Journal of Mathematical Systems, Estimation, and Control
Vol. 8, No. 1, 1998, pp. 1–22
© 1998 Birkhäuser-Boston

Robust Diffusion Approximation for Nonlinear Filtering*

Robert Liptser† Ofer Zeitouni‡

Abstract

In this paper, we consider the filtering of diffusion processes observed in non-Gaussian noise, when diffusion approximations for the system apply. Standard continuity results show then that the filtering error using the optimal filter for the limit model is close to the error for the limit system. However, this procedure is known to be in general suboptimal. We show that for a certain class of models where the observation is in discrete time and corrupted by i.i.d. (non Gaussian) noise, a pointwise pre-processing is enough to recover optimality. This strengthens some recent results of Goggin. We further exhibit the role of the "signal-to-noise" ratio in the analysis of the performance of the system, and prove monotonicity (in this ratio) of the filtering error. Finally, we provide a filtering lower bound for a class of wide bandwidth observation processes.

1 Introduction

There are only a few stochastic filtering models of Kalman's and Kushner-Zakai's types for which the optimal filtering estimates have a convenient form for computer implementation. A lot of effort has therefore been put into developing approximation techniques for filtering models which are close in some sense to the above mentioned models, the general goal being to construct nearly optimal filters for the original model, based on the solution to the simpler model.

Here, we consider the filtering problem for the nonlinear model in which the unobservable signal X_t is a diffusion process. It can be observed at

*Received December 4, 1996; received in final form March 22, 1997. Summary appeared in Volume 8, Number 1, 1998.
†This work was supported by the ONR Grant N00014-96-1-0413
‡Partially supported by the ONR Grant N00014-91-J-1526 and by a grant from the Israel Science Foundation

times $t_k^n, k = 0, 1 \ldots$ $(t_{k+1}^n - t_k^n \equiv \Delta^n)$ so that the observation process $Y_{t_k^n}$ is defined as:

$$Y_{t_0^n} = 0, \;\; Y_{t_k^n} - Y_{t_{k-1}^n} = h(X_{t_{k-1}^n})\Delta^n + \sqrt{\Delta^n}\xi_k \tag{1.1}$$

where $h(x)$ is some continuous function and $(\xi_k)_{k\geq 1}$ is an i.i.d. sequence of random variables with $E\xi_1 = 0$ and $E\xi_1^2 = B^2$.

The attractiveness of this model is based on the following fact. Independently of the distribution for ξ_1, the sequence of random processes $(Y_t^n)_{t\geq 0}, n \geq 1$, where $Y_t^n = Y_{t_k^n}, t_k^n < t \leq t_{k+1}^n$, converges in the distribution sense, as $\Delta^n \to 0$, to a diffusion type process with respect to a Wiener process W_t independent of the X_t process:

$$Y_t = \int_0^t h(X_s)ds + BW_t \, .$$

This allows us to apply Kushner-Zakai's filtering equation corresponding to the limit model for the prelimit observation. Specifically, for some continuous function $f(x)$, consider the limit pair $(f(X_t), Y_t)$, for which one can compute the functional $\pi_t(y)$ defined on continuous functions $y_s, s \leq t$ such that P-a.s., $\pi_t(Y) = E\Big(f(X_t)\Big|Y_s, s \leq t\Big)$. Assuming $\pi_t(y)$ is defined also for piecewise constant functions and is continuous in some sense, take $\pi_t(Y^n)$ as a filtering estimate for the prelimit pair $(f(X_t), Y_t^n)$. Note that, due to the weak convergence of (X_t, Y_t^n) to (X_t, Y_t), and the continuity of the functional $\pi_t(\cdot)$, the distributions of $f(X_t), \pi_t(Y^n)$ converges to the distribution of the limit $f(X_t), \pi_t(Y)$ (c.f. Section 3 below for a justification of these claims under suitable assumptions). Hence, for any bounded function $f(x)$,

$$\lim_{\Delta^n \to 0} E\Big(f(X_t) - \pi_t(Y^n)\Big)^2 = E\Big(f(X_t) - \pi_t(Y)\Big)^2.$$

Many approximation results of this type with different models of observation noises are well known (see e.g. [6],[7], [9]). However, if the distribution of ξ_1 is not Gaussian, the resulting filter might be far from optimal, even asymptotically, when $\Delta^n \to 0$. On the other hand, using Bayes' formula, one can find the optimal in the mean square filtering estimate $\pi_t^n(Y^n) = E\Big(f(X_t)\Big|Y_s^n, s \leq t\Big)$ for the prelimit model, which may be asymptotically better than $\pi_t(Y^n)$, i.e. it may happen that

$$\limsup_{\Delta^n \to 0} E\Big(f(X_t) - \pi_t^n(Y^n)\Big)^2 < E\Big(f(X_t) - \pi_t(Y)\Big)^2.$$

To remedy this situation, we make a preliminary nonlinear transformation of the observation by some smooth function $G(x)$, hereafter called "limiter", and show that filtering via the diffusion approximation implementation for the transformed signal might be asymptotically better and even optimal.

We mention that one could easily adapt the above procedure and results to situations in which the state process X_t also depends on n. For the sake of simplicity, we chose not to do so here. On the other hand, different observation approximation lead to new difficulties and different challenges, and the results there are less satisfactory. We comment in Section 6 below on some such extensions.

Letting $Y^G_{t^n_0}=0$ and

$$Y^G_{t^n_k}-Y^G_{t^n_{k-1}}=\sqrt{\Delta^n}G\Big(\frac{1}{\sqrt{\Delta^n}}[Y_{t^n_k}-Y_{t^n_{k-1}}]\Big) \tag{1.2}$$

and taking into account that $Y^G_{t^n_k}-Y^G_{t^n_{k-1}}\approx\sqrt{\Delta^n}G(\xi_k)+G'(\xi_k)h(X_{t^n_{k-1}})\Delta^n$, we arrive at another diffusion limit for the sequence of random processes $(Y^{n,G}_t)_{t\ge0}, n\ge1$, with $Y^{n,G}_t=Y^G_{t^n_k}, t^n_k<t\le t^n_{k+1}$:

$$Y^G_t=\int_0^t A_G h(X_s)ds+B_GW_t,$$

where $B^2_G=EG^2(\xi_1)$ and $A_G=EG'(\xi_1)$. This type of diffusion approximation allows us to compare limiters using the parameter $SN_G=A^2_G/B^2_G$ which naturally can be called the "signal-to-noise" ratio. Put $\pi^{G_i}_t(Y^{G_i})=E\Big(f(X_t)\Big|Y^{G_i}_s, s\le t\Big)$ and $\mathcal{E}_{G_i}=E\Big(f(X_t)-\pi^{G_i}_t(Y^{G_i})\Big)^2$, $i=1,2$ where G_1,G_2 are different limiters. We establish the following important implication:

$$SN_{G_1}\le SN_{G_2}\Longrightarrow\mathcal{E}_{G_1}\ge\mathcal{E}_{G_2}\,.$$

Hence, maximization of the "signal-to-noise" ratio by choosing an appropriate limiter is a reasonable goal. To this end, assume that the distribution of ξ_1 possesses a continuously differentiable density $p(x)$ with finite Fisher's information $I_p=\int_{\mathrm{R}}\frac{(p'(x))^2}{p(x)}dx$. Then, for every limiter G

$$SN_G\le I_p$$

and, moreover, the equality is attained at the limiter $G^\circ(x)=-\frac{p'(x)}{p(x)}$, resulting with $A_{G^\circ}=I_p$, $B^2_{G^\circ}=I_p$ (c.f. Lemma 3). Therefore, with I_p finite, the best possible limiter is G° and the asymptotic mean square filtering error (for any bounded f) is defined by

$$\lim_{\Delta^n\to0}E\Big(f(X_t)-\pi^{G^\circ}_t(Y^{n,G^\circ})\Big)^2=E\Big(f(X_t)-\pi^{G^\circ}_t(Y^{G^\circ})\Big)^2. \tag{1.3}$$

On the other hand, slightly modifying the result of Goggin [4] (in the case of signal X_t independent of n), we have

$$\lim_{\Delta^n \to 0} E\Big(f(X_t) - \pi_t^n(Y^n)\Big)^2 = E\Big(f(X_t) - E(f(X_t)|Y_s, Y_s^{G^\circ}, s \le t)\Big)^2, \quad (1.4)$$

where $Y_t, Y_t^{G^\circ}$ is the diffusion limit for the pair Y_t^n, Y_t^{n,G°. Due to the obvious inequality $E\Big(f(X_t) - \pi_t^{G^\circ}(Y^{G^\circ})\Big)^2 \ge E\Big(f(X_t) - E\big(f(X_t)|Y_s, Y_s^{G^\circ}, s \le t\big)\Big)^2$, it is a-priori unclear whether the limiter G° guarantees asymptotically optimal filtering. However, we establish in Theorem 2 that in fact it does.

