the different possible cases based upon Proposition 7.8—that is, upon the detectability

of (a, h) and the stabilizability of $(\bar{a}, \bar{c}^{1/2})$. A slightly more direct classification scheme of the scalar problem, however, will emerge, yielding four possibilities parametrized by h, $\bar{a}$ and $\bar{c}$.

Our first case is an obvious degeneracy:

Proposition 9.1. *If either $\bar{a} = 0$ or $h = 0$, then $\varepsilon_t = 0$ for all $t = 1, 2, \ldots$ and all distributions F in $\mathcal{D}(\mathbb{R})$.*

Proof. Fix F in $\mathcal{D}^\circ(\mathbb{R})$. If $\bar{a} = 0$, then $q_t^\star = 0$ for all $t = 1, 2, \ldots$ by (9.2) and (9.3), and therefore (9.5) implies $\varepsilon_t = 0$ for all $t = 1, 2, \ldots$. On the other hand, if $h = 0$, then (9.4) yields $r_t^\star = 0$ for all $t = 0, 1, \ldots$, so again $\varepsilon_t = 0$ for all $t = 1, 2, \ldots$, this time by direct evaluation of (7.3). We translate these results from $\mathcal{D}^\circ(\mathbb{R})$ to $\mathcal{D}(\mathbb{R})$ by making use of the translation arguments of (7.1)–(7.2). ∎

We now consider the more interesting situation where both $\bar{a} \neq 0$ *and* $h \neq 0$, in which case $q_t^\star \neq 0$, $r_t^\star > 0$ and $\varepsilon_t > 0$ for all $t = 1, 2, \ldots$. We rewrite (9.1) as $p_{t+1} = T(p_t)$ where the mapping $T : [0, \infty) \to \mathbb{R}$ is given by

$$T(p) := \frac{\bar{a}^2 \sigma^v p}{h^2 p + \sigma^v} + \bar{c}, \quad p \geq 0. \tag{9.6}$$

Note that since

$$T'(p) = \frac{\bar{a}^2 (\sigma^v)^2}{(h^2 p + \sigma^v)^2}, \quad p \geq 0,$$

the mapping T is concave and nondecreasing on $[0, \infty)$. Hence, the iterates $\{p_t,\ t = 0, 1, \ldots\}$ form a nondecreasing and 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 $\mathbf{k}$ has a limit k_∞ given by

$$k_\infty := \frac{\bar{a} \sigma^v}{h^2 p_\infty + \sigma^v}$$

with $|k_\infty| > 0$ since $\bar{a} \neq 0$ and $p_\infty < \infty$. By Cesaro convergence, this implies via (9.2) that

$$\lim_t \frac{1}{t} \ln (q_t^\star)^2 = 2 \ln |k_\infty|.$$

It is then easy to see from (7.13) that if $|k_\infty| < 1$, then $r_\infty^\star := \lim_t r_t^\star$ is well defined and finite, whereas if $|k_\infty| \geq 1$, then $\lim_t r_t^\star = \infty$. These observations give us the second case of interest:

Proposition 9.2. *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 $\lim_t \varepsilon_t = 0$ with $\lim_t \frac{1}{t} \ln \varepsilon_t = 2 \ln |k_\infty| < 0$ for all non–Gaussian distributions F in $\mathcal{D}(\mathbb{R})$.*

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 $t = 0, 1, \ldots$, 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 $\bar{c} > 0$ by the first remark following Theorem 7.8 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 relation $T'(p_\infty) = k_\infty^2$.

As pointed out earlier, if both $h \neq 0$ and $\bar{a} \neq 0$, then $q_t^\star \neq 0$ and $r_t^\star > 0$ for all $t = 0, 1, \ldots$, whence $r_\infty^\star > 0$ since $\{r_t^\star,\ t = 0, 1, \ldots\}$ is an increasing sequence. On the other hand, we saw earlier that $|k_\infty| < 1$ implies $r_\infty^\star < \infty$. Therefore, from Propositions 7.3 and 7.5, we obtain $0 < \underline{\lim}_t\ I_F(1, r_t^\star) \leq \overline{\lim}_t\ I_F(1, r_t^\star) < \infty$ for every non–Gaussian F in $\mathcal{D}^\circ(\mathbb{R})$. As a result, $\lim_t \frac{1}{t} \ln \varepsilon_t = \lim_t \frac{1}{t} \ln (q_t^\star)^2 = 2 \ln |k_\infty| < 0$ for all F non–Gaussian in $\mathcal{D}^\circ(\mathbb{R})$, and thus in $\mathcal{D}(\mathbb{R})$ by translation. ∎

