

A SIMPLE ASYMPTOTICALLY OPTIMAL FILTER OVER AN INFINITE HORIZON

P. CHIGANSKY, R. LIPTSER and B.Z. BOBROVSKY

Tel Aviv University

Department of Electrical Engineering-Systems

69978 Tel Aviv, Israel

E-mail: pavelm@eng.tau.ac.il, lipster@eng.tau.ac.il, bobrov@eng.tau.ac.il

(Received October, 1999; Revised October, 2000)

A filtering problem over an infinite horizon for a continuous time signal and discrete time observation in the presence of non-Gaussian white noise is considered. Conditions are presented, under which a nonlinear Kalman type filter with limiter is asymptotically optimal in the mean square sense for long time intervals given provided the sampling frequency is sufficiently high.

Key words: Kalman Filter, Limiter, Asymptotic Optimality, Infinite Horizon, Lower Error Bound, Non-Gaussian Noise.

AMS subject classifications: 60G35, 93E11.

1. Introduction

Consider the infinite horizon filtering problem for a continuous time signal $(X_t)_{t \geq 0}$ observed only at time points $t_k, k = 0, 1, \dots$ where $t_k - t_{k-1} \equiv \Delta$. Suppose the signal is distorted by a linear transformation $AX_{t_{k-1}}$ and corrupted by white noise proportional to $\frac{1}{\sqrt{\Delta}}$, that is $y_k^\Delta = AX_{t_{k-1}} + \frac{1}{\sqrt{\Delta}}\xi_k$. For convenience, we introduce a continuous time process

$$Y_t^\Delta = \begin{cases} 0 & t = 0, \\ Y_{t_{k-1}}^\Delta & t_{k-1} \leq t < t_k, k \geq 1, \end{cases}$$

where $Y_{t_k}^\Delta = Y_{t_{k-1}}^\Delta + y_k^\Delta \Delta$ with similar signal information as $(y_k^\Delta)_{k \geq 1}$.

Both (X_t) and (Y_t^Δ) are assumed to be vector processes of sizes n and ℓ respectively. The signal is a homogeneous diffusion process with respect to the vector Wiener process $(W_t)_{t \geq 0}$ with independent components. The initial condition X_0 is the random vector with $E \| X_0 \|^2 < \infty$. Then the filtering model

$$dX_t = aX_t dt + b dW_t \tag{1.1}$$

$$Y_{t_k}^\Delta - Y_{t_{k-1}}^\Delta = AX_{t_{k-1}}^\Delta + \xi_k \sqrt{\Delta}, \quad (1.2)$$

where matrices A , a and b are known may be analyzed as follows. The noise $(\xi_k)_{k \geq 0}$ forms an i.i.d. sequence of random vectors independent of $(W_t)_{t \geq 0}$ and X_0 . Components of the vector ξ_1 are independent random variables and their distributions obey densities so that the Fisher information of each component of ξ_1 is well defined and positive. The assumptions made about the distribution of ξ_1 imply that the Fisher information matrix \mathcal{I} of ξ_1 is scalar and nonsingular.

If $\mathbf{E} \|\xi_1\|^2 < \infty$, the Kalman filter with continuous time signal and discrete time observation, that is being adapted to the setting, might be used as an optimal linear filter in the mean square sense. If $\mathbf{E} \|\xi_1\|^2$ is infinite or even if $\mathbf{E} \|\xi_1\|^2$ is finite but too large, the Kalman filter becomes useless though the optimal filtering estimate

$$\pi_t(Y^\Delta) = \mathbf{E}(X_t | Y_{t_k}^\Delta, t_k \leq t)$$

would probably be reasonable from an applied view of filtering quality. Such situations, where $\mathbf{E} \|\xi_1\|^2$ is large because of “heavy tails” of the distributions of ξ_1 components are typical in engineering practice.

An essential role in the verification of a nonlinear filter quality is to be the lower bound for the mean square filtering errors matrix

$$V_t^\Delta = \mathbf{E}(X_t - \pi_t(Y^\Delta))(X_t - \pi_t(Y^\Delta))^T$$

where T is the transpose operator.

The lower bound for V_t^Δ is found by a method borrowed from Bobrovsky and Zakai [2,3], Bobrovsky, Zakai, and Zeitouni [4], Bobrovsky, Mayer-Wolf, and Zeitouni [5]. Under the assumption that the Fisher information matrix \mathcal{I}_0 of the random vector X_0 exists, we show that

$$V_t^\Delta \geq P_t^\Delta, \quad t \geq 0 \quad (1.3)$$

in the sense that $V_t^\Delta - P_t^\Delta$ is a nonnegative definite matrix. Here, P_t^Δ is the filtering mean square error matrix for Gaussian filtering model

$$d\check{X}_t = a\check{X}_t dt + b dW_t$$

$$\check{Y}_{t_k}^\Delta - \check{Y}_{t_{k-1}}^\Delta = A\check{X}_{t_{k-1}}^\Delta + \mathcal{I}^{-1/2} \check{\xi}_k \sqrt{\Delta}$$

subject to $\check{Y}_0^\Delta = 0$, \check{X}_0 is the Gaussian vector with $\mathbf{E}\check{X}_0 = \mathbf{E}X_0$, the covariance matrix is equal to \mathcal{I}_0^+ (\mathcal{I}_0^+ is the Moore-Penrose pseudoinverse matrix [1]) and $(\check{\xi}_k)_{k \geq 1}$ is an i.i.d. sequence of zero mean Gaussian vectors with unit covariance matrices.

In general, this lower bound might be unattainable. However, in this paper we show that it can be closely approached if the parameter Δ is small enough. Moreover, we use this fact to construct a nearly optimal filtering estimate by applying a Kalman type filter to a nonlinearly preprocessed observation. The use of a preliminary nonlinear transformation (limiter), to improve filtering accuracy is well known from Kushner [8], Kushner and Runggaldier [9], Liptser and Lototsky [10], Liptser and Runggaldier [12], Liptser and Zeitouni [14], and Liptser and Muzhikanov [11]. In [11] and [14], the choice of an appropriate limiter depends on the diffusion approximation of the observation process so that the drift and the diffusion parameters in the associated diffusion limit determine the signal to noise ratio, and

thus, the filtering quality. Moreover, the filter with the limiter which gives the largest signal to noise ration (for the limit model) turns out to be optimal in the mean square sense, as $\Delta \rightarrow 0$ for any finite interval. This is readily verified by applying Goggin's approximation [7] for the conditional expectation.

In this paper, we consider the filtering problem on the infinite time interval (infinite horizon) in the situations when Goggin's approach can not be effectively extended for the analysis of $\lim_{\Delta \rightarrow 0} \lim_{t \rightarrow \infty}$ asymptotic. Our approach exploits the lower bound from (1.3) and conditions under which

$$\lim_{\Delta \rightarrow 0} \lim_{t \rightarrow \infty} P_t^\Delta = P, \tag{1.4}$$

where P is a positive definite matrix. We show that the estimate $\tilde{\pi}_t(Y^\Delta)$, generated by the Kalman filter with limiter, satisfies the asymptotic optimal property

$$\lim_{\Delta \rightarrow 0} \lim_{t \rightarrow \infty} E(X_t - \tilde{\pi}_t(Y^\Delta))(X_t - \tilde{\pi}_t(Y^\Delta))^T = P. \tag{1.5}$$

This paper is organized as follows. In Section 2, we make a revision of the finite horizon case and show that the filter, with the relevant limiter, guarantees the asymptotic optimal property in the sense that

$$\lim_{\Delta \rightarrow 0} V_t^\Delta = \lim_{\Delta \rightarrow 0} P_t^\Delta, \quad \forall t > 0.$$

In Section 3, the main infinite horizon result is presented. Several practical aspects are discussed and demonstrated with computer simulations in Section 4.

2. Preliminaries

2.1 A Diffusion Approximation and the Filter

As in [11], to keep the heredity of property of asymptotic optimality for a wider range of sampling intervals Δ , the limiter is applied to the innovation difference.

The limiter is chosen to be a column vector-function G with components $G_i(x) = -\frac{p'_i(x)}{p_i(x)}$, $i = 1, \dots, \ell$, where $p_i(x)$ is the distribution density of the i th component of ξ_1 . To incorporate the limiter into the filtering algorithm, Y_t^Δ is transformed into

$$\tilde{Y}_t^\Delta = \sum_{k=1}^{[t/\Delta]} \sqrt{\Delta} \mathcal{T}^{-1} G \left(\frac{Y_{t_k}^\Delta - Y_{t_{k-1}}^\Delta - A \tilde{\pi}_{t_{k-1}}(Y^\Delta) \Delta}{\sqrt{\Delta}} \right). \tag{2.1}$$

Here, $[x]$ stands for the integer part of x , $\tilde{\pi}_t(Y^\Delta)$ is a random process defined by the linear equation

$$d\tilde{\pi}_t(Y^\Delta) = a\tilde{\pi}_t(Y^\Delta)dt + P_t A^T \mathcal{T} d\tilde{Y}_t^\Delta, \quad \tilde{\pi}_0 = EX_0, \tag{2.2}$$

and P_t is a solution of the Riccati equation

$$\dot{P}_t = aP_t + P_t a^T + bb^T - P_t A^T \mathcal{T} A P_t, \tag{2.3}$$

subject to $P_0 = \text{cov}(X_0, X_0)$ in which the Fisher information \mathcal{T} is a scalar matrix with diagonal elements

$$\mathcal{T}_{ii} = \int_{\mathbb{R}} \frac{(p'_i(x))^2}{p_i(x)} dx, \quad i = 1, \dots, \ell.$$

The process $\tilde{\pi}_t(Y^\Delta)$ is suggested as the nonlinear filtering estimate for the signal X_t given $\{Y_s^\Delta, 0 \leq s \leq t\}$. It should be noted that $Y_{t_k}^\Delta - Y_{t_{k-1}}^\Delta - A\tilde{\pi}_{t_{k-1}}(Y^\Delta)\Delta$ is regarded as “innovation difference” and the proposed filter can be seen as a Kalman filter with a limiter which is applied to the innovation process (see the block diagram on Figure 1).

