Efficiency of an Approximate Filter for a Particular Class of Nonlinear Diffusions with Observations Corrupted by Small Noise

Paula Milheiro-Oliveira
Faculdade de Engenharia da U.P., Rua dos Bragas, P-4050-123 Porto, Portugal, poliv@fe.up.pt

Jean Picard
LMA, Univ. Blaise Pascal, F-63177 Aubière Cedex, France, picard@ucfma.univ-bpclermont.fr

Abstract

The asymptotic behaviour of a nonlinear continuous time approximate filter when the variance of the observation noise tends to 0 is investigated. We consider a particular class of signals modeled by a two-dimensional quasi-linear diffusion from which only one of the components is noisy, and we assume that a one-dimensional linear function of the signal, depending only of the unnoisy component, is observed in a low noise channel. Under some detectability assumptions the unobserved signal can be restored by means of an approximate nonlinear filter. We establish that the filtering error converges to 0 and we give an upper bound for the convergence rate. The efficiency of the approximate filter is compared with the efficiency of the optimal filter and the order of magnitude of the error between the two filters, as the observation noise vanishes, is obtained. A more general case is briefly presented.

1 Introduction

The problem of filtering a random signal based on noisy observations has been considered by several authors. In particular, the case of small observation noise has been widly studied, some articles being devoted to the research of approximate filters asymptotically efficient when the observation noise vanishes. In [4, 6, 1] the case of a nonlinear one-dimensional system observed through an injective observation function is studied. In [7, 8] the research is extended to the multidimensional case. Again, an assumption of injectivity of the observation function is required. The extended Kalman filter (EKF) is studied in [8]. In the case of a non-injective observation function the unobserved process can not always be restored [3]. However, this is possible in some situations such as [2, 9, 10, 12] and efficient suboptimal filters can be derived. Within the present work, we are concerned by another of these situations where the state process can be restored. We consider the framework of [11], that is the problem of estimating $X_t = (x_t^{(1)}, x_t^{(2)})$ given by the Itô equation

$$\left\{ \begin{array}{rcl} dx_t^{(1)} & = & f_1(x_t^{(1)}, x_t^{(2)})dt \\ dx_t^{(2)} & = & f_2(x_t^{(1)}, x_t^{(2)})dt + \sigma(x_t^{(1)}, x_t^{(2)})dw_t \end{array} \right. \quad (1)$$

with initial condition $X_0 = (x_0^{(1)}, x_0^{(2)})$, when the observation process is modeled by

$$dy_t = h(x_t^{(1)})\, dt + \varepsilon\, d\bar{w}_t\,, \quad (2)$$

$\{w_t\}$ and $\{\bar{w}_t\}$ being standard independent Wiener processes with values in $\mathbf{R}$ and ε being a small nonnegative parameter.

In [11] this problem has been dealt with by means of a formal asymptotic expansion of the optimal filter in a stationary situation and the solution $M_t = (m_t^{(1)}, m_t^{(2)})$ of

$$dM_t = f(M_t)dt + R_t[dy_t - h(m_t^{(1)})dt]\,, \quad (3)$$

$$R_t \overset{def}{=} \left[\begin{array}{c} \sqrt{\dfrac{2\sigma(M_t)F_{12}(M_t)}{h'(m_t^{(1)})\,\varepsilon}} \\ \dfrac{\sigma(M_t)}{\varepsilon} \end{array} \right], \quad (4)$$

with $F_{12} = \partial f_1/\partial x_2$ and with initial condition $M_0 = E[X_0]$, is proposed as a finite dimension approximate filter. Our goal is to prove that, under certain regularity assumptions on the functions involved in the model, X_t can be estimated by means of M_t and that this approximate filter is asymptotically as efficient as the EKF. Notice that (3-4) corresponds to the EKF with stationary gain if one neglects the contribution of the derivatives of f other than $\partial f_1/\partial x_2$.