Notice that Proposition 9.2 is almost a direct consequence of Theorems 7.4 and 7.6 since in the scalar case, we have $\lambda_{\min}(k_\infty) = \rho(k_\infty) = |k_\infty|$, and we thus would need only to establish that conditions **(C.1)** and **(C.2)** hold true under the assumptions of Proposition 9.2. However, we found the more direct argument involving (9.6) to be an interesting calculation 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 in this case the initial state distribution F has a nontrivial effect on the large time asymptotics of $\boldsymbol{\varepsilon}$. *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 LLSE filters. However, in both cases considered thus far in Propositions 9.1 and 9.2, the effect of the system parameters (a, h, Σ) 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 giving a complete analysis for two specific initial distributions F, and by noting the different asymptotics of

$\boldsymbol{\varepsilon}$. We first verify a general result which complements Proposition 7.5.

Lemma 9.3. *For any distribution F in $\mathcal{D}(\mathbb{R})$, we have $I_F(1,r) \leq \frac{4}{r}$ for all $r > 0$ so that $\lim_r I_F(1,r) = 0$.*

Proof. Since the functional $I_F(1,\cdot)$ is *independent* of the system triple (a,h,Σ), we can study it by choosing the system (1.1a)–(1.1b) at our convenience. In particular, we shall take the system (1.1a)–(1.1b) to be

$$x_t^\circ = \xi, \quad y_t = \xi + v_{t+1}^\circ, \qquad\qquad t = 0,1,\ldots$$

i.e., $a = h = 1$ and $\sigma^w = \sigma^{wv} = 0$, in which case $q_t^\star = 1$, $r_t^\star = t/\sigma^v$ and $\varepsilon_t = I_F(1, t/\sigma^v)$ for all $t = 0,1,\ldots$. We now define a process $\check{\mathbf{x}} = \{\check{x}_t;\, t = 1,2,\ldots\}$ of linear estimates of the process $\mathbf{x}^\circ$ on the basis of $\mathbf{y}$ as

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

Invoking (7.31), we get that

$$\begin{aligned} &\mathbf{E}^\circ[|\hat{x}_{t+1} - x_{t+1}^K|^2] \\ &\quad \leq \mathbf{E}^\circ[|x_{t+1}^K - x_{t+1}^\circ|^2] \\ &\quad \leq \mathbf{E}^\circ[|\check{x}_{t+1} - x_{t+1}^\circ|^2], \qquad t = 0,1,\ldots \end{aligned} \tag{9.7}$$

where the last inequality follows from the minimizing definition of the LLSE estimator and the fact that $\check{\mathbf{x}}$ is a sequence of linear estimates. Therefore, we conclude that

$$I_F\left(1, \frac{t}{\sigma^v}\right) = \varepsilon_t \leq \mathbf{E}^\circ\left[\left|\frac{1}{t}\sum_{s=0}^{t-1} v_{s+1}^\circ\right|^2\right] = \frac{\sigma^v}{t}, \qquad t = 1,2\ldots$$

and the result now follows since σ^v is arbitrary. ∎

Whereas Lemma 9.3 gives us a uniform upper bound similar to that of Proposition 7.5, our real interest in the next case is to show that if the system triple (a,h,Σ) is of a specific form, then the asymptotics of $\boldsymbol{\varepsilon}$ are *not* uniform over all non–Gaussian initial distributions F—the asymptotics of $\boldsymbol{\varepsilon}$ depend *nontrivially* on F. We shall use the following two types of non–Gaussian distributions as examples of this:

Distribution $\mathbf{F_1}$: Distribution F_1 admits a density with respect to Lebesgue measure λ on $\mathbb{R}$ of the form

$$\frac{dF_1}{d\lambda}(z) = \sum_{i=1}^{m} \alpha_i \frac{1}{\sqrt{2\pi\rho^2}} \exp\left[-\frac{1}{2}\frac{(z-\mu_i)^2}{\rho^2}\right], \quad z \in \mathbb{R},$$

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

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

Distributions of the type F_1 have been considered before in the context of filtering theory [1], [2, Sec. 8.4], [26]. The following two important facts about F_1 and F_2 are proved in [27]:

Fact 1. *We have*

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

for some $K > 0$.