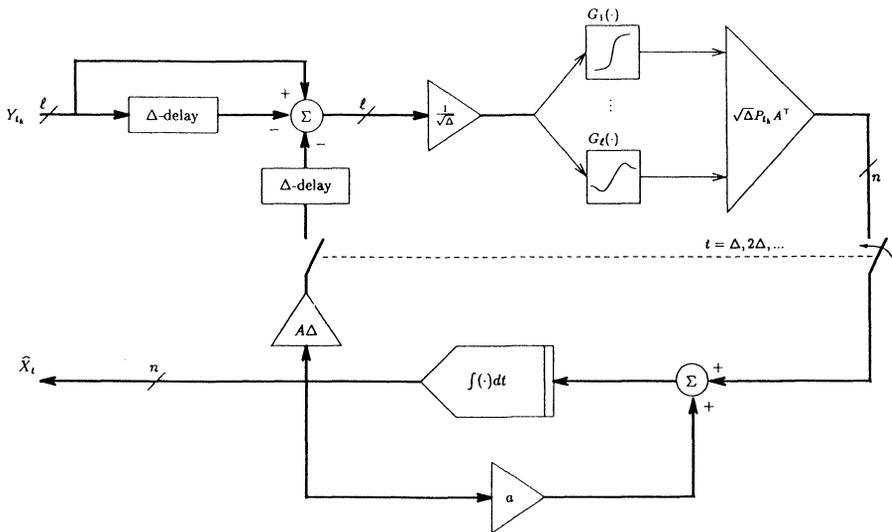


Figure 1. Block diagram of the proposed filter.

The choice of such a nonlinear filter is warranted by the diffusion approximation arguments given below. We fix the following assumptions

$G_i(x)$ is continuously differentiable

$$E[G'_i(\xi_1^{(i)})]^2 < \infty$$

$$EG_i^4(\xi_1) < \infty$$

$$|G_i(x)| \leq C(1 + |z|), \text{ for some } C > 0, \tag{2.4}$$

and extend the proof of Theorem 2.1 from [11] to the vector case. As in [11], we have

$$(X_t, \tilde{Y}_t^\Delta, \tilde{\pi}_t(Y^\Delta)) \xrightarrow[\Delta \rightarrow 0]{\text{law}} (X_t, \tilde{Y}_t, \tilde{\pi}_t(\tilde{Y})), \tag{2.5}$$

where $\tilde{Y}_0 = 0, \tilde{\pi}_0 = EX_0$ and

$$dX_t = aX_t dt + bdW_t \tag{2.6}$$

$$d\tilde{Y}_t = A(X_t - \tilde{\pi}_t(\tilde{Y}))dt + \mathcal{T}^{-1/2}dV_t$$

with Wiener process $(V_t)_{t \geq 0}$ independent of $(X_t)_{t \geq 0}$ and

$$d\tilde{\pi}(\tilde{Y})_t = a\tilde{\pi}(\tilde{Y})_t dt + P_t A^T \mathcal{T} d\tilde{Y}_t. \quad (2.7)$$

Despite the trajectories of the prelimit triple $(x_t, \tilde{Y}_t^\Delta, \tilde{\pi}_t(Y^\Delta))$ which belong to ${}^1\mathbb{C}_{[0,\infty)}^n \times \mathbb{D}_{[0,\infty)}^\ell \times \mathbb{D}_{[0,\infty)}^n$, the trajectories of the limit $(X_t, \tilde{Y}_t, \tilde{\pi}_t(\tilde{Y}))$ belong to $\mathbb{C}_{[0,\infty)}^n \times \mathbb{C}_{[0,\infty)}^\ell \times \mathbb{C}_{[0,\infty)}^n$, so that “ $\xrightarrow{\text{law}}$ ” holds in the local supremum topology with the metric

$$\rho(u', u'') = \sum_{m=1}^{\infty} 2^{-m} \{1 \wedge \sup_{t \leq m} h(t)\}, \quad u', u'' \in \mathbb{C}_{[0,\infty)}^n \times \mathbb{D}_{[0,\infty)}^\ell \times \mathbb{D}_{[0,\infty)}^n,$$

where $u' = (x'_t, y'_t, z'_t)$, $u'' = (x''_t, y''_t, z''_t)$ and

$$h(t) = \sum_{i=1}^n |x'_t(i) - x''_t(i)| + \sum_{i=1}^{\ell} |y'_t(i) - y''_t(i)| + \sum_{i=1}^n |z'_t(i) - z''_t(i)|.$$

Consider now the filtering problem for the limit pair (X_t, Y_t) (see (2.6)) in which X_t is treated as the signal and \tilde{Y}_t as the observation. Clearly, the optimal, in the mean square sense, linear filtering estimate is generated by the Kalman filter given by equations (2.7) and (2.3). Note also that the nonlinear filter proposed in (2.2) is nothing but this Kalman filter with \tilde{Y}_t replaced by \tilde{Y}_t^Δ .

2.2 A Lower Bound Over a Finite Horizon

Let us introduce the essential notation. Set

$$\begin{aligned} x_k &= X_{t_k}, \quad \text{and} \quad y_k = Y_{t_k}^\Delta - Y_{t_{k-1}}^\Delta \\ \hat{x}_k &= \mathbf{E}(x_k | y_j, 1 \leq j \leq k) \\ V_{t_k}^\Delta &= \mathbf{E}(x_k - \hat{x}_k)(x_k - \hat{x}_k)^\top \\ x_k &= e^{a\Delta} x_{k-1} + \int_{t_{k-1}}^{t_k} e^{a(t_k-s)} b dW_s \\ P_{t_k}^\Delta &= e^{a\Delta} P_{t_{k-1}}^\Delta e^{a^\top \Delta} + \int_0^\Delta e^{as} b b^\top e^{a^\top s} ds \\ &- e^{a\Delta} P_{t_{k-1}}^\Delta A^\top (\mathcal{T}^{-1} + A P_{t_{k-1}}^\Delta A^\top \Delta)^{-1} A P_{t_{k-1}}^\Delta e^{a^\top \Delta} \Delta, \quad k \geq 1 \\ P_0^\Delta &= \mathcal{T}_0^+. \end{aligned} \quad (2.8)$$

Lemma 2.1: *Assume*

$$\Gamma_{(ba)} = \begin{pmatrix} b & ab & \dots & a^{n-1}b \end{pmatrix}_{n \times (n \cdot m)} \quad (FR)$$

${}^1\mathbb{C}_{[0,\infty)}^r, \mathbb{D}_{[0,\infty)}^r$ are the spaces of continuous and right continuous (having limits to the left) vector-functions of size r .

where the block-matrix has the full rank.

Then, for every $k \geq 0$

$$V_{t_k}^\Delta \geq P_{t_k}^\Delta.$$

Proof: Under (FR), for every $t > 0$ the distribution of X_t obeys a density (e.g., see the proof of Lemma 16.3 in [13]). Hence, the joint distribution of $\zeta_k = (x_k, \dots, x_1)$ has a density as well. For convenience, introduce the vector $\theta_k = (y_k, \dots, y_1)$. The signal and noise independence guarantees the existence of a smooth and strictly positive distribution density $p(u, v)$ for (ζ_k, θ_k) . Set

$$\begin{aligned} f(u) &= \int_{\mathbb{R}^{k\ell}} p(u, v) dv \quad (\text{density of } \zeta_k) \\ g(v) &= \int_{\mathbb{R}^{kn}} p(u, v) du \quad (\text{density of } \theta_k) \\ q(u | v) &= \frac{p(u, v)}{g(v)} \quad (\text{conditional density of } \zeta_k \text{ given } \theta_k) \\ r(v | u) &= \frac{p(u, v)}{f(u)} \quad (\text{conditional density of } \theta_k \text{ given } \zeta_k) \end{aligned} \quad (2.9)$$

and denote by $\nabla_u q(u | v)$ the row gradient vector with respect to u . Let us define a nonnegative definite matrix

$$I_k = \int_{\mathbb{R}^{k\ell}} \int_{\mathbb{R}^{kn}} \frac{\nabla_u^T q(u | v) \nabla_u q(u | v)}{q(u | v)} g(v) du dv. \quad (2.10)$$

We show that the mean square error for the estimate $\widehat{\zeta}_k = \mathbf{E}(\zeta_k | \theta_k)$ of ζ_k given θ_k is bounded below (a version of the Rao-Crammer inequality):

$$\mathbf{E}(\zeta_k - \widehat{\zeta}_k)(\zeta_k - \widehat{\zeta}_k)^T \geq I_k^+, \quad (2.11)$$

where I_k^+ is the Moore-Penrose pseudoinverse matrix for I_k . The inequality (2.11) becomes an equality if (ζ_k, θ_k) is Gaussian pair.