The structure of this article is as follows: in Section 2, we state the main diffusion approximation and continuity results we use. Section 3 deals with the convergence of the filter obtained by applying a pointwise nonlinear transformation to the discrete observation, and using the optimal filter for the limit model. Section 4 addresses the choice of the optimal nonlinearity for the limit model, as well as some min-max characterization of the optimal nonlinear transformation. Section 5 is devoted to the proof of asymptotic optimality for the pre-limit model. Finally, Section 6 provides a discussion of continuous time extensions as well as a lower bound on the performance for such systems.

2 Assumptions

We assume that the signal X_t is a diffusion process defined by the Itô equation

$$dX_t = a(t, X_t)dt + b(t, X_t)dV_t \quad (2.1)$$

with respect to a Wiener process V_t subject to the initial condition X_0. Denote by $\mathcal{L}$ the Fokker-Planck-Kolmogorov operator, corresponding to (2.1): $\mathcal{L} = a(t,x)\frac{\partial}{\partial x} + \frac{b^2(t,x)}{2}\frac{\partial^2}{\partial x^2}$, and for each $\lambda > 0$ define a nonlinear operator $\mathcal{D}^\lambda$ acting on twice differentiable functions

$$\mathcal{D}^\lambda g = \lambda|\mathcal{L}g| - \frac{1}{2}g^2. \quad (2.2)$$

Hereafter, we fix also the following assumptions.

(A-1) For each $\lambda \in \mathbf{R}$, $Ee^{\lambda h(X_0)} < \infty$.

(A-2) The functions $a(t,x)$ and $b(t,x)$ are continuous and Lipschitz continuous in x uniformly in t; the functions $a(t,0)$ and $b(t,x)$ are bounded.

(A-3) $f(x)$ is a continuous function such that for each $t \geq 0$, $Ef^2(X_t) < \infty$.

(A-4) The function $h(x)$ is twice continuously differentiable, having bounded derivatives $h'(x)$, $h''(x)$, and for each $\lambda > 0$ there exists a constant $C(\lambda)$, depending on λ, such that

$$\mathcal{D}^\lambda h \leq C(\lambda).$$

(A-5) $(\xi_k)_{k\geq 1}$ is an independent of $(X_t)_{t\geq 0}$ sequence of i.i.d. random variables, $E\xi_1^2 < \infty$, $E\xi_1 = 0$.

(A-6) G is a continuously differentiable function such that $E\Big(G'(\xi_1)\Big)^2 < \infty$ and $EG(\xi_1) = 0$.

3 Preliminaries: Diffusion Approximations, Filter Continuity

Introduce the random processes Y_t^n and $Y_t^{n,G}$ letting $Y_t^n = Y_{t_k^n}, Y_t^{n,G} = Y_{t_k^n}^G, t_k^n \leq t < t_{k+1}^n$, where $Y_{t_k^n}$ and $Y_{t_k^n}^G$ are defined in (1.1) and (1.2) respectively. For brevity, the notation $\mathcal{W} - \lim_{n\to\infty}$ is used below for denoting weak convergence in the Skorohod - Lindvall and the local supremum topologies (see e.g. Ch. 6 in [10]).

Theorem 1 *Assume (A-5), (A-6), and that $h(x)$ is continuous. Then*

$$\mathcal{W} - \lim_{n\to\infty}(X_t, Y_t^n, Y_t^{n,G})_{t\geq 0} = (X_t, Y_t, Y_t^G)_{t\geq 0},$$

where $(Y_t, Y_t^G)_{t\geq 0}$ are diffusion processes with respect to independent Wiener processes W_t, W_t', independent of the process X_t, with $Y_0 = 0$, $Y_0^G = 0$, and

$$dY_t = h(X_t)dt + B_1 dW_t' + B_2 dW_t\,, \qquad dY_t^G = A_G h(X_t)dt + B_G dW_t\,.$$

Here,

$$\begin{aligned} A_G &= EG'(\xi_1)\,, B_G = \sqrt{EG^2(\xi_1)} \\ B_1 &= \sqrt{E\xi_1^2 - \frac{(E\xi_1 G(\xi_1))^2}{EG^2(\xi_1)}}\,, B_2 = \frac{E\xi_1 G(\xi_1)}{\sqrt{EG^2(\xi_1)}}. \end{aligned}$$

Proof: Define the increasing function $L_t^n = \Delta^n[t/\Delta^n]$, where $[t]$ is the integer part of t, and the random process $M_t^n = \sqrt{\Delta^n}\sum_{k=1}^{L_t^n/\Delta^n}\xi_k$. Then, the process Y_t^n can be represented as:

$$Y_t^n = \int_0^t h(X_s)dL_s^n + M_t^n. \tag{3.1}$$

Analogously, introduce $M_t^{n,G} = \sqrt{\Delta^n}\sum_{k=1}^{L_t^n/\Delta^n} G(\xi_k)$, and with some $0 \le \theta_k \le 1$,

$$\begin{aligned} u^n(t) &= G'(\xi_k),\ t_{k-1}^n < t \le t_k^n \\ U_t^n &= \Delta^n \sum_{k=1}^{L_t^n/\Delta^n} h(X_{t_{k-1}^n})\Big[G'(\theta_k\Delta^n h(X_{t_{k-1}^n}) + \xi_k) - G'(\xi_k)\Big]. \end{aligned} \tag{3.2}$$

Taking into account the mean value theorem, we arrive at a description for $Y_t^{n,G}$ similar to (3.1): $Y_t^{n,G} = \int_0^t u^n(s)h(X_s)dL_s^n + M_t^n + U_t^n$. Introduce next the process $\widetilde{Y}_t^{n,G}$:

$$\widetilde{Y}_t^{n,G} = \int_0^t A_G h(X_s)dL_s^n + M_t^n\,. \tag{3.3}$$

We show that for every $T > 0$, $P - \lim_n \sup_{t\le T}\left|Y_t^{n,G} - \widetilde{Y}_t^{n,G}\right| = 0\,$. This holds provided that

$$\begin{aligned} P - \lim_n \sup_{t\le T}\left|U_t^n\right| &= 0 \\ P - \lim_n \sup_{t\le T}\left|\int_0^t [u^n(s) - A_G]h(X_s)dL_s^n\right| &= 0. \end{aligned} \tag{3.4}$$