Fact 2. *We also have*

$$I_{F_2}(1, r) = \frac{1 + o(1)}{r}, \quad r > 0.$$

We now can prove the following results, which concern the third case:

Proposition 9.4. *If $h \neq 0$, $|\bar{a}| = 1$ and $\bar{c} = 0$, then $\lim_t \varepsilon_t = 0$ for any distribution F in $\mathcal{D}(\mathbb{R})$, with $\overline{\lim}_t \frac{1}{t} \ln \varepsilon_t \leq 0$. This convergence takes place at a rate which depends nontrivially upon F for non–Gaussian F.*

Proof. Under the stated hypothesis, we have $p_t = 0$, $(q_t^\star)^2 = 1$, $r_t^\star = h^2 t/\sigma^v$ and $\varepsilon_t = I_F(1, h^2 t/\sigma^v)$ for all $t = 0, 1, \ldots$ and all F in $\mathcal{D}^\circ(\mathbb{R})$, the extension to $\mathcal{D}(\mathbb{R})$ being as before. The conclusions $\lim_t \varepsilon_t = 0$ and $\overline{\lim}_t \frac{1}{t} \ln \varepsilon_t \leq 0$ are immediate consequences of Lemma 9.3. However, direct calculations show that if the initial distribution is F_1, then $\lim_t t^2 \varepsilon_t = K(\sigma^v)^2/(\rho^4 h^4)$, whereas if the initial distribution is F_2, then $\lim_t t\varepsilon_t = \sigma^v/h^2$ (thus $\lim_t \frac{1}{t} \ln \varepsilon_t = 0$ in both cases.) ∎

And finally, the fourth case, which like Proposition 9.4 displays a nontrivial dependence on the initial distribution, is

Proposition 9.5. *If $h \neq 0$, $|\bar{a}| > 1$ and $\bar{c} = 0$, then $\overline{\lim}_t \varepsilon_t < \infty$ for all distributions F in $\mathcal{D}(\mathbb{R})$, the asymptotic behavior depending nontrivially upon F for non-Gaussian F.*

Proof. Under the stated hypotheses on (a, h, Σ), $p_t = 0$, $(q_t^\star)^2 = \bar{a}^{2t}$, $r_t^\star = \frac{h^2}{\sigma^v} \frac{\bar{a}^{2t}-1}{\bar{a}^2-1}$ for all $t = 0, 1, \ldots$. Thus $\lim_t r_t^\star = \infty$ with $\lim_t (q_t^\star)^2/r_t^\star = \sigma^v(\bar{a}^2 - 1)/h^2$ and we are led to write

$$\varepsilon_t = \frac{(q_t^\star)^2}{r_t^\star}(r_t^\star I_F(1, r_t^\star)) \leq \frac{\sigma^v}{h^2}(\bar{a}^2 - 1) \cdot \frac{\bar{a}^{2t}}{\bar{a}^{2t} - 1}, \qquad t = 1, 2, \ldots \quad (9.8)$$

where the inequality follows from Lemma 9.3. We now see that $\overline{\lim}_t \varepsilon_t < \infty$ for all F in $\mathcal{D}^\circ(\mathbb{R})$, and thus for all distributions F in $\mathcal{D}(\mathbb{R})$. However, if ξ has distribution F_1, then $\lim_t \varepsilon_t = 0$, whereas if ξ has distribution F_2, then $\lim_t \varepsilon_t = \sigma^v(\bar{a}^2 - 1)/h^2$. ∎

This essentially completes our analysis of the scalar case, as all possible combinations of $\bar{a}$, $\bar{c}$ and h have now been considered.