Integrating by parts we find $\int_{\mathbb{R}^{kn}} u^T \nabla_u q(u | v) du = -\mathbf{I}$, where \mathbf{I} is the unit matrix of relevant size, so that

$$\begin{aligned} -\mathbf{I} &= \int_{\mathbb{R}^{k\ell}} \int_{\mathbb{R}^{kn}} (u - \widehat{\zeta}_k(v))^T \nabla_u q(u | v) g(v) du dv \\ &= \int_{\mathbb{R}^{k\ell}} \int_{\mathbb{R}^{kn}} (u - \widehat{\zeta}_k(v))^T \nabla_u q(u | v) \frac{g(v)}{p(u, v)} p(u, v) du dv. \end{aligned} \quad (2.12)$$

Since $\frac{g^2(v)}{p^2(u, v)} p(u, v) = \frac{p(u, v)}{q^2(u | v)} = \frac{g(v)}{q(u | v)}$, we have

$$I_k = \int_{\mathbb{R}^{k\ell}} \int_{\mathbb{R}^{kn}} \nabla_u^T q(u | v) \nabla_u q(u | v) \frac{g^2(v)}{p^2(u, v)} p(u, v) du dv.$$

Now, the matrix Cauchy-Schwarz inequality, applied to the right side of (2.12), gives

$$\begin{aligned} & \left(\int_{\mathbb{R}^{k\ell}} \int_{\mathbb{R}^{kn}} (u - \widehat{\zeta}_k(v))^T \frac{\nabla_u q(u|v)}{q(u|v)} q(u|v) g(v) dudv \right)^T \\ & \times I_k^+ \int_{\mathbb{R}^{k\ell}} \int_{\mathbb{R}^{kn}} (u - \widehat{\zeta}_k(v))^T \frac{\nabla_u q(u|v)}{q(u|v)} q(u|v) g(v) dudv \\ & \leq \int_{\mathbb{R}^{k\ell}} \int_{\mathbb{R}^{kn}} (u - \widehat{\zeta}_k(v))(u - \widehat{\zeta}_k(v))^T p(u, v) dudv. \end{aligned}$$

Thus, (2.11) is implied by (2.12).

We now use another representation for I_k (comparable to (2.10)). To this end, introduce the operator ∇_u^2 , that is, for every smooth scalar function $h(u)$ with vector argument $u \in \mathbb{R}^{kn}$,

$$\nabla_u^2 h(u) = \begin{pmatrix} \text{matrix} \\ \text{with elements} \\ \frac{\partial^2 h(u)}{\partial u_i \partial u_j} \end{pmatrix}.$$

The direct verification shows $\int_{\mathbb{R}^{k\ell}} \int_{\mathbb{R}^{kn}} p(u, v) \nabla_u^2 \log p(u, v) dudv = -I_k$ that, taking into account $p(u, v) = r(v|u)f(u)$ (see (2.9)), leads

$$\begin{aligned} I_k &= - \int_{\mathbb{R}^{kn}} f(u) \left(\int_{\mathbb{R}^{k\ell}} r(v|u) \nabla_u^2 \log r(v|u) dv \right) du \\ &\quad - \int_{\mathbb{R}^{kn}} f(u) \nabla_u^2 \log f(u) du \\ &= \int_{\mathbb{R}^{kn}} f(u) \left(\int_{\mathbb{R}^{k\ell}} \frac{\nabla_u^T f(u|u) \nabla_u r(v|u)}{r(v|u)} dv \right) du \\ &\quad + \int_{\mathbb{R}^{kn}} \frac{\nabla_u^T f(u) \nabla_u f(u)}{f(u)} du. \end{aligned} \tag{2.13}$$

Since $r(v|u)$ is the conditional distribution density of θ_k given $\zeta_k(=u)$, the matrix

$$\int_{\mathbb{R}^{kn}} f(u) \left(\int_{\mathbb{R}^{k\ell}} \frac{\nabla_u^T r(v|u) \nabla_u f(v|u)}{r(v|u)} dv \right) du$$

can be explicitly expressed in terms of matrices A and \mathcal{T} , or more exactly, as a block

diagonal matrix with blocks $A^T \mathcal{T} A$. The detailed computation is omitted here whereas the rest of the proof does not rely on a specific structure of this matrix.

Consistent with (ζ_k, θ_k) , consider an auxiliary Gaussian pair $(\tilde{\zeta}_k, \tilde{\theta}_k)$, where $\tilde{\theta}_k = (\tilde{y}_k, \dots, \tilde{y}_1)$ and $\tilde{\zeta}_k = (\tilde{x}_k, \dots, \tilde{x}_1)$ with

$$\begin{aligned} \tilde{x}_j &= e^{a\Delta} \tilde{x}_{j-1} + \int_{t_{j-1}}^{t_j} e^{a(t_j-s)} b dW_s \\ \tilde{y}_j &= A \tilde{x}_{j-1} \Delta + \mathcal{T}^{-1/2} \tilde{\xi}_j \sqrt{\Delta}, \end{aligned} \quad (2.14)$$

where $(\tilde{\xi}_j)_{j \geq 1}$ is a sequence i.i.d. zero mean Gaussian vectors with i.i.d. elements of unit variance and \tilde{x}_0 is a Gaussian random vector with $\mathbf{E} \tilde{x}_0 = \mathbf{E} X_0$ and covariance matrix \mathcal{T}_0^+ . Denoted by $\tilde{r}(v|u)$ the conditional density of $\tilde{\theta}_k$ given $\tilde{\zeta}_k (= u)$ and by $\tilde{f}(u)$ the distribution density of $\tilde{\zeta}_k$. Using the fact that under a Gaussian distribution, the covariance matrix coincides with the pseudoinverse of the Fisher information, we get

$$\begin{aligned} & \int_{\mathbb{R}^{kn}} f(u) \left(\int_{\mathbb{R}^{k\ell}} \frac{\nabla_u^T f(v|u) \nabla_u f(v|u)}{r(v|u)} dv \right) du \\ &= \int_{\mathbb{R}^{kn}} \tilde{f}(u) \left(\int_{\mathbb{R}^{k\ell}} \frac{\nabla_u^T \tilde{r}(v|u) \nabla_u \tilde{r}(v|u)}{\tilde{r}(v|u)} dv \right) du \end{aligned}$$

and hence, with $\check{\zeta}_k = \mathbf{E}(\tilde{\zeta}_k | \tilde{\theta}_k)$, we have $\mathbf{E}(\zeta_k - \check{\zeta})(\zeta_k - \check{\zeta}_k)^T = I_k^+$. Thus, the above equality and (2.11) imply

$$\mathbf{E}(\zeta_k - \hat{\zeta}_k)(\zeta_k - \hat{\zeta}_k)^T \geq \mathbf{E}(\zeta_k - \check{\zeta}_k)(\zeta_k - \check{\zeta}_k)^T. \quad (2.15)$$

It is clear that the matrix $V_{t^k}^\Delta$ is the sub-block of $\mathbf{E}(\zeta_k - \hat{\zeta}_k)(\zeta_k - \hat{\zeta}_k)^T$ located on the left upper corner. Denote the same sub-block of $\mathbf{E}(\zeta_k - \check{\zeta}_k)(\zeta_k - \check{\zeta}_k)^T$ by $P_{t^k}^\Delta$. Recall that the Gaussian pair $(\check{\zeta}_k, \tilde{\theta}_k)$ is generated by the recursion (2.14), so that the matrix $P_{t^k}^\Delta$ gives the mean square filtering error and is defined by the recursion given in (2.8) as a part of the Kalman filter corresponding to model (2.14).