By virtue of assumption (A-6) the function G' is continuous. Therefore $\sup_{t\le T}\left|U_t^n\right| \le T\sup_{t\le T}|h(X_t)|\sup_k|G'(\theta_k\Delta^n h(X_{t_{k-1}^n}) + \xi_k) - G'(\xi_k)| \to 0, n \to \infty$, i.e. the first part of (3.4) holds. To verify the validity of the second part note that

$$\begin{aligned} \int_0^t [u^n(s) - A_G]h(X_s)dL_s^n &= \Delta^n \sum_{k=1}^{[t/\Delta^n]} h(X_{t_{k-1}^n})\Big(G'(\xi_k) - EG'(\xi_k)\Big) \\ &:= \mathcal{M}_t^n \end{aligned}$$

and that the process $\mathcal{M}_t^n$ is a square integrable martingale with respect to the filtration $(\mathcal{F}_t^n)_{t\ge 0}$, with $\mathcal{F}_t^n = \sigma\{X_{t_{k-1}^n}, \xi_k; t_k^n \le t, k \le [t/\Delta^n]\}$. $\mathcal{M}_t^n$ possesses the predictable quadratic variation

$$\Big\langle \mathcal{M}^n \Big\rangle_t = (\Delta^n)^2 \sum_{k=1}^{[t/\Delta^n]} h^2(X_{t_{k-1}^n})E\Big(G'(\xi_k) - EG'(\xi_k)\Big)^2.$$

This and assumption (A-6) allow us to conclude that

$$\Big\langle \mathcal{M}^n \Big\rangle_T \le \Delta^n T \sup_{t\le T} h^2(X_t) E\Big(G'(\xi_1)\Big)^2$$

which in turn implies $P - \lim_n \Big\langle \mathcal{M}^n \Big\rangle_T = 0$. Then, due to Problem 1.9.2 in [10], we have $P - \lim_n \sup_{t\le T} \Big|\mathcal{M}^n_t\Big| = 0$, i.e. the second part in (3.4) holds as well.

By Theorem 4.1 from [1][Ch. 1 §4], Theorem 1 holds provided that

$$\mathcal{W} - \lim_{n\to\infty} (X_t, Y^n_t, \widetilde{Y}^{n,G}_t)_{t\ge 0} = (X_t, Y_t, Y^G_t)_{t\ge 0}. \tag{3.5}$$

For checking the validity of (3.5), note that (X_t) and $(\xi_k)_{k\ge 1}$ are independent objects, implying the independence of the processes $\int_0^t A_G H(X_s) dL^n_s$ and $M^n_t, M^{n,G}_t$. Hence, we get (3.5) by checking the convergence

$$\begin{aligned} &\mathcal{W} - \begin{cases} \lim_n (M^n_t)_{t\ge 0} = (B_1 W'_t + B_2 W_t)_{t\ge 0} \\ \lim_n (M^{n,G}_t)_{t\ge 0} = (B_G W_t)_{t\ge 0} \end{cases} \\ &P - \lim_n \sup_{t\le T} \Big| \int_0^t A_G h(X_s) d(L^n_s - s)\Big| = 0. \end{aligned} \tag{3.6}$$

The vector $(M^n_t, M^{n,G}_t)$ is a vector of square integrable martingales with independent homogeneous increments (with respect to the filtration generated by itself), with matrix of predictable quadratic variations elements $\Big\langle M^n \Big\rangle_t \equiv E\xi_1^2 L^n_t$, $\Big\langle M^{n,G} \Big\rangle_t \equiv EG(\xi_1)^2 L^n_t$, and $\Big\langle M^n, M^{n,G} \Big\rangle_t \equiv E\xi_1 G(\xi_1) L^n_t$. Since $\lim_n L^n_t = t$, it holds that

$$\begin{aligned} \lim_n \Big\langle M^n \Big\rangle_t &= (B_1^2 + B_2^2)t \\ \lim_n \Big\langle M^{n,G} \Big\rangle_t &= B_G^2 t \\ \lim_n \Big\langle M^n, M^{n,G} \Big\rangle_t &= B_2 B_G t. \end{aligned}$$

Therefore, by a vector version of the Donsker theorem (see e.g. Theorems 9.1.1 and 9.1.2 in [10] for the scalar version) the vector $M^n_t, M^{n,G}_t$ converges weakly (in the Skorohod-Lindvall topology), with the limit M_t, M^G_t being a continuous Gaussian vector martingale with predictable variations matrix elements $\Big\langle M \Big\rangle_t \equiv (B_1^2 + B_2^2)t$, $\Big\langle M^G \Big\rangle_t \equiv B_G^2 t$, and $\Big\langle M, M^G \Big\rangle_t \equiv B_2 B_G t$. Hence, using the orthogonality of M^G_t and $M_t - B_2 M^G_t / B_G$ and putting $W_t = (1/B^G) M_t$ and $W'_t = (1/B_1)[M_t - B_2 M^G_t / B_G]$, one can conclude that W_t, W'_t are independent Wiener processes and the first part of (3.6) holds.

The proof of the second part of (3.6) uses the fact that X_t is a continuous process. Letting $X_t^m = X_{[mt]/m}$, $m = 1, 2, \ldots$, introduce a sequence of random processes $(X_t^m)_{t\geq 0}, m \geq 1$. Since $\lim_m \frac{[mt]}{m} = t$, we get $P - \lim_m \sup_{j/m\leq s\leq (j+1)/m} \Big|h(X_s) - h(X_{j/m})\Big| = 0$. Hence, using $L_t^n \leq t$, we obtain

$$\begin{aligned}
&\sup_{t\leq T}\Big|\int_0^t A_G\{h(X_s) - h(X_{s-}^m)\}d(L_s^n - s) \\
\leq\ & 2|A_G|T| \sup_{j/m\leq s\leq (j+1)/m} \Big|h(X_s) - h(X_{j/m})\Big| := J^m ,
\end{aligned}$$

where $P - \lim_m J^m = 0$, that is (3.6) now holds provided that for every m, $P - \lim_n \sup_{t\leq T} \Big|\int_0^t A_G h(X_{s-}^m)d(L_s^n - s)\Big| = 0$. The last limit follows from

$$\sup_{t\leq T}\Big|\int_0^t A_G h(X_{s-}^m)d(L_s^n - s)\Big| \leq Tm|A_G| \sup_{t\leq T}|h(X_t)| \sup_{t\leq T}|L_t^n - t| = 0 .\Box$$

In the remainder of this section, we describe some (essentially known) continuity properties of the functional $\pi_t^G(y)$. Throughout, we let $\mathbf{D}$ denote the Skorkhod space of right continuous having limits from the left functions on $[0,\infty)$, equipped with the local supremum topology. Let the pair (X_t, Y_t^G) be defined on a probability space $(\Omega, \mathcal{F}, P)$ and $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{P})$ be its copy. Define a new pair $(\widetilde{X}_t, Y_t^G)$ on $(\Omega\times\widetilde{\Omega}, \mathcal{F}\otimes\widetilde{\mathcal{F}}, P\times)\widetilde{P})$. By the Kallianpur-Striebel formula [5], the conditional expectation $\pi_t^G(Y^G)$ is defined as

$$\pi_t^G(Y^G) = \frac{\widetilde{E}f(\widetilde{X}_t)\exp\Big\{\frac{1}{B_G^2}\Big(\int_0^t h(\widetilde{X}_s)dY_s^G - \frac{1}{2}\int_0^t h^2(\widetilde{X}_s)ds\Big)\Big\}}{\widetilde{E}\exp\Big\{\frac{1}{B_G^2}\Big(\int_0^t h(\widetilde{X}_s)dY_s^G - \frac{1}{2}\int_0^t h^2(\widetilde{X}_s)ds\Big)\Big\}}.$$

By Itô's formula,

$$\begin{aligned}
h(\widetilde{X}_t)Y_t^G &= \int_0^t h(\widetilde{X}_s)dY_s^G + \int_0^t Y_s^G\mathcal{L}h(s,\widetilde{X}_s)ds \\
&\quad + \int_0^t Y_s^G h'(\widetilde{X}_s)b(s,\widetilde{X}_s)d\widetilde{V}_s,
\end{aligned}$$

where $\widetilde{V}_t$, defined on $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{P})$, is a copy of the Wiener process V_t. Again applying Itô's formula we find

$$h(\widetilde{X}_t) = h(\widetilde{X}_0) + \int_0^t \mathcal{L}h(s,\widetilde{X}_s)ds + \int_0^t h'(\widetilde{X}_s)b(s,\widetilde{X}_s)d\widetilde{V}_s.$$