A review of the arguments thus far reveals that a large portion of our calculations relied upon scalar manipulations which are not readily extendable to the multivariable case. Several remarks, however, do give us some information and intuition concerning the multivariable case. First, from (7.31) we see by an argument similar to the one leading to (9.7) that

$$\varepsilon_t \leq \mathbf{E}^\circ[|x_t^\circ - x_t^K|^2] = p_t^K \qquad t = 1, 2, \ldots$$

where the error variances $\{p_t^K,\ t = 0, 1, \ldots\}$ are generated through the recursion (9.1) (which is nothing but (1.7)) with initial condition $p_0^K = \delta$. The sequence $\{p_t^K,\ t = 0, 1, \ldots\}$ is either monotone nondecreasing or monotone nonincreasing, and thus convergent, with limit point p_∞^K. Therefore, whenever $p_\infty^K < \infty$, we conclude by inspection that

$$\varepsilon_t \leq \max\{\delta, p_\infty^K\}. \qquad t = 1, 2, \ldots \quad (9.9)$$

In particular, under the conditions of Proposition 9.5, i.e., $h \neq 0$, $|\bar{a}| > 1$ and $\bar{c} = 0$, we have (9.9) with $p_\infty^K = \sigma^v(\bar{a}^2 - 1)/h^2$ (a fact which is of course in agreement with (9.8)). Analogous calculations using (7.31) can be made in the multivariable case to yield a bound corresponding to (9.9).

A second comment on the multivariable case is the following. Our analysis suggests the following classification: For any matrices $\tilde{A}$ and $\tilde{C}$ in $\mathcal{M}_n$, the pair $(\tilde{A}, \tilde{C})$ is said to be *marginally stabilizable* if all modes which are neither stable nor critically stable, are in the controllable subspace. 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 9.6. *We have the following convergence results:*

1a. If the pair $(\bar{a}, \bar{c})$ is marginally stabilizable, then $\lim_t \varepsilon_t = 0$ for any distribution F in $\mathcal{D}(\mathbb{R})$; and

1b. If the pair $(\bar{a}, \bar{c})$ is not marginally stabilizable, then the asymptotic behavior of ε depends nontrivially upon F in $\mathcal{D}(\mathbb{R})$.

Moreover we also have the following:

2a. If $(\bar{a}, \bar{c})$ is stabilizable, then $\lim_t \varepsilon_t = 0$ at an exponential rate independent of F for non–Gaussian F in $\mathcal{D}(\mathbb{R})$; and

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

X. ACKNOWLEDGMENTS

The first author was supported during the preparation of this work by ONR Grant N00014-91-J-1526. The work of the second author was supported partially through the NSF Engineering Research Centers Programs under grant CDR 88-03012 and partially through NSF grant ECS 83-51836.

XI. REFERENCES

[1] D. Alspach, "Gaussian sum approximations in nonlinear filtering," *Information Sciences* **7** (1974), pp. 271–290.

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

[3] V. Beneš and I. Karatzas, "Estimation and control for linear partially observed systems with non–Gaussian initial distribution," *Stochastics Processes and Their Applications* **14** (1983), pp. 233–248.

[4] P.E. Caines, *Linear Stochastic Systems*, John Wiley & Sons, New York (NY) (1988).

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

[6] M.H.A. Davis, *Linear Estimation and Stochastic Control*, Chapman and Hall, London (United Kingdom) (1977).

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

[8] R. Elliott, *Stochastic Calculus and Applications*, Springer–Verlag, New York (NY) (1982).

[9] I. Girsanov, "On transforming a certain class of stochastic processes by absolutely continuous substitution of measures," *Theory of Probability and Its Applications* **5** (1960), pp. 285–301.

[10] I. Gohberg and S. Goldberg, *Basic Operator Theory*, Birkhäuser, Boston (MA) (1981).

[11] U.G. Haussman and E. Pardoux, "A conditionally almost linear filtering problem with non–Gaussian initial condition," *Stochastics* **23** (1988), pp. 241–275.