The property given in (2.15) is nothing but the statement of the lemma for $k \geq 1$ (for $k = 0$ the proof uses the same type arguments and is omitted). \square

2.3 An Asymptotic Optimal Property Over a Finite Horizon

X_0 is Gaussian vector, so $P_0 = \mathcal{T}_0^+$. By virtue of (2.5) we have

$$(X_t - \tilde{\pi}_t(Y^\Delta))(X_t - \tilde{\pi}_t(Y^\Delta))^T \stackrel{\text{law}}{\underset{\Delta \rightarrow 0}{\rightarrow}} (X_t - \tilde{\pi}_t(\tilde{Y}))(X_t - \tilde{\pi}_t(\tilde{Y}))^T. \quad (2.16)$$

Moreover, under (2.4) and (FR), for every $t > 0$ the trace of the matrix on the left side of (2.16) is uniformly integrable for $\Delta > 0$. Therefore

$$\lim_{\Delta \rightarrow 0} \mathbf{E}(X_t - \tilde{\pi}_t(Y^\Delta))(X_t - \tilde{\pi}_t(Y^\Delta))^T = \mathbf{E}(X_t - \tilde{\pi}_t(\tilde{Y}))(X_t - \tilde{\pi}_t(\tilde{Y}))^T. \quad (2.17)$$

We now show that with P_t the solution of (2.3)

$$\mathbf{E}(X_t - \tilde{\pi}_t(\tilde{Y}))(X_t - \tilde{\pi}_t(\tilde{Y}))^\top = P_t, \quad \forall t > 0. \tag{2.18}$$

To this end, introduce $\hat{X}_t = \mathbf{E}(X_t | \tilde{Y}_s, s \leq t)$ and $\Gamma_t = \mathbf{E}(X_t - \hat{X}_t)(X_t - \hat{X}_t)^\top$. Taking into accounting (2.6) and applying the conditionally Gaussian filter (see Ch. 11 in [13]) we find $\hat{X}_0 = \mathbf{E}X_0$ and $\Gamma_0 = P_0$ and

$$\begin{aligned} d\hat{X}_t &= a\hat{X}_t dt + \Gamma_t A^\top \mathcal{T}(d\tilde{Y}_t - A(\hat{X}_t - \tilde{\pi}_t(\tilde{Y}))dt) \\ \dot{\Gamma}_t &= a\Gamma_t + \Gamma_t a^\top + bb^\top - \Gamma_t A^\top \mathcal{T} A \Gamma_t. \end{aligned}$$

Hence, for $\delta_t = \hat{X}_t - \tilde{\pi}_t(\tilde{Y})$, we have $\delta_0 = 0$ and $d\delta_t = (a - \Gamma_t A^\top \mathcal{T} A)dt$. That is $\delta_t \equiv 0$. Since $P_t \equiv \Gamma_t$, (2.18) holds.

Let us show also that

$$\lim_{\Delta \rightarrow 0} \mathbf{E}(X_t - \pi_t(Y^\Delta))(X_t - \pi_t(Y^\Delta))^\top \equiv P_t. \tag{2.19}$$

Since $\pi_t(Y^\Delta)$ is the optimal filtering estimate, we have the inequality

$$\mathbf{E}(X_t - \pi_t(Y^\Delta))(X_t - \pi_t(Y^\Delta))^\top \leq \mathbf{E}(X_t - \tilde{\pi}_t(Y^\Delta))(X_t - \tilde{\pi}_t(Y^\Delta))^\top$$

and thus, by virtue of (2.17) and (2.18), to verify (2.19), it remains to show only that for every $t \geq 0$

$$\lim_{\Delta \rightarrow 0} \mathbf{E}(X_t - \pi_t(Y^\Delta))(X_t - \pi_t(Y^\Delta))^\top \geq P_t. \tag{2.20}$$

For fixed $t > 0$, let k be an integer such that $t_k \leq t < t_{k+1}$. Then for such t ,

$$\begin{aligned} X_t &= e^{a(t-t_k)} X_{t_k} + \int_{t_k}^t e^{a(s-s)} b dW_s \\ \pi_t(Y^\Delta) &= e^{a(t-t_k)} \pi_{t_k}(Y^\Delta) \end{aligned}$$

and, therefore,

$$X_t - \pi_t(Y^\Delta) = e^{a(t-t_k)} (X_{t_k} - \pi_{t_k}(Y^\Delta)) + \int_{t_k}^t e^{a(t-s)} b dW_s. \tag{2.21}$$

Consequently,

$$\begin{aligned} &\mathbf{E}(X_t - \pi_t(Y^\Delta))(X_t - \pi_t(Y^\Delta))^\top \\ &= e^{a(t-t_k)} V_{t_k}^\Delta e^{a^\top(t-t_k)} + \int_{t_k}^t e^{a(t-s)} b b^\top e^{a^\top(t-s)} ds. \end{aligned}$$

By Lemma 2.1, we have $V_{t_k}^\Delta \geq P_{t_k}^\Delta$ and with $P_t^\Delta \equiv P_{t_k}^\Delta$ for $t_k \leq t < t_{k+1}$, the inequality

$$\begin{aligned} &\mathbf{E}(X_t - \pi_t(Y^\Delta))(X_t - \pi_t(Y^\Delta))^\top \\ &\geq e^{a(t-t_k)} P_{t_k}^\Delta e^{a^\top(t-t_k)} + \int_{t_k}^t e^{a(t-s)} b b^\top e^{a^\top(t-s)} ds \end{aligned} \tag{2.22}$$

holds. Now, the proof of (2.20) is reduced to verifying

$$\lim_{\Delta \rightarrow 0} P_t^\Delta = P_t. \quad (2.23)$$

Proof of (2.23): The matrix $P_{t_k}^\Delta$, the filtering error of X_{t_k} , is upper bounded by the matrix $Q_{t_k} = \mathbf{cov}(X_{t_k}, X_{t_k})$ with $P_{t_k}^\Delta \leq Q_{t_k}$. That is, for every fixed $t > 0$, $\sup_{t' \leq t} \|P_{t'}^\Delta\|$ is bounded by a constant independent of Δ ; in particular, $\sup_{t' \leq t} \|P_{t'}^\Delta - P_{t'}^\Delta\|$ inherits the same property. Setting $U_t^\Delta = P_t - P_t^\Delta$, we show that for every $t > 0$

$$\lim_{\Delta \rightarrow 0} \|U_t^\Delta\| = 0. \quad (2.24)$$

From (2.3), it follows

$$\begin{aligned} P_{t_k} &= e^{a\Delta} P_{t_{k-1}} e^{a^\top \Delta} + \int_{t_{k-1}}^{t_k} e^{a(t_k-s)} b b^\top e^{a^\top(t_k-s)} ds \\ &= - \int_{t_{k-1}}^{t_k} e^{a(t_k-s)} P_s A \mathcal{T} A^\top P_s e^{a^\top(t_k-s)} ds \\ &= e^{a\Delta} P_{t_{k-1}} e^{a^\top \Delta} + \int_0^\Delta e^{as} b b^\top e^{a^\top s} ds \\ &\quad - e^{a\Delta} P_{t_{k-1}} A^\top \mathcal{T} A P_{t_{k-1}} e^{a^\top \Delta} + r'_{t_{k-1}}(\Delta), \end{aligned} \quad (2.25)$$

where $r'_{t_{k-1}}(\Delta)$ is the residual matrix with $\|r'_{t_{k-1}}(\Delta)\| \propto o(\Delta)$. On the other hand, taking into account

$$\begin{aligned} &- e^{a\Delta} P_{t_{k-1}}^\Delta A^\top (\mathcal{T}^{-1} + A P_{t_{k-1}}^\Delta A^\top \Delta)^{-1} A P_{t_{k-1}}^\Delta e^{a^\top \Delta} \Delta \\ &= - e^{a\Delta} P_{t_{k-1}}^\Delta A^\top \mathcal{T} A P_{t_{k-1}}^\Delta e^{a^\top \Delta} \Delta + r''_{t_{k-1}}(\Delta) \end{aligned}$$

with residual matrix $r''_{t_{k-1}}(\Delta)$ with $\|r''_{t_{k-1}}(\Delta)\| \propto o(\Delta)$, by virtue of (2.8), we have

$$\begin{aligned} P_{t_k}^\Delta &= e^{a\Delta} P_{t_{k-1}}^\Delta e^{a^\top \Delta} + \int_0^\Delta e^{as} b b^\top e^{a^\top s} ds \\ &\quad - e^{a\Delta} P_{t_{k-1}}^\Delta A^\top (\mathcal{T}^{-1} + A P_{t_{k-1}}^\Delta A^\top \Delta)^{-1} A P_{t_{k-1}}^\Delta e^{a^\top \Delta} \Delta \\ &= e^{a\Delta} P_{t_{k-1}}^\Delta e^{a^\top \Delta} + \int_0^\Delta e^{as} b b^\top e^{a^\top s} ds \\ &\quad - e^{a\Delta} P_{t_{k-1}}^\Delta A^\top \mathcal{T} A P_{t_{k-1}}^\Delta e^{a^\top \Delta} \Delta + r'''_{t_{k-1}}(\Delta), \end{aligned} \quad (2.26)$$

where $\|r'''_{t_{k-1}}(\Delta)\| \propto o(\Delta)$ as well. Therefore (2.25) and (2.26) imply

$$U_{t_k}^\Delta = e^{a\Delta} U_{t_{k-1}}^\Delta e^{a^\top \Delta} - e^{a\Delta} U_{t_{k-1}}^\Delta A^\top \mathcal{T} A P_{t_{k-1}}^\Delta e^{a^\top \Delta} \Delta$$

$$-e^{a\Delta} P_{t_k-1}^\Delta A^T \mathcal{T} A U_{t_k-1}^\Delta e^{a^T \Delta} \Delta + r_{t_k-1}(\Delta) \tag{2.27}$$

with $\|r_{t_k-1}(\Delta)\| \propto o(\Delta)$ (for fixed $t > 0$ and $t_k \leq t$, $o(\Delta)$ depends on t only).