Therefore, taking into account the independence of the processes $\widetilde{X}_t, \widetilde{V}_t$ and Y_t^G, we arrive at

$$\begin{aligned}\int_0^t h(\widetilde{X}_s)dY_s^G &= h(\widetilde{X}_0)Y_t^G + \int_0^t [Y_t^G - Y_s^G]\mathcal{L}h(s,\widetilde{X}_s)ds \\ &\quad + \int_0^t [Y_t^G - Y_s^G]h'(\widetilde{X}_s)b(s,\widetilde{X}_s)d\widetilde{V}_s.\end{aligned}$$

For $y \in \mathbf{D}$, define

$$\begin{aligned}\Phi_t(\widetilde{X}, y) &= \exp\Big(\frac{1}{B_G^2}\Big[y_t h(\widetilde{X}_0) + \int_0^t \Big[(y_t - y_s)\mathcal{L}h(\widetilde{X}_s) - \frac{1}{2}h^2(\widetilde{X}_s)\Big]ds\Big) \\ &\quad \times \exp\Big(\frac{1}{B_G^2}\int_0^t (y_t - y_s)h'(\widetilde{X}_s)b(s,\widetilde{X}_s)d\widetilde{V}_s\Big) \qquad (3.7)\end{aligned}$$

and for $y \in \mathbf{D}$ such that it is defined, let

$$\pi_t^G(y) = \frac{\widetilde{E}f(\widetilde{X}_t)\Phi_t(\widetilde{X}, y)}{\widetilde{E}\Phi_t(\widetilde{X}, y)}. \qquad (3.8)$$

We will need the

Lemma 1 *Assume (A-1)–(A-4). Then $\pi_t^G(\cdot), t \ge 0$ is well defined and continuous on $\mathbf{D}$.*

Proof: Results of this kind, especially for continuous function $y \in \mathbf{D}$, are well known (see e.g. [12], [3],[11]). Thus, we give only a sketch of the proof.

By Jensen's inequality, for each $y \in \mathbf{D}$,

$$\begin{aligned}\widetilde{E}\Phi_t(\widetilde{X}, y) &\ge \exp\Big(\frac{1}{B_G^2}\widetilde{E}\Big[y_t h(\widetilde{X}_0) + \int_0^t \Big[(y_t - y_s)\mathcal{L}h(\widetilde{X}_s) \\ &\quad - \frac{1}{2}h^2(\widetilde{X}_s)\Big]ds\Big)\Big\} > 0.\end{aligned}$$

On the other hand, putting $\beta(s) = \frac{4}{B_G^2}((y_t - y_s)h'(\widetilde{X}_s)b(s,\widetilde{X}_s)$ and noticing that there exists a constant ℓ such that $|\beta(s)| \le \sup_{s\le t}|y_t - y_s|\ell$, we get $\widetilde{E}\exp\{\int_0^t \beta(s)d\widetilde{V}_s - \frac{1}{2}\int_0^t \beta^2(s)ds\Big\} = 1$. Coupled with $\mathcal{D}^\lambda h \le 0, \lambda = \sup_{s\le t}|y_t - y_s|$, we arrive at the following upper bound

$$\widetilde{E}\Phi_t^4(\widetilde{X}, y) \le \exp\Big\{\frac{\ell^2}{2}\sup_{s\le t}|y_t - y_s|^2\Big\}\widetilde{E}\exp\Big\{\frac{4|y_t||h(\widetilde{X}_0|}{B_G^2}\Big\}(<\infty). \qquad (3.9)$$

This upper bound jointly with the above-mentioned lower one guarantees that the functional $\pi_t^G(y)$ is well defined. Moreover, since $\Phi(\widetilde{X}, y)$ considered as a function of y is uniformly continuous in probability at any point $y \in \mathbf{D}$, one can conclude, using (3.9) for checking the uniform integrability, that the numerator and denominator in the right hand side of (3.8) are uniformly continuous in y. $\square$

4 Analysis of Limiting Model: Choice of Optimal Limiter

Theorem 1 shows that the limit model depends on the choice of limiter G. For a fixed G, the limit filtering model is characterized by two parameters A_G, B_G which define a natural "signal-to-noise" ratio

$$SN_G = \frac{A_G^2}{B_G^2}. \tag{4.1}$$

The next lemma shows that to larger values of the "signal-to-noise" ratio there correspond smaller value of the filtering error $\mathcal{E}_G(t) = E\Big(f(X_t) - \pi_t^G(Y^G)\Big)^2$.

Lemma 2 *The following implication holds: for every $t > 0$,*

$$SN_{G^1} \le SN_{G^2} \Longrightarrow \mathcal{E}_{G^1}(t) \ge \mathcal{E}_{G^2}(t).$$

Remark: For a related result, see [13].

Proof: For every limiter G, define a new observation $\widetilde{Y}_t^G = Y_t^G / A_G$ and note that since the σ-algebras generated by $\{Y_s^G, s \le t\}$ and $\{\widetilde{Y}_s^G, s \le t\}$ coincide, the mean square filtering errors corresponding to the observations $\{Y_s^G, s \le t\}$ and $\{\widetilde{Y}_s^G, s \le t\}$ coincide as well. Due to the definition of the process Y_t^G given in Theorem 1 we have

$$\begin{aligned} d\widetilde{Y}_t^G &= h(X_t)dt + \frac{B_G}{A_G} dW_t \\ \widetilde{Y}_0^G &= 0. \end{aligned} \tag{4.2}$$

Therefore, for the comparison of $\mathcal{E}_{G^1}(t)$ and $\mathcal{E}_{G^2}(t)$ one can use $\widetilde{Y}_t^{G^1}$ and $\widetilde{Y}_t^{G^2}$ as observation processes. To simplify notations, put $\gamma' = \frac{B_{G^1}}{A_{G^1}}$ and $\gamma'' = \frac{B_{G^2}}{A_{G^2}}$. Since

$$(\gamma')^2 = \frac{1}{SN_{G'}} \quad \text{and} \quad (\gamma'')^2 = \frac{1}{SN_{G''}},$$

we have that $(\gamma')^2 \geq (\gamma'')^2$. Take $(\gamma')^2 > (\gamma'')^2$ and consider two observation signals:

$$\begin{aligned}
dY_t^{''} &= h(X_t)dt + \gamma'' dW_t \\
dY_t^{'} &= h(X_t)dt + \gamma'' dW_t + \sqrt{(\gamma')^2 - (\gamma'')^2}\, d\widetilde{W}_t,
\end{aligned}$$

where $\widetilde{W}_t$ is independent of the process (X_t, W_t). It is clear that the diffusion parameter of the observation process for the first model is $(\gamma'')^2$ while for the second $(\gamma')^2$. Denote

$$\begin{aligned}
\mathcal{E}_{\gamma'}(t) &= E\Big(f(X_t) - E\Big(f(X_t)\Big|Y_s^{'}, s \leq t\Big)\Big)^2 \\
\mathcal{E}_{\gamma''}(t) &= E\Big(f(X_t) - E\Big(f(X_t)\Big|Y_s^{''}, s \leq t\Big)\Big)^2
\end{aligned}$$

and note that $\mathcal{E}_{\gamma'}(t) \equiv \mathcal{E}_{G^1}(t)$, $\mathcal{E}_{\gamma''}(t) \equiv \mathcal{E}_{G^2}(t)$. Therefore, it remains to check only the validity of the implication:

$$\gamma' > \gamma'' \Longrightarrow \mathcal{E}_{\gamma'}(t) \geq \mathcal{E}_{\gamma''}(t). \tag{4.3}$$

To this end, taking into account that $\sigma\{Y_s^{'}, \widetilde{W}_s, s \leq t\} \supseteq \sigma\{Y_s^{'}, s \leq t\}$, we get $\mathcal{E}_{\gamma'}(t) \geq \mathcal{E}(t)$, where $\mathcal{E}(t) = E\Big(f(X_t) - E\Big(f(X_t)\Big|Y_s^{'}, \widetilde{W}_s, s \leq t\Big)\Big)^2$. We next show that ($P$-a.s.)