[12] G. Kallianpur, *Stochastic Filtering Theory*, Springer–Verlag, New York (NY) (1980).

[13] R.E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Engr., Trans. ASME, Series D* **82** (1960), pp. 35–45.

[14] R.E. Kalman and R.S. Bucy, "New results in linear filtering and prediction theory," *J. Basic Engr., Trans. ASME, Series D* **83** (1961), pp. 95–108.

[15] H. Kunita, "Asymptotic behavior of the nonlinear filtering errors of Markov processes," *Journal of Multivariate Analysis* **1** (1971), pp. 365–393.

[16] H. Kunita, "Ergodic properties of nonlinear filtering processes," in *Spatial Stochastic Processes: Festschrift in Honor of Ted Harris on His Seventieth Birthday*, Eds. K. Alexander and J. Watkins,

Progress in Probability **19** (1992), Birkhauser, Boston (MA), pp. 233–256.

[17] R. Liptser and A. Shiryayev, *Statistics of Random Processes I: General Theory*, Springer–Verlag, New York (NY) (1977).

[18] M. Loéve, *Probability Theory II*, Fourth Edition, Springer-Verlag, New York (NY) (1978).

[19] A.M. Makowski, "Results on the filtering problem for linear systems with non–Gaussian initial condition," in *Proceedings of the 21st IEEE Conference on Decision and Control*, Orlando (FL), December 1982, pp. 201–214.

[20] A.M. Makowski, W.S. Levine and M. Asher, "The nonlinear MMSE filter for tracking partially observable randomly maneuvering targets," in *Proceedings of the 7th MIT/ONR Workshop on C^3 Systems*, San Diego (CA), June 1984, pp. 191–196.

[21] A.M. Makowski, W.S. Levine and M. Asher, "The nonlinear MMSE filter for partially observed systems driven by non–Gaussian white noise, with applications to failure estimation," in *Proceedings of the 23rd IEEE Conference on Decision and Control*, Las Vegas (NV), December 1984, pp. 644–650.

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

[23] J. Neveu, *Discrete–Parameter Martingales*, Translated by T.P. Speed, North–Holland, New York (NY) (1975).

[24] E. Pardoux, "Filtrage non linéaire et équations aux dérivées partielles stochastiques associées," in *Ecole d'Eté de Probabilités de Saint Flour XIX–1989*, Ed. P.–L. Hennequin, Springer *Lecture Notes in Mathematics* **1464** (1991), Springer, New York (NY), pp. 71–158.

[25] B.L. Rozovskii, *Stochastic Evolution Systems: Linear Theory and Applications to Nonlinear Filtering*, Kluwer Academic Publishers, Boston (MA) (1990).

[26] H. Sorenson and D. Alspach, "Recursive Bayesian estimation using Gaussian sums," *Automatica* **7** (1971), pp. 465–479.

[27] R.B. Sowers, *New Discrete–Time Filtering Results*, M.S. Thesis, Electrical Engineering Department, University of Maryland, College Park (MD), August 1988. Also available as Systems Research Center Technical Report **TR 88–85** (1988).

[28] R.B. Sowers and A.M. Makowski, "Discrete–time filtering for linear systems in correlated noise with non–Gaussian initial conditions," in *Proceedings of the 23rd Annual Conference on Information Sciences and Systems*, The Johns Hopkins University, Baltimore (MD), March 1989, pp. 308–313.

[29] R.B. Sowers and A.M. Makowski, "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*, Eds. E M.A. Kaashoek, J.H. van Schuppen and A.C.M. Ran, Birkhäuser Boston Inc., Cambridge (MA) (1990), Proceedings of the 1989 International Symposium on the Mathematical Theory of Networks and Systems, Amsterdam (The Netherlands), June 1989, pp. 407–419.

[30] 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," *IEEE Transactions on Automatic Control* **AC–37** (1992), pp. 114–120.

[31] G.W. Stewart, *Introduction to Matrix Computations*, Academic Press, New York (NY) (1973).

[32] M. Zakai, "On the optimal filtering of diffusion processes," *Z. Wahrsch. verw. Gebiete* **11** (1969), pp. 230–243.