Hence, there is a positive constant $C > 0$ such that for every $t_k \leq t$ we have

$$\|U_{t_k}^\Delta\| \leq \|U_{t_k-1}^\Delta\| (1 + C\Delta) + o(\Delta).$$

This recursion and $\|U_0^\Delta\| = 0$ imply

$$\|U_t\| \leq o(\Delta) \sum_{j=1}^{\lceil t/\Delta \rceil} (1 + C\Delta)^{j-1} \leq te^{tC\frac{o(\Delta)}{\Delta}} \rightarrow 0, \Delta \rightarrow 0.$$

Thus (2.24) holds. □

3. The Infinite Horizon Case

3.1 Formulation of the Main Result

An essential role in establishing the asymptotic optimal property for the filtering estimate $\tilde{\pi}_t(Y^\Delta)$ over the infinite time interval is played by the relation

$$\lim_{\Delta \rightarrow 0} \lim_{t \rightarrow \infty} P_t^\Delta = \lim_{t \rightarrow \infty} P_t = P \tag{3.1}$$

with positive definite matrix P which is the unique solution (in the class of nonnegative definite matrices) of the algebraic Riccati equation

$$aP + Pa^T + bb^T - PA^T \mathcal{T} AP = 0. \tag{3.2}$$

To clarify a role of the matrix P , we mention here that the optimal filtering estimate $\pi_t(Y^\Delta)$ obeys a lower bound (Lemma 3.2)

$$\lim_{\Delta \rightarrow 0} \lim_{t \rightarrow \infty} \mathbf{E}(X_t - \pi_t(Y^\Delta))(X_t - \pi_t(Y^\Delta))^T \geq P \tag{3.3}$$

while the filtering estimate $\tilde{\pi}_t(Y^\Delta)$ given in (2.2) obeys the asymptotically optimal property

$$\lim_{\Delta \rightarrow 0} \lim_{t \rightarrow \infty} \mathbf{E}(X_t - \tilde{\pi}_t(Y^\Delta))(X_t - \tilde{\pi}_t(Y^\Delta))^T = P. \tag{3.4}$$

It is known (see [13], Theorem 16.2) that at least the second part of (3.1) is provided by (FR) (see Lemma 2.1) and

$$\begin{pmatrix} A \\ Aa \\ \vdots \\ Aa^{n-1} \end{pmatrix}_{(\ell \cdot n) \times n} \text{ is the matrix of full rank.} \tag{FR'}$$

In this setting, for the verification of the first part of (3.1), and especially (3.4), another (FR') condition is required. Even if (FR') holds, we assume that eigenvalues of the matrix a have negative real parts (see (INF.3) in Theorem 3.1). It can be proved that under (FR) and (INF.3) the second equality in (3.1) holds with the posi-

tive definite matrix P .

It is natural when filtering over a long time interval to replace P_t in the equation for $\tilde{\pi}_t(Y^\Delta)$ (see (2.2)), by its limit P , that is, to use the modified version of (2.2)

$$d\tilde{\pi}_t(Y^\Delta) = a\tilde{\pi}_t(Y^\Delta)dt + PA^\top \mathcal{T} d\tilde{Y}_t^\Delta. \quad (3.5)$$

Moreover, the exact knowledge of the initial condition $\tilde{\pi}_0(Y^\Delta)$ is not essential, since it is forgotten by the filter dynamics. It is well known from Makowski [15], Makowski and Sowers [16], Ocone and Pardoux [17] (see also Budhiraja and Kushner [6]) that, under (FR) and (FR'), the Kalman filter is asymptotically optimal (among nonlinear filters) over the infinite time interval even if the first and second moments of an inappropriate distribution for X_0 are used as the initial conditions for the filter. We establish the same property for $\tilde{\pi}_t(Y^\Delta)$ defined in (3.5).

Theorem 3.1: *Assume (FR) and*

(INF.1) G_i is twice continuous differentiable; G_i, G_i', G_i'' 's are bounded.

(INF.2) X_0 is an arbitrary distributed vector with $E\|X_0\|^2 < \infty$.

(INF.3) Eigenvalues of the matrix a have negative real parts.

Then (3.3) is valid and (3.4) holds for $\tilde{\pi}_t(Y^\Delta)$ defined in (3.5) under any fixed initial condition.

3.2 Auxiliary Lemmas

It can be shown that not only under (FR) and (FR') but also under (FR) and (INF.3), $\lim_{t \rightarrow \infty} P_t = P$ exists. Let us denote $P_k^\Delta = P_{t_k}^\Delta$. Then (2.8) may be rewritten as

$$P_k^\Delta = e^{a\Delta} P_{k-1}^\Delta e^{a^\top \Delta} + \int_0^\Delta e^{as} bb^\top e^{a^\top s} ds \\ - e^{a\Delta} P_{k-1}^\Delta A^\top (\mathcal{T}^{-1} + AP_{k-1}^\Delta A^\top \Delta)^{-1} AP_{k-1}^\Delta e^{a^\top \Delta} \Delta.$$

Taking into account the eigenvalues of the matrix $e^{a\Delta}$ have absolute values strictly less than 1 and the matrix $\int_0^\Delta e^{as} bb^\top e^{a^\top s} ds$ is positive definite, one can modify the proof of Theorem 14.3 from [13] to show that under (FR) and (INF.3)

$$\lim_{k \rightarrow \infty} P_k^\Delta = P^\Delta \quad (3.6)$$

with

$$P^\Delta = e^{a\Delta} P^\Delta e^{a^\top \Delta} + \int_0^\Delta e^{as} bb^\top e^{a^\top s} ds \\ - e^{a\Delta} P^\Delta A^\top (\mathcal{T}^{-1} + AP^\Delta A^\top \Delta)^{-1} AP^\Delta e^{a^\top \Delta} \Delta. \quad (3.7)$$

Lemma 3.1: *Under (FR), (INF.3), $\lim_{\Delta \rightarrow 0} P^\Delta = P$.*

Proof: By defining the following matrices,

$$r'(\Delta) = aP^\Delta + P^\Delta a^\top - \frac{1}{\Delta} (e^{a\Delta} P^\Delta e^{a^\top \Delta} - P^\Delta)$$

$$r''(\Delta) = \int_0^\Delta \frac{1}{\Delta} (bb^\top - e^{as} bb^\top e^{a^\top s}) ds$$

$$r'''(\Delta) = e^{a\Delta} P^\Delta A (\mathcal{T}^{-1} + A P^\Delta A^\top \Delta)^{-1} A^\top P^\Delta e^{a^\top \Delta} - P^\Delta A \mathcal{T} A^\top P^\Delta$$

$$r(\Delta) = r'(\Delta) + r''(\Delta) + r'''(\Delta),$$

(3.7) can be rewritten as

$$aP^\Delta + P^\Delta a^\top + bb^\top - P^\Delta A \mathcal{T} A^\top P^\Delta = r(\Delta). \quad (3.8)$$

Under assumptions made previously, the norm of the matrix P^Δ is bounded from above by a positive constant independent of Δ . Hence $\|r(\Delta)\| \propto O(\Delta)$. Therefore,

$$\lim_{\Delta \rightarrow 0} (aP^\Delta + P^\Delta a^\top + bb^\top - P^\Delta A \mathcal{T} A^\top P^\Delta) = 0.$$

Since matrices P^Δ , $\Delta > 0$ are bounded, any infinite sequence $(\Delta_i)_{i \geq 1}$ decreasing to zero contains subsequences $(\Delta_{i'})_{i' \geq 1}$ such that the limit $\lim_{i' \rightarrow \infty} P^{\Delta_{i'}} = P'$ exists. Then P' is a solution of (3.2) and thus $P' = P$. (Recall that (3.2) has the unique solution in the class of nonnegative definite matrices.) This fact allows us to conclude the existence of the limit $\lim_{\Delta \rightarrow 0} P^\Delta = P$. \square

Lemma 3.2: Under (FR) and (INF.3), (3.3) holds.

Proof: By virtue of (2.21), we have

$$\begin{aligned} & \lim_{\Delta \rightarrow 0} \lim_{t \rightarrow \infty} \mathbf{E}(X_t - \pi_t(Y^\Delta))(X_t - \pi_t(Y^\Delta))^\top \\ &= \lim_{\Delta \rightarrow 0} \lim_{k \rightarrow \infty} \mathbf{E}(X_{t_k} - \pi_{t_k}(Y^\Delta))(X_{t_k} - \pi_{t_k}(Y^\Delta))^\top, \end{aligned}$$

while by Lemma 2.1, $\mathbf{E}(X_{t_k} - \pi_{t_k}(Y^\Delta))(X_{t_k} - \pi_{t_k}(Y^\Delta))^\top \geq P_{t_k}^\Delta$.

Hence the desired result is implied by (3.6) and Lemma 3.1. \square

Lemma 3.3: Under (FR) and (INF.3), eigenvalues of the matrix $a - PA^\top \mathcal{T} A$ have negative real parts.