$$\mathcal{E}(t) \equiv \mathcal{E}_{\gamma''}(t). \tag{4.4}$$

Indeed, noticing that $\sigma\{Y_s^{'}, \widetilde{W}_s, s \leq t\} = \sigma\{Y_s^{''}, \widetilde{W}_s, s \leq t\}$, one can conclude that

$$E\Big(f(X_t)\Big|Y_s^{'}, \widetilde{W}_s, s \leq t\Big) = E\Big(f(X_t)\Big|Y_s^{''}, \widetilde{W}_s, s \leq t\Big).$$

Taking into account now that $(X_t, Y_t^{''})$ and $(\widetilde{W}_t)$ are independent processes, we arrive at the following chain of equalities: for every bounded random variables α and β, which are measurable with respect to σ-algebras $\sigma\{Y_s^{''}, s \leq t\}$ and $\sigma\{\widetilde{W}_s, s \leq t\}$ respectively, and every bounded and measurable function $g(x)$,

$$E\Big(\alpha\beta E\Big(g(X_t)\Big|Y_s^{''}, \widetilde{W}_s, s \leq t\Big)\Big) \quad = \quad E\Big(\alpha\beta g(X_t)\Big)$$

$$
\begin{aligned}
&= E\Big(\alpha g(X_t)\Big)E\Big(\beta\Big) \\
&= E\Big(\alpha E\Big(g(X_t)\Big|Y_s^{''}, s \le t\Big)\Big)E(\beta) \\
&= E\Big(\alpha\beta E\Big(g(X_t)\Big|Y_s^{''}, s \le t\Big)\Big)
\end{aligned}
$$

which in turn implies (P-a.s.) $E\Big(f(X_t)\Big|Y_s^{''}, \widetilde{W}_s, s \le t\Big) = E\Big(f(X_t)\Big|Y_s^{''}, s \le t\Big)$. Then also (4.4) holds. □

Disregarding the constraint (A-6), one can find the limiter maximizing the "signal-to-noise" ratio.

Lemma 3 *Assume that the distribution of the random variable ξ_1 has a density $p(x)$ which is twice continuously differentiable, and possesses a finite Fisher information $I_p = \int_{\mathbf{R}} \frac{\left(p'(x)\right)^2}{p(x)}dx$. Then the limiter $G^\circ(x) = -\frac{p'(x)}{p(x)}$ has maximal "signal-to-noise" ratio among all limiters G which are smooth functions with $\int_{\mathbf{R}} G(x)p(x)dx = 0$ and $\int_{\mathbf{R}} G^2(x)p(x)dx < \infty$:*

$$
I_p = \frac{A^2_{G^\circ}}{B^2_{G^\circ}} \ge \frac{A^2_G}{B^2_G}.
$$

Proof: Let G be an admissible limiter. Under the assumptions of the lemma $SN_G = \frac{\left(\int_{\mathbf{R}} G'(x)p(x)dx\right)^2}{\int_{\mathbf{R}} G^2(x)p(x)dx}$. Integrating by parts and applying Cauchy-Schwartz's inequality we obtain

$$
\begin{aligned}
\Big(\int_{\mathbf{R}} G'(x)p(x)dx\Big)^2 &= \Big(\int_{\mathbf{R}} G(x)p'(x)dx\Big)^2 \\
&\le I_p \int_{\mathbf{R}} G^2(x)p(x)dx,
\end{aligned}
$$

that is $SN_G \le I_p$. On the other hand, since $\int_{\mathbf{R}}(G^\circ(x))^2p(x)dx = I_p$ and $(G^\circ)'(x) = \frac{p''(x)p(x)-(p'(x))^2}{p^2(x)}$, it holds that

$$
SN_{G^\circ} = \frac{\Big(\int_{\mathbf{R}} \Big(p''(x) - \frac{(p'(x))^2}{p(x)}dx\Big)^2}{\int_{\mathbf{R}}(G^\circ)(x))^2p(x)dx} = I_p.
$$

□

It is interesting to comment here on the relation of Lemma 3 and [4]: indeed, it is proved in [4] that the optimal nonlinear filter for the prelimit model converges to the optimal filter given both the limit observation Y_t^G *and* the limit observation Y_t. We show in the lemma below that the second component of the observation is actually superfluous.

Lemma 4 *Assume that the distribution of the random variable ξ_1 has a density $p(x)$ which is twice continuously differentiable and possesses a finite Fisher information $I_p = \int_{\mathbf{R}} \frac{\left(p'(x)\right)^2}{p(x)} dx$. Moreover, assume that the limiter $G^\circ(x) = -\frac{p'(x)}{p(x)}$ satisfies assumption (A-6). Then for any $t \geq 0$, a.s.*

$$E\left(f(X_t)\Big|Y_s, Y_s^{G^\circ}, s \leq t\right) = E\left(f(X_t)\Big|Y_s^{G^\circ}, s \leq t\right).$$

Proof: Due to Theorem 1, the observable processes Y_t, $Y_t^{G^\circ}$ are defined by as:

$$\begin{aligned} dY_t &= h(X_t)dt + B_1^\circ dW_t' + \frac{1}{\sqrt{I_p}} dW_t \\ dY_t^{G^\circ} &= I_p h(X_t)dt + \sqrt{I_p} dW_t, \end{aligned}$$

where $B_1^\circ = \sqrt{E\xi_1^2 - \frac{1}{I_p}}$. Evidently $I_p Y_t = Y_t^{G^\circ} + I_p B_1^\circ W_t'$, where the process $I_p B_1^\circ W_t'$ is independent of the processes $X_t, Y_t^{G^\circ}$, and thereby the desired statement is obtained similarly to the proof of Lemma 2. ☐

4.1 Minmax problems, and Gaussian worst case

As seen above, the "signal-to-noise" ration attached to the limiter, A_G^2/B_G^2, plays a decisive role in assessing the filtering performance. The choice of optimal limiter, however, requires the knowledge of the density $p(\cdot)$ (c.f. Lemma 3). Often, this knowledge is not available a-priori, and all that one knows is that for some class of densities $\mathcal{P}$ possessing zero mean and finite variance, $p \in \mathcal{P}$.

Although $p(\cdot)$ can be estimated from the data, this results in a cumbersome algorithm, both from the implementation and analysis aspects. It is natural therefore to define the "minmax" performance as the one for which the worst case performance over $\mathcal{P}$ is optimized by the choice of limiter G. Namely, define

$$SN_{\max} = \max_G \min_{p \in \mathcal{P}} \frac{A_G^2}{B_G^2},$$

where we take the ratio above to be 0 for such p, G with either $B_G = \infty$ or $A_G = B_G = 0$. Use σ_p to denote the variance of $p \in \mathcal{P}$.

Lemma 5

$$\inf_{p\in\mathcal{P}} I_p \geq SN_{\max} \geq \inf_{p\in\mathcal{P}} \frac{1}{\sigma_p^2}\,.$$

In particular, with $\sigma_{\max}^2 = \max_{p\in\mathcal{P}} \sigma_p^2$*, if* $N(0,\sigma_{\max}^2)\in\mathcal{P}$*, then* $SN_{\max} = \frac{1}{\sigma_{\max}^2}$*, and* $G^o(x) = cx$ *for any* $c \neq 0$*.*

Proof: Clearly,

$$SN_{\max} \leq \inf_{p\in\mathcal{P}} \sup_G \frac{A_G^2}{B_G^2} = \inf_{p\in\mathcal{P}} I_p\,.$$

On the other hand, the choice of $G(x)=x$ in the definition of $SN_{\max}$ leads to

$$SN_{\max} \geq \inf_{p\in\mathcal{P}} \frac{1}{\int x^2 p(x)dx} = \inf_{p\in\mathcal{P}} \frac{1}{\sigma^2}\,.$$

The conclusion of the lemma follows at once. $\square$

5 Asymptotically Optimal Filter

We have now completed the preliminaries needed to state our main results. Let assumptions (A-1)-(A-6) be fulfilled. For a fixed limiter G, consider the filtering problem for the prelimit model $(X_t, Y_t^{n,G})$ with $f(X_t)$ the signal which has to be filtered by the observation $Y_t^{n,G}$. By Theorem 1, the pair $(X_t, Y_t^{n,G})$ converges to (X_t, Y_t^G) in the Skorohod-Lindvall metric. If $\pi_t^G(Y^{n,G})$ is chosen as filtering estimate, it is reasonable to compare its mean square filtering error $E\Big(f(X_t) - \pi_t^G(Y^{n,G})\Big)^2$ with the corresponding optimal filtering error for the limit model $E\Big(f(X_t) - \pi_t^G(Y^G)\Big)^2$. Here, $\pi_t^G(Y^G) = E\Big(f(X_t)|Y_s^G, s\leq t)$.

For $f(\cdot)$ bounded continuous, Lemma 1 and Theorem 1 imply that

$$\lim_n E\Big(f(X_t) - \pi_t^G(Y^{n,G}\Big)^2 = E\Big(f(X_t) - \pi_t^G(Y^G)\Big)^2. \qquad (5.1)$$

This property reflects an asymptotical filtering equivalence for the prelimit model to the limit one.

Furthermore, for $f(x)$ satisfying only assumptions (A-3), (5.1) remains true under the uniform integrability condition for the family $(\pi_t^G(Y^{n,G})^2$, $n\geq 1$. To avoid a verification of the uniform integrability, let us introduce δ-asymptotically equivalent and δ-asymptotically optimal filtering estimates.

Definition 1 *For a fixed limiter G, a filtering estimate $\pi_t^{G,\delta}(y)$ is said to be asymptotically δ-equivalent to the filtering estimate $\pi_t^G(y)$, if for each $\delta > 0$ there exists a positive constant $N(t,\delta)$, depending on t and δ, such that $|\pi_t^{G,\delta}(y)| \le N(t,\delta)$ and*

$$\lim_n E\Big(f(X_t) - \pi_t^{G,\delta}(Y^{n,G})\Big)^2 \le E\Big(f(X_t) - \pi_t^G(Y^G)\Big)^2 + \delta. \tag{5.2}$$

Definition 2 *A filtering estimate $\pi_t^{\circ,\delta}(y)$ is said to be asymptotically δ-optimal, if for each $\delta > 0$ there exists a positive constant $N(t,\delta)$, depending on t and δ, such that $|\pi_t^{\circ,\delta}(y)| \le N(t,\delta)$ and*

$$\lim_n E\Big(f(X_t) - \pi_t^{\circ,\delta}(Y^{n,G})\Big)^2 \le \liminf_n E\Big(f(X_t) - E(f(X_t)|Y_s^n, s \le t)\Big)^2 + \delta. \tag{5.3}$$

Let us define $\pi_t^{G,\delta}(y)$ in the following way. Put

$$f^\delta(t,x) = \begin{cases} f(x) & |f(x)| \le N(t,\delta) \\ N(t,\delta) & f(x) \ge N(t,\delta) \\ -N(t,\delta) & f(x) \le -N(t,\delta), \end{cases} \tag{5.4}$$

where $N(t,\delta)$ is chosen such that $Ef^2(X_t)I(|f(X_t)| \ge N(t,\delta)) \le \delta$. We take $\pi_t^{G,\delta}(y)$ as the functional in (3.8) corresponding to the filtering estimate

$$\pi_t^{G,\delta}(Y^G) = E\Big(f^\delta(t,X_t)\Big|Y_s^G, s \le t\Big).$$

The function $f^\delta(t,x)$ is bounded by $N(t,\delta)$. Hence $\pi_t^{G,\delta}(y)$ is bounded by the same constant and we have

$$\lim_n E\Big(f(X_t) - \pi_t^{G,\delta}(Y^{n,G})\Big)^2 = E\Big(f(X_t) - \pi_t^{G,\delta}(Y^G)\Big)^2. \tag{5.5}$$

On the other hand, since

$$E\Big(\pi_t^G(Y^G) - \pi_t^{G,\delta}(Y^G)\Big)^2 \le E\Big(f(X_t) - f^\delta(X_t)\Big)^2 \le \delta,$$

we get

$$\begin{aligned} E\Big(f(X_t) - \pi_t^{G,\delta}(Y^G)\Big)^2 &= E\Big(f(X_t) - \pi_t^G(Y^G)\Big)^2 \\ &\quad + E\Big(\pi_t^G(Y^G) - \pi_t^{G,\delta}(Y^G)\Big)^2 \\ &\le E\Big(f(X_t) - \pi_t^G(Y^G)\Big)^2 + \delta. \end{aligned} \tag{5.6}$$

Assume now that the distribution of the random variable ξ_1 has a density $p(x)$ which is strictly positive everywhere, twice continuously differentiable, and such that (see [4])

(B-1) $\int_{\mathbf{R}} x^2 p(x)dx < \infty$;

(B-2) $|x|p(x) \to 0$ as $|x| \to \infty$;

(B-3) $\int_{\mathbf{R}} \frac{[p'(x)]^2}{p(x)} dx < \infty$;

(B-4) $\int_{\mathbf{R}} |(\ln p(x+\varepsilon))'' - (\ln p(x))''|dx \to 0$ as $\varepsilon \to 0$.

Note that (B-1)–(B-4) imply that the limiter $G^\circ(x) = -\frac{p'(x)}{p(x)}$ satisfies assumption (A-6). We are now in the position to revise Goggin's result from [4].

Theorem 2 *Assume (A-1)–(A-5) and (B-1)–(B-4).*

1. If $f(x)$ is bounded, then the filtering estimate $\pi_t^{G^\circ}(Y^{n,G^\circ})$ is asymptotically optimal, i.e. for each $t \ge 0$

$$\begin{aligned}\lim_n E\Big(f(X_t) - \pi_t^{G^\circ}(Y^{n,G^\circ})\Big)^2 &= \lim_n E\Big(f(X_t) - E\big(f(X_t)\big|Y_s^n, s \le t\big)\Big)^2 \\ &= E\Big(f(X_t) - \pi_t^{G^\circ}(Y^{G^\circ})\Big)^2\end{aligned}$$

2. If $f(x)$ satisfies (A-3) only, then the filtering estimate $\pi_t^{G^\circ,\delta}(y)$, chosen as $\pi_t^{G,\delta}(y)$ by replacing G with G°, is asymptotically δ-optimal, i.e. for each $t \ge 0$

$$\begin{aligned}&\lim_n E\Big(f(X_t) - \pi_t^{G^\circ,\delta}(Y^{n,G^\circ})\Big)^2 \le \\ &\liminf_n E\Big(f(X_t) - E\big(f(X_t)\big|Y_s^n, s \le t\big)\Big)^2 + \delta.\end{aligned}$$

Proof: 1. The desired conclusion follows from

$$\begin{aligned}&\lim_n E\Big(f(X_t) - E\big(f(X_t)\big|Y_s^n, s \le t\big)\Big)^2 = \\ &E\Big(f(X_t) - E\big(f(X_t)\big|Y_s, Y_s^{G^\circ}, s \le t\big)\Big)^2,\end{aligned}$$

(see [4]), from (5.1), and from Lemma 4.