Proof: Set $K = a - PA^\top \mathcal{T} A$ and rewrite (3.2) as

$$KP + PK^\top + bb^\top + PA^\top \mathcal{T} AP = 0. \quad (3.9)$$

Let $\varphi(\varphi^\top)$ be the right eigenvector of K^\top (the left eigenvector of K) with eigenvalue λ . ($Re(\lambda)$ is the real part of λ .) Multiplying (3.9) from the right by φ and from the left by φ^\top , we obtain

$$2Re(\lambda)\varphi^\top P\varphi + \varphi^\top(bb^\top + PA^\top \mathcal{T} AP)\varphi = 0.$$

Because P is positive definite and $bb^\top + PA^\top \mathcal{T} AP$ is nonnegative definite, then $Re(\lambda) \leq 0$. Assume $Re(\lambda) = 0$. Then $\varphi^\top PA^\top \mathcal{T} AP\varphi = 0$, so $\varphi^\top PA^\top \mathcal{T}^{1/2} = 0$ and, in turn, $\varphi^\top PA^\top \mathcal{T} = 0$ and $\varphi^\top PA^\top \mathcal{T} A = 0$ (or $A^\top \mathcal{T} AP\varphi = 0$). Then

$$\lambda\varphi = K^\top \varphi = (a^\top - A^\top \mathcal{T} AP)\varphi = a^\top \varphi,$$

i.e., φ is the right eigenvector of the matrix a^\top so the eigenvalue of a^\top has a zero real part. The latter contradicts the assumption $Re(\lambda) = 0$. Consequently, $Re(\lambda) < 0$. \square

3.3 Proof of Theorem 3.1

The proof is divided into two parts.

3.3.1: X_0 is a Gaussian vector with known expectation and covariance.

Set $u_t^\Delta = X_t - \tilde{\pi}_t(Y^\Delta)$. It suffices to verify

$$\lim_{\Delta \rightarrow 0} \lim_{k \rightarrow \infty} \mathbf{E} u_{t_k}^\Delta (u_{t_k}^\Delta)^\top = P. \quad (3.10)$$

From (1.1) and (2.2) with P_s replaced by P , it follows

$$u_t^\Delta = u_0^\Delta + \int_0^t a u_s^\Delta ds + \int_0^t b dW_s - \int_0^t P A^\top \mathcal{T} d\tilde{Y}_s^\Delta$$

and, in turn, for every t_k we find

$$u_{t_k}^\Delta = e^{a\Delta} u_{t_{k-1}}^\Delta + \int_{t_{k-1}}^{t_k} e^{a(t_k-s)} b dW_s - P A^\top \mathcal{T} (\tilde{Y}_{t_k}^\Delta - \tilde{Y}_{t_{k-1}}^\Delta). \quad (3.11)$$

Due to (2.1)

$$\begin{aligned} (\tilde{Y}_{t_k}^\Delta - \tilde{Y}_{t_{k-1}}^\Delta) &= \sqrt{\Delta} \mathcal{T}^{-1} G \left(\frac{Y_{t_k}^\Delta - Y_{t_{k-1}}^\Delta - A \tilde{\pi}_{t_{k-1}}(Y^\Delta) \Delta}{\sqrt{\Delta}} \right) \\ &= \sqrt{\Delta} \mathcal{T}^{-1} G(\xi_k + A u_{t_{k-1}}^\Delta \sqrt{\Delta}) \\ &= \sqrt{\Delta} \mathcal{T}^{-1} \begin{pmatrix} G_1((\xi_k)^{(1)} + (A u_{t_{k-1}}^\Delta)^{(1)} \sqrt{\Delta}) \\ \vdots \\ G_\ell((\xi_k)^{(\ell)} + (A u_{t_{k-1}}^\Delta)^{(\ell)} \sqrt{\Delta}) \end{pmatrix}. \end{aligned}$$

In addition, since G_i is twice continuously differentiable, we have

$$\begin{aligned} G_i((\xi_k)^{(i)} + (A u_{t_{k-1}}^\Delta)^{(i)} \sqrt{\Delta}) &= G_i((\xi_k)^{(i)}) + G_i'((\xi_k)^{(i)}) \cdot (A u_{t_{k-1}}^\Delta)^{(i)} \sqrt{\Delta} \\ &\quad + \int_0^{(A u_{t_{k-1}}^\Delta)^{(i)} \sqrt{\Delta}} \int_0^z G_i''((\xi_k)^{(i)} + z') dz' dz. \end{aligned}$$

For convenience, denote by

$$r'_k(\Delta) = \sqrt{\Delta} P A^\top \mathcal{T}^{-1} \begin{pmatrix} \int_0^{(A u_{t_{k-1}}^\Delta)^{(1)} \sqrt{\Delta}} \int_0^z G_1''((\xi_k)^{(1)} + z') dz' dz \\ \vdots \\ \int_0^{(A u_{t_{k-1}}^\Delta)^{(\ell)} \sqrt{\Delta}} \int_0^z G_\ell''((\xi_k)^{(\ell)} + z') dz' dz \end{pmatrix},$$

$$G(\xi_k) = \begin{pmatrix} G_1((\xi_k)^{(1)}) \\ \vdots \\ G_\ell((\xi_k)^{(\ell)}) \end{pmatrix}$$

$$G'(\xi_k) = \text{diag}(G'_1(\xi_k)^{(1)} \quad G'_2(\xi_k)^{(2)} \quad \dots \quad G'_\ell(\xi_k)^{(\ell)}) \quad (3.12)$$

where $\text{diag}(\ast)$ designates a scalar matrix with \ast as the diagonal. Then (3.11) may be rewritten as

$$\begin{aligned} u_{t_k}^\Delta &= (e^{a\Delta} - PA^\top G'(\xi_k)A\Delta)u_{t_{k-1}}^\Delta \\ &+ \int_{t_{k-1}}^{t_k} e^{a(t_k-s)} b dW_s - PA^\top G'(\xi_k)\sqrt{\Delta} - r'_k(\Delta) \end{aligned} \quad (3.13)$$

where $r'_k(\Delta)$ is the residual matrix with norm

$$\|r'_k(\Delta)\| \propto O(\Delta^{3/2}). \quad (3.14)$$

Set $\Upsilon_{t_k}^\Delta = \mathbf{E}u_{t_k}^\Delta(u_{t_k}^\Delta)^\top$ and note that (3.13) implies

$$\begin{aligned} \Upsilon_{t_k}^\Delta &= \mathbf{E}\left\{ (e^{a\Delta} - PA^\top G'(\xi_1)A\Delta)\Upsilon_{t_{k-1}}^\Delta (e^{a\Delta} - PA^\top G'(\xi_1)A\Delta)^\top \right\} \\ &+ \int_0^\Delta e^{as} b b^\top e^{a^\top s} ds + PA^\top \mathbf{E}\{G(\xi_1)G^\top(\xi_1)\}AP\Delta \\ &- \mathbf{E}\{u_{t_k}^\Delta (r'_k(\Delta))^\top\} - \mathbf{E}\{r'_k(\Delta)(u_{t_k}^\Delta)^\top\}. \end{aligned} \quad (3.15)$$

A few additional computations are now required. Write

$$\begin{aligned} &\mathbf{E}\left\{ (e^{a\Delta} - PA^\top G'(\xi_1)A\Delta)\Upsilon_{t_{k-1}}^\Delta (e^{a\Delta} - PA^\top G'(\xi_1)A\Delta)^\top \right\} \\ &= e^{a\Delta}\Upsilon_{t_{k-1}}^\Delta e^{a^\top\Delta} - (e^{a\Delta}\Upsilon_{t_{k-1}}^\Delta A^\top \mathbf{E}G'(\xi_1)AP + PA^\top \mathbf{E}G'(\xi_1)A\Upsilon_{t_{k-1}}^\Delta e^{a^\top\Delta})\Delta + r''_k(\Delta), \end{aligned}$$

where $r''_k(\Delta) = \mathbf{E}\{PA^\top G'(\xi_1)A\Upsilon_{t_{k-1}}^\Delta (P_{t_k} A^\top G'(\xi_1)A)^\top\}\Delta^2$. Assumptions (INF.1) and (INF.3) imply that $\Upsilon_{t_k}^\Delta$'s are bounded. Therefore,

$$\|r''_k(\Delta)\| \propto \Delta^2. \quad (3.16)$$

Note that $\mathbf{E}G'_i(\xi_1^{(i)}) = -\int_{\mathbb{R}} \left(\frac{p'_i(x)}{p_i(x)}\right)' dx = \int_{\mathbb{R}} \frac{(p'_i(x))^2}{p_i(x)} = \mathcal{T}_{ii}$. Therefore, $\mathbf{E}G'(\xi_1) = \mathcal{T}$ and thus

$$\begin{aligned} &\mathbf{E}\left\{ (e^{a\Delta} - PA^\top G'(\xi_1)A\Delta)\Upsilon_{t_{k-1}}^\Delta (e^{a\Delta} - PA^\top G'(\xi_1)A\Delta)^\top \right\} \\ &= e^{a\Delta}\Upsilon_{t_{k-1}}^\Delta e^{a^\top\Delta} - (e^{a\Delta}\Upsilon_{t_{k-1}}^\Delta A^\top \mathcal{T}AP + PA^\top \mathcal{T}A\Upsilon_{t_{k-1}}^\Delta e^{a^\top\Delta})\Delta + r''_k(\Delta). \end{aligned}$$