2. The required statement follows from

$$E\Big(f(X_t) - E\big(f(X_t)\big|Y_s^n, s \le t\big)\Big)^2 \ge E\Big(f(X_t) - E\big(f^\delta(t,X_t)\big|Y_s^n, s \le t\big)\Big)^2 - \delta$$

and from

$$\lim_n E\Big(f(X_t) - E\big(f^\delta(t,X_t)\big|Y_s^n, s\le t\big)\Big)^2 \quad = \quad E\Big(f(X_t) - \pi_t^{G^\circ,\delta}(Y^{G^\circ}\Big)^2$$
$$= \lim_n E\Big(f(X_t) - \pi_t^{G^\circ,\delta}(Y^{n,G^\circ}\Big)^2.$$

□

6 Wide Bandwidth Noise

In the setup of Section 3, let $(\xi_t)_{t\in\mathrm{R}}$ denote a stationary in the strict sense, ergodic process, and consider the sequence of observations:

$$\dot{Y}_t^\epsilon = h(X_t) + \epsilon^{-1}\xi_{t\epsilon^{-2}}\,. \tag{6.1}$$

For a limiter G, define the transformed observation

$$Y_t^{\epsilon,G} = \int_0^t \epsilon^{-1}G(\epsilon\dot{Y}_s^\epsilon)ds\,. \tag{6.2}$$

We now have the analog of Theorem 1:

Theorem 3 *Assume (A-1)-(A-4),(A-6), and weak dependence conditions for the process $\xi(t)$:*

(A.7) $EG^4(\xi_0) < \infty$, $EG(\xi_0) = 0$

(A.8) $\int_0^\infty E^{1/4}\Big[E\big(G(\xi_t)\big|\xi_s\le 0\big)\Big]^4 dt < \infty$.

Then $\left(X_t, Y_t^{\epsilon,G}\right)_{t\ge 1}$ *converges, as* $\epsilon\to 0$, *weakly in the local uniform topology of* $\mathbf{D}$, *to* $(X_t, Y_t)_{t\ge 0}$ *with*

$$Y_t = A_G\int_0^t h(X_s)ds + B_G W_t,$$

where $A_G = EG'(\xi_0)$ *and* $B_G^2 = 2\int_0^\infty EG(\xi_0)G(\xi_s)ds$.

The proof of this theorem is similar to corresponding diffusion approximation results from [9] and so is omitted here.

As can be seen from the above, the limit process for the wide bandwidth noise case has the some structure as for the discrete time observation one. Moreover, for any fixed G, it can be shown that the Kallianpur-Striebel functional $\pi_t(y)$ applied to $Y_t^{\epsilon,G}$ results in filtering estimates satisfying the same continuity properties as in discrete time. Under some additional conditions on the joint distribution density for ξ_0 and ξ_t one can obtain an upper bound for the "signal-to-noise" ratio (see [8])

$$SN_G = \frac{E^2 G'(\xi_0)}{2\int_0^\infty EG(\xi_0)G(\xi_s)ds}.$$

Showing optimality for the wide bandwidth noise case is much harder. In fact, without further assumptions on ξ_t, one cannot hope for optimality, as the following example amply demonstrates.

Example 1 *Let ν_t be an (independent of the process X_t) Wiener process, let ξ_t denote the stationary ergodic diffusion process solution of the Itô equation $d\xi_t = -\xi_t dt + \sqrt{q(\xi_t)}d\nu_t$, where*

$$q(x) = \begin{cases} 1/4 + x^2 & |x| \le 1/2 \\ |x| & |x| > 1/2. \end{cases}$$

Assume $\dot{Y}_t^\epsilon = X_t + \epsilon^{-1}\xi_{t\epsilon^{-2}}$. It does not seem possible to construct a limiter with arbitrarily high "signal-to-noise" ratio. Hence, the filtering error for the limit model is bounded away from 0. On the other hand, one may construct filters whose errors converge to 0 with ϵ. Indeed, note that

$$d\dot{Y}_t^\epsilon = [a(t,X_t) - \epsilon^{-3}\xi_{t\epsilon^{-2}}]dt + b(t,X_t)dV_t + \epsilon^{-2}q(\xi_{t\epsilon^{-2}})d(\epsilon\nu_{t\epsilon^{-2}}), \quad (6.3)$$

where the diffusion parameter, which can be measured with zero error from $\dot{Y}_t^\epsilon$, is $b^2(t,X_t) + \epsilon^{-4}q^2(\xi_{t\epsilon^{-2}})$. Therefore for small ϵ, $q^2(\xi_{t\epsilon^{-2}})$ can be reconstructed with error converging to 0 with ϵ. It implies the same for $|\epsilon^{-1}\xi_{t\epsilon^{-2}}|$, and hence, using again $\dot{Y}_t^\epsilon$, an estimate can be constructed with arbitrarily small filtering error.

To avoid such pathologies, we introduce some structure to the model in continuous time. Assume that the process X_t satisfies the stochastic differential equation:

$$dX_t = a(X_t)dt + dV_t, \quad (6.4)$$

where X_0 is distributed with a density p_o. Assume also that the stationary process ξ_t is defined by the stochastic differential equation

$$d\xi_t = g(\xi_t)dt + d\nu_t,$$

where ν is a standard Wiener processes independent of the process X_t, and g is some smooth function. Evidently, the transformed process $\xi_t^\epsilon = \epsilon^{-1}\xi_{t\epsilon^{-2}}$ is defined as

$$d\xi_t^\epsilon = \epsilon^{-3}g(\epsilon\xi_t^\epsilon)dt + \epsilon^{-2}d(\epsilon\nu_{t\epsilon^{-2}}), \quad (6.5)$$

where $(\epsilon\nu_{t\epsilon^{-2}})$ is again a standard Wiener process independent of the process X_t. As before, let the observation satisfy

$$\dot{Y}_t^\epsilon = X_t + \xi_t^\epsilon. \tag{6.6}$$

Assuming a and g are smooth functions and putting

$$\begin{aligned} A_t &= Ea'(X_t) \\ C_t &= \sqrt{E(a'(X_t)))^2 - A_t^2 + E(g'(\xi(0))^2)} \end{aligned}$$

we associate with the original non linear model a linear one defined by stochastic differential equations with respect to independent Wiener processes v_t, w_t

$$\begin{aligned} dx_t &= A_t x_t dt + dv_t \\ dy_t &= C_t x_t dt + dw_t \end{aligned} \tag{6.7}$$

with x_0 a Gaussian random variable, independent of the processes w_t, v_t, of zero mean and variance σ_0^2, and $y_0 = 0$.

Theorem 4 *Assume that p_o, a, and g are continuously differentiable functions (p_o once; a and g twice, having bounded derivatives, g is bounded). Assume also $I_{p_o} := \int_{\mathbf{R}} \frac{[p_o'(x)]^2}{p_o(x)} dx < \infty$ and $\sigma_0^2 = 1/I_{p_0}$. Then for each $t \geq 0$*

$$\lim_{\epsilon \to 0} E(X_t - E(X_t | \dot{Y}_s^\epsilon, s \leq t))^2 \geq E(x_t - E(x_t | y_s, s \leq t))^2.$$

Proof: The proof uses the ideas in [2]. For $T > 0$, consider the filtering problem on the time interval $[0, T]$. For a deterministic $\phi \in \{\psi : \int_0^T \dot{\psi}^2(t)dt < \infty\}$, let $\mu_{\delta,\phi}^\epsilon$ denote the law induced on $(C[0,T])^2$ by the processes

$$\begin{aligned} dx_t^{\delta,\phi} &= a(x_t^{\delta,\phi} - \delta\phi_t)dt + \delta\dot{\phi}_t dt + dv_t\,, \quad x_0^{\delta,\phi} = x_0 + \delta\phi(0)\,, \\ d\xi_t^{\epsilon,\delta,\phi} &= \left(\frac{g(\epsilon(\xi_t^{\epsilon,\delta,\phi} + \delta\phi_t))}{\epsilon^3} - \delta\dot{\phi}_t \right) dt + \frac{1}{\epsilon^2} d(\epsilon\nu_{t\epsilon^{-2}}). \end{aligned} \tag{6.8}$$