Since $\mathbf{E}G(\xi_1)G^\top(\xi_1) = \mathcal{T}$, we have $PA^\top \mathbf{E}\{G(\xi_1)G^\top(\xi_1)\}AP\Delta = PA^\top \mathcal{T}AP\Delta$. With the residual matrix $r'''_k(\Delta) = -\mathbf{E}\{u_{t_k}^\Delta (r_k(\Delta))^\top\} - \mathbf{E}\{r_k(\Delta)(u_{t_k}^\Delta)^\top\}$ with norm

$$\|r'''_k(\Delta)\| \propto O(\Delta^{3/2}), \quad (3.17)$$

and, by virtue of (3.16) and by (INF.1) and (INF.3), we find

$$\begin{aligned} \Upsilon_{t_k}^\Delta &= e^{a\Delta} \Upsilon_{t_{k-1}}^\Delta e^{a^\top \Delta} - (e^{a\Delta} \Upsilon_{t_{k-1}}^\Delta A^\top \mathcal{J} A P + P A^\top \mathcal{J} A \Upsilon_{t_{k-1}}^\Delta e^{a^\top \Delta}) \Delta \\ &\quad + \int_0^\Delta e^{as} b b^\top e^{a^\top s} ds + P A^\top \mathcal{J} A P \Delta \\ &\quad + r_k''(\Delta) + r_k'''(\Delta). \end{aligned} \quad (3.18)$$

To simplify notations define the matrix $r_k(\Delta) := r_k''(\Delta) + r_k'''(\Delta)$ with norm

$$\|r_k(\Delta)\| \propto O(\Delta^{3/2}). \quad (3.19)$$

Then, (3.18) may be rewritten in the form

$$\begin{aligned} \Upsilon_{t_k}^\Delta &= e^{a\Delta} \Upsilon_{t_{k-1}}^\Delta e^{a^\top \Delta} - (e^{a\Delta} \Upsilon_{t_{k-1}}^\Delta A^\top \mathcal{J} A P + P A^\top \mathcal{J} A \Upsilon_{t_{k-1}}^\Delta e^{a^\top \Delta}) \Delta \\ &\quad + \int_0^\Delta e^{as} b b^\top e^{a^\top s} ds + P A^\top \mathcal{J} A P \Delta + r_k(\Delta). \end{aligned} \quad (3.20)$$

Further, we use $r_k(\Delta)$ for designating a generic residual matrix with the norm of type (3.19). Taking this into account, the right side of (3.20) can be transformed again so that

$$\begin{aligned} \Upsilon_{t_k}^\Delta &= e^{(a - P A^\top \mathcal{J} A) \Delta} \Upsilon_{t_{k-1}}^\Delta e^{(a - P A^\top \mathcal{J} A)^\top \Delta} \\ &\quad + \int_0^\Delta e^{(a - P A^\top \mathcal{J} A)s} (b b^\top + P A^\top \mathcal{J} A P) e^{(a - P A^\top \mathcal{J} A)^\top s} ds + r_k(\Delta). \end{aligned} \quad (3.21)$$

Now, the equivalent presentation of (3.2) is used which yields

$$(a - P A^\top \mathcal{J} A) P + P (a - P A^\top \mathcal{J} A)^\top + b b^\top + P A^\top \mathcal{J} A P = 0.$$

The solution of the differential equation

$$\dot{\Upsilon}_t = (a - P A^\top \mathcal{J} A) \Upsilon_t + \Upsilon_t (a - P A^\top \mathcal{J} A)^\top + b b^\top + P A^\top \mathcal{J} A P$$

subject to $\Upsilon_0 = P$ is P . In particular, for every k , $\Upsilon_{t_k} = P$. On the other hand, for every k

$$\begin{aligned} \Upsilon_{t_k} &= e^{(a - P A^\top \mathcal{J} A) \Delta} \Upsilon_{t_{k-1}} e^{(a - P A^\top \mathcal{J} A)^\top \Delta} \\ &\quad + \int_0^\Delta e^{(a - P A^\top \mathcal{J} A)s} (b b^\top + P A^\top \mathcal{J} A P) e^{(a - P A^\top \mathcal{J} A)^\top s} ds. \end{aligned}$$

Hence, $U_{t_k}^\Delta = \Upsilon_{t_k}^\Delta - \Upsilon_{t_k}$ is defined as: $U_0^\Delta = P_0 - P$ and

$$U_{t_k}^\Delta = e^{(a - P A^\top \mathcal{J} A) \Delta} U_{t_{k-1}}^\Delta e^{(a - P A^\top \mathcal{J} A)^\top \Delta} + r_k(\Delta).$$

Let integer k' be fixed and $k > k'$. Then

$$\begin{aligned}
 U_k^\Delta &= e^{(a - PA^T \mathcal{J} A) \Delta (k - k')} U_{t_{k'}}^\Delta e^{(a - PA^T \mathcal{J} A)^T \Delta (k - k')} \\
 &+ \sum_{j=k'}^k e^{(a - PA^T \mathcal{J} A) \Delta (k - j - 1)} r_{k'}(\Delta) e^{(a - PA^T \mathcal{J} A)^T \Delta (k - k')}.
 \end{aligned} \tag{3.22}$$

Taking Lemma 3.3 into account, let us denote $-\mu$ ($\mu > 0$) the maximal real part among all real parts of eigenvalues of the matrix $a - PA^T \mathcal{J} A$. Then (recall $a - PA^T \mathcal{J} A$ is an $(n \times n)$ -matrix) there is a positive constant c such that for every integer $q > 0$

$$\| e^{(a - PA^T \mathcal{J} A) \Delta q} \| \leq c(1 + (\Delta q)^{n-1}) e^{-\mu \Delta q}.$$

Therefore, due to U_{k-1}^Δ being bounded, there is a positive constant C such that

$$\begin{aligned}
 \| U_{t_k}^\Delta \| &\leq C \{ e^{-\mu \Delta (k - k')} + \Delta^{3/2} \sum_{j=0}^{k - k' - 1} e^{-\mu \Delta j} \} \\
 &\leq C \{ e^{-\mu \Delta (k - k')} + \frac{\Delta^{3/2}}{1 - e^{-\mu \Delta}} \}.
 \end{aligned}$$

Hence, $\overline{\lim}_{k \rightarrow \infty} \| U_{t_k}^\Delta \| \leq C \frac{\Delta^{3/2}}{1 - e^{-\mu \Delta}} \rightarrow 0, \Delta \rightarrow 0$.

3.3.2: X_0 has an unknown distribution, $\mathbf{E} \| X_0 \|^2 < \infty$.

Consistent with

$$\hat{x}_k = \mathbf{E}(x_k | y_j, 1 \leq j \leq k) \quad (= \pi_{t_k}(Y^\Delta)),$$

let us introduce the filtering estimate

$$\hat{x}_k^\circ = \mathbf{E}(x_k | y_j, 1 \leq j \leq k, X_0).$$

Note that $V_{t_k}^\Delta \geq \mathbf{E}(x_k - \hat{x}_k^\circ)(x_k - \hat{x}_k^\circ)^T, k \geq 1$. On the other hand, since the conditional distribution for x_1 given X_0 is Gaussian with nonsingular covariance, the conditional (given X_0) Fisher information matrix $\mathcal{J}_{1|0}$ of x_1 is well defined and nonsingular. Applying the same arguments used to prove Lemma 2.1, it can be shown that

$$\mathbf{E}(x_k - \hat{x}_k^\circ)(x_k - \hat{x}_k^\circ)^T \geq \bar{P}_{t_k}^\Delta, \quad k \geq 1,$$

where $\bar{P}_{t_k}^\Delta$ is defined by the recursion given in (2.8) subject to $\bar{P}_{t_1}^\Delta = \mathcal{J}_{1|0}^{-1}$. Therefore, we have $V_{t_k}^\Delta \geq \bar{P}_{t_k}^\Delta$. Now, applying the same arguments used in the proof of (3.3), we obtain

$$\lim_{\Delta \rightarrow 0} \lim_{t \rightarrow \infty} \mathbf{E}(X_t - \pi_t(Y^\Delta))(X_t - \pi_t(Y^\Delta))^T \geq \bar{P},$$

where $\bar{P} = \lim_{t \rightarrow \infty} \bar{P}_t$ and \bar{P}_t is a solution of the Riccati equation from (2.2) subject to $\bar{P}_0 = \mathcal{J}_{1|0}^{-1}$. Under the assumption made previously, $\bar{P} = P$. Thus

$$\lim_{\Delta \rightarrow 0} \lim_{t \rightarrow \infty} \mathbf{E}(X_t - \pi_t(Y^\Delta))(X_t - \pi_t(Y^\Delta))^T \geq P \tag{3.23}$$

Analogously, we find

$$\lim_{\Delta \rightarrow 0} \lim_{t \rightarrow \infty} E(X_t - \tilde{\pi}_t(Y^\Delta))(X_t - \tilde{\pi}_t(Y^\Delta))^T = P. \quad \square$$

4. Concluding Remarks

The proof of Theorem 3.1 uses extensively the bounded property of the limiter G and the stability assumption that eigenvalues of the matrix a have negative real parts. Neither can be weakened: with an unbounded limiter and an unstable model, the computer simulation with a reasonable small Δ exhibits catastrophic failure of the signal tracking (see Figure 2).

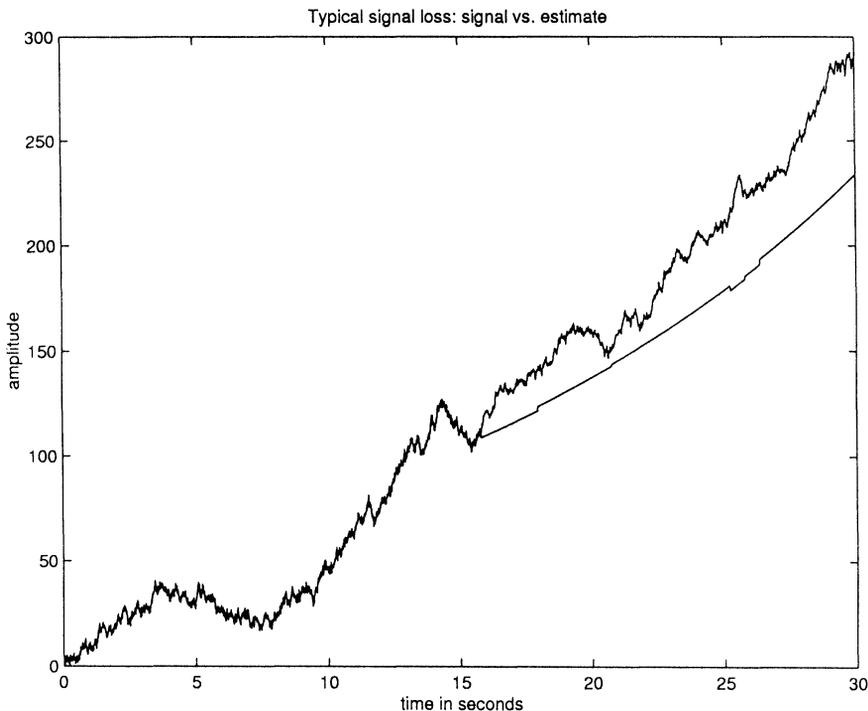


Figure 2. An unstable signal model and the observation signal noise with the following simulation settings: scalar observation noise $\xi_n = \xi_n^\alpha \theta_n + \xi_n^\beta (1 - \theta_n)$, where ξ_n^α and ξ_n^β are independent i.i.d. zero mean Gaussian sequences with $\text{var}(\xi_n^\alpha) = 0.1$ and $\text{var}(\xi_n^\beta) = 10$, and θ_n a binary i.i.d. sequence with $\text{P}\{\theta_n = 1\} = 0.95$. The signal is an unstable scalar diffusion with drift coefficient $a = 0.05$. The estimate was generated for the signal sampled at $\Delta = 10$ msec. The signal loss occurs at ≈ 15 sec.

Similar phenomena occur on long time intervals even for stable signals when the filter uses an unbounded limiter. Tracking failures occasionally occur even if our assumptions are satisfied, i.e., the limiter is bounded and the signal is stable. However, these failures are localized (see Figure 3) and the averaged performance is close to the optimal.

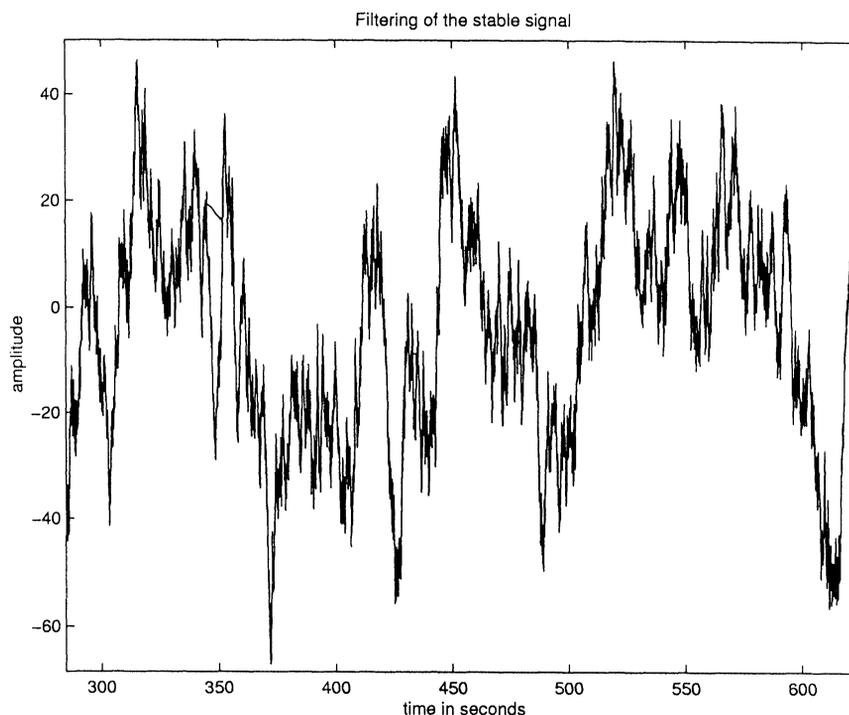


Figure 3. A stable signal model (with drift coefficient $a = -0.01$) and observation noise with Cauchy distribution. In this case the limiter is bounded. With the sampling step Δ as in Figure 2, local tracking failures still occur (as can be seen at $t \approx 350$ sec), but they do not affect drastically the overall performance: empirical mean square error is still close to the theoretical lower bound.

The requirement of bounded limiter may turn out to be too restrictive for certain noise models (e.g., if p_i is a density of Gaussian mixture, then $\frac{p_i}{\bar{p}_i}$ is unbounded). In such cases, it makes sense to approximate G_i by some bounded $\frac{p_i}{\bar{p}_i}$ function G_i^δ , which gives an asymptotically δ -optimal filtering estimate $\tilde{\pi}_i^\delta(Y^{\Delta, \delta})$.

References

- [1] Albert, A., *Regression and Moore-Penrose Pseudoinverse*, Academic Press, New York, London 1972.
- [2] Bobrovsky, B.Z. and Zakai, M., A lower bound on the estimation error for certain diffusion processes, *IEEE Trans. Inform. Theory* **22** (1976), 45-52.
- [3] Bobrovsky, B.Z. and Zakai, M., Asymptotic a priori estimates for the error in the nonlinear filtering problem, *IEEE Trans. Inform. Theory* **2** (1982), 371-376.
- [4] Bobrovsky, B.Z., Zakai, M., and Zeitouni, O., Error bounds for the nonlinear filtering of signals with the small diffusion coefficients, *IEEE Trans. Inform. Theory* **34**:4 (1988), 710-721.

- [5] Bobrovsky, B.Z., Mayer-Wolf, E., and Zeitouni, O., Some classes of global Cramér-Rao bounds, *Ann. Statist.* **4** (1987), 1421-1438.
- [6] Budhiraja, A. and Kushner, H.J., Robustness of nonlinear filters over the infinite time interval, *SIAM J. Control Optim.* **36:5** (1998), 1618-1637.
- [7] Goggin, E.M., Convergence of filters with applications to the Kalman-Bucy case, *IEEE Trans. Inform. Theory* **38** (1992), 1091-1100.
- [8] Kushner, H.J., *Weak Convergence Method and Singular Perturbed Stochastic Control and Filtering Problems*, Birkhauser 1990.
- [9] Kushner, H.J. and Runggaldier, W.J., Filtering and control for wide bandwidth noise driven systems, *IEEE Trans. on Autom. Contr.* **AC-23** (1987), 123-133.
- [10] Liptser, R.S. and Lototsky, S.V., Diffusion approximation and robust Kalman filter, *J. of Math. Syst., Estim. and Contr.* **2:3** (1992), 263-274.
- [11] Liptser, R.S. and Muzhikanov, P., Filtering with a limiter (improved performance), *J. of Appl. Math & Stoch. Anal.* **11:3** (1998), 289-300.
- [12] Liptser, R.S. and Runggaldier, W.J., Diffusion approximations for filtering, *Stoch. Proc. and their Appl.* **38** (1991), 205-238.
- [13] Liptser, R.S. and Shyryaev, A., *Statistics of Random Processes II. Applications*, Applications of Mathematics **6**, Springer-Verlag, New York, Heidelberg, Berlin 1978.
- [14] Liptser, R. and Zeitouni, O., Robust diffusion approximation for nonlinear filtering, *J. of Math. Sys., Est. and Contr.* **8** (1998), 139-142.
- [15] Makowski, A.M., Filtering formulae for partially observed linear systems with non-Gaussian initial conditions, *Stochastics* **16** (1986), 1-24.
- [16] Makowski, A.M. and Sowers, R.B., Discrete-time filtering for linear systems with non-Gaussian initial conditions: asymptotic behaviors of the difference between the MMSE and the LMSE estimates, *IEEE Trans. Autom. Contr.* **37** (1992), 114-121.
- [17] Ocone, D. and Pardoux, E., Asymptotic stability of the optimal filter with respect to its initial condition, *SIAM J. Contr. Optim.* **34:1** (1986), 226-243.