We use $\mu^\epsilon = \mu_{0,\phi}^\epsilon$ throughout. Note that the law of $\{(x_t^{0,\phi}, \xi_t^{\epsilon,0,\phi})\}$ is identical to the law of $\{(x_t, \xi_t^\epsilon)\}$, and that $x_t + \delta\phi_t = x_t^{\delta,\phi}, \xi_t^\epsilon - \phi_t = \xi_t^{\epsilon,\delta,\phi}$. Further, by our assumptions on a, g, and ϕ, $d\mu_{\delta,\phi}^\epsilon / d\mu^\epsilon$ exists, and

$$\frac{d\mu^\epsilon_{\delta,\phi}}{d\mu^\epsilon} = \exp\left(\int_0^T \alpha_s dv_s - \frac{1}{2}\int_0^T \alpha_s^2 ds + \int \beta_s d(\epsilon\nu_{s\epsilon^{-2}}) - \frac{1}{2}\int_0^T \beta_s^2 ds\right),$$

where

$$\begin{aligned}
\alpha_t &= a(x_t^{\delta,\phi} - \delta\phi_t) + \delta\dot{\phi}_t - a(x_t^{\delta,\phi}) \\
\beta_t &= \frac{g(\epsilon(\xi_t^{\epsilon,\delta,\phi} + \delta\phi_t)) - g(\epsilon\xi_t^{\epsilon,\delta,\phi})}{\epsilon} - \epsilon^2\delta\dot{\phi}_t.
\end{aligned}$$

Also, $\left(1 - \frac{d\mu^\epsilon_{\delta,\phi}}{d\mu^\epsilon}\right)^2/\delta^2$ is easily checked to be uniformly integrable (in δ, with ϵ, ϕ fixed). Let $\{\phi_i\}$ denote a complete, smooth orthonormal basis of $L^2[0,T]$. Then, by a direct computation, as in Theorem 4 of [2],

$$\begin{aligned}
K^\epsilon_{i,j} &:= \lim_{\delta\to 0} E\left(\frac{1}{\delta^2}\left(1 - \frac{d\mu^\epsilon_{\delta,\phi_i}}{d\mu^\epsilon}\right)\left(1 - \frac{d\mu^\epsilon_{\delta,\phi_j}}{d\mu^\epsilon}\right)\right) \\
&= \frac{\phi_i(0)\phi_j(0)}{\sigma_0^2} \\
&\quad + E\left\{\int_0^T (\dot{\phi}_i(t) - a'(x_t)\phi_i(t))(\dot{\phi}_j(t) - a'(x_t)\phi_j(t))dt\right. \\
&\quad \left. + \int_0^T (g'(\epsilon\xi_t^\epsilon)\phi_i(t) - \epsilon^2\dot{\phi}_i(t))(g'(\epsilon\xi_t^\epsilon)\phi_j(t) - \epsilon^2\dot{\phi}_j(t))dt\right\} \\
&= \frac{\phi_i(0)\phi_j(0)}{\sigma_0^2} + E\left\{(1+\epsilon^4)\int_0^T \left[\dot{\phi}_i(t) - \phi_i(t)\frac{a'(x_t) + \epsilon^2 g'(\epsilon\xi_t^\epsilon)}{1+\epsilon^4}\right]\right. \\
&\quad \times\left[\dot{\phi}_j(t) - \phi_j(t)\frac{a'(x_t) + \epsilon^2 g'(\epsilon\xi_t^\epsilon)}{1+\epsilon^4}\right]dt \\
&+ \left.\int_0^T \phi_i(t)\phi_j(t)\left(g'(\epsilon\xi_t^\epsilon)^2 + a'(x_t)^2 - \frac{(a'(x_t) + \epsilon^2 g'(\epsilon\xi_t^\epsilon))^2}{1+\epsilon^4}\right)dt\right\}.
\end{aligned}$$

The $K^\epsilon_{i,j}$'s are the elements of the (infinite) information matrix associated with the nonlinear diffusion X_t and observation $\{\dot{Y}^\epsilon_t, 0 \le t \le T\}$. Consider the linear, Gaussian system:

$$\begin{aligned}
dx_t^\epsilon &= A_t^\epsilon x_t^\epsilon dt + B_t^\epsilon dv_t\,, \\
dy_t^\epsilon &= C_t^\epsilon \bar{x}_t^\epsilon dt + d(\epsilon\nu_{t\epsilon^{-2}}),
\end{aligned} \tag{6.9}$$

where $A_t^\epsilon = E(a'(X_t) + \epsilon^2 g'(\epsilon\xi_t^\epsilon)), B^\epsilon = \sqrt{1+\epsilon^4}, (C_t^\epsilon)^2 = E(g'(\epsilon\xi_t^\epsilon)^2 + a'(X_t)^2) - (A_t^\epsilon)^2$, and x_o^ϵ is Gaussian with $Ex_0^\epsilon = 0, E(x_0^\epsilon)^2 = \sigma_0^2, y_0^\epsilon = 0$.

By Theorem 5 of [2], the linear, Gaussian system (6.9) possesses the same information matrix $\{K^{\epsilon}_{i,j}\}$, and furthermore, by choosing as the $\{\phi_i\}$ the Karhunen-Loeve base of x^{ϵ}_t, the filtering error matrix associated with the coefficients of the Karhunen Loeve expansion equals precisely this information matrix. The standard Cramér–Rao bound then implies (c.f. Theorem 6 of [2]) that the filtering error of X_T with respect to the observation $\dot{Y}^{\epsilon}_t$ is bounded below by that of the linear model (6.9). Taking now $\epsilon \to 0$, the conclusion of the theorem follows. $\square$

Acknowledgement

This work was initiated during a visit of the authors to the Center for Applied Mathematical Sciences at U.S.C. We thank Prof. Boris Rozovskii for inviting us to visit there, for posing some of the questions leading to this work, and for many useful discussions. We also thank the people of CAMS for their kind hospitality.

References

[1] P. Billingsley. *Convergence of Probability Measures.* New York: J. Wiley, 1979.

[2] B.Z. Bobrovsky and M. Zakai. A lower bound on the estimation error for certain diffusion processes, *IEEE Trans. Inf. Theory*, **22** (1976), 45–52.

[3] M. Chaleyat-Maurel and D. Michel. “Une propriété de continuité en filtrage non linéaire” *Stochastics,* **19** (1986), 11–40.

[4] E.M. Goggin. Convergence of filters with applications to the Kalman–Bucy case, *IEEE Trans. Inf. Theory*, **38** (1992), 1091–1100.

[5] G. Kallianpur and C. Striebel. Estimation of stochastic systems: Arbitrary system process with additive white noise observation errors, *AMS,* **39** (1968), 785–801.

[6] H.J. Kushner. *Weak Convergence Method and Singular Perturbed Stochastic Control and Filtering Problems.* Boston: Birkhäser, 1990.

[7] H.J. Kushner and W.J. Runggaldier. Filtering and control for wide bandwidth noise driven systems, *IEEE Transactions on Automatic Control,* **AC-23** (1987), 123–133.

[8] R.S. Liptser and S.V. Lototsky. Diffusion approximation and robust Kalman filter *Journal of Mathematical Systems, Estimation, and Control,* **2(3)** (1992), 263–274.

[9] R. Sh. Liptser and W.J. Runggaldier. Diffusion approximations for filtering, *Stochastic Processes and Their Applications,* **38** (1991), 205–238.

[10] R. Sh. Liptser and A.N. Shiryayev. *Theory of Martingales.* Amsterdam: Kluwer Acad. Publ., 1989.

[11] J. Picard. Robustness de la solution des problemes de filtrage avec bruit blanc independant, *Stochastics,* **13** (1994), 229–245.

[12] B. Rozovskii. *Stochastic Evolution Systems. Linear Theory and Application to Nonlinear Filtering.* Dordrecht: D. Reidel Publ. Co., 1990.

[13] O. Zeitouni and M. Zakai. On the optimal tracking problem, *SIAM J. Control*, **30** (1992), 426-439. Erratum, **32** (1994), 1194.

Department of Electrical Engineering, Tel Aviv University, Tel Aviv 69978, Israel

Department of Electrical Engineering, Technion, Haifa 32000, Israel

Communicated by Clyde Martin