In this work we consider the particular case in which σ, h' and F_{12} are constant, so that the system (1-2) is

$$\left\{ \begin{array}{rcl} dx_t^{(1)} & = & \left(f_1^0(x_t^{(1)}) + F_{12}x_t^{(2)}\right) dt \\ dx_t^{(2)} & = & f_2(x_t^{(1)}, x_t^{(2)})\, dt + \sigma\, dw_t \\ dy_t & = & h' x_t^{(1)}\, dt + \varepsilon\, d\bar{w}_t \end{array} \right. . \quad (5)$$

It will be referred as the almost linear case.

This class of problems includes the case of real time estimation of the position and speed of a moving body on $\mathbf{R}$ submitted to a dynamical force (described by f_2) plus a random force (described by σ), based on noisy observations of the position. Indeed, if one denotes by $x_t^{(1)}$ the position of the moving body and its speed by $x_t^{(2)}$ model (5) applies with $f_1^0 = 0$, $F_{12} = 1$. For instance, in the case of free fall of a body through the atmosphere, $f_2(x^{(1)}, x^{(2)}) = \rho_0 e^{-x^{(1)}/k}(x^{(2)})^2/(2\beta) - g$, where ρ_0 is the reference air density, k is the atmosphere thickness, β is the ballistic coefficient of the body and g is the accelaration due to gravity.

In Section 3 we present an extention to the general non-linear model (1-2). A detailed study of this general case can be found in [5].

2 The almost linear case

We consider system (5) and the filter M_t given by (3). In this particular case R_t does not depend on t.

The following assumptions are used:

(H1) X_0 is a random variable with finite moments;

(H2) $\{w_t\}$ and $\{\bar{w}_t\}$ are standard independent Wiener processes independent from X_0;

(H3) h' is a positive constant;

(H4) the function f is C^3 with bounded partial derivatives and F_{12} is a positive constant;

and notations:

$f = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}$, $\Sigma = \begin{bmatrix} 0 \\ \sigma \end{bmatrix}$, $F = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix}$ is the Jacobian matrix of f;

$\nabla_0 \Phi = \dfrac{\partial \Phi}{\partial X_0}$ is a line-vector;

$\mathcal{F}_t$ and $\mathcal{Y}_t$ are the filtrations generated by $\{X_0, w_t, \bar{w}_t\}$ and by $\{y_t\}$, respectively.

The symbol * is used for the transposition of matrices.

When describing the behaviour of approximate filters, we will write asymptotic expressions with the meaning given by the following definition.

Definition Consider a (real or vector-valued) stochastic process $\{\xi_t\}$. If β is real and $p \geq 1$, we will write that $\xi_t = \mathcal{O}(\varepsilon^\beta)$ in L^p when for some $q, \alpha > 0$ and some positive constants C_1, C_2, c_3, one has $E\big[\|\xi_t\|^p\big]^{1/p} \leq C_1 e^{-c_3 t/\varepsilon^\alpha}/\varepsilon^q + C_2 \varepsilon^\beta$, for $t \geq 0$ and ε small. Then $\{\xi_t\}$ is said to converge to 0 with rate of order ε^β.

We establish that the errors $X_t - M_t$ and $\hat{X}_t - M_t$ converge to 0 in L^p as the observation noise vanishes. The rates of convergence are stated in the Theorems bellow. The proofs follow the method of [8] and details can be found in [5].

Theorem 2.1 *Assume* **(H1)** *to* **(H4)** *in model* (5). *Then the filter M_t given by equation (3) verifies*

$$x_t^{(1)} - m_t^{(1)} = \mathcal{O}(\varepsilon^{3/4}), \quad x_t^{(2)} - m_t^{(2)} = \mathcal{O}(\varepsilon^{1/4})$$

in L^p for any $p \geq 1$.

Proof First note that the system can be re-scaled, by replacing the processes $x_t^{(1)}$, $x_t^{(2)}$ and y_t by $x_t^{(1)}/(\sigma F_{12})$, $x_t^{(2)}/\sigma$ and $y_t/(\sigma F_{12} h')$, respectively. Then the problem can be reduced to the case $\sigma = F_{12} = h' = 1$.

Consider a change of basis defined by the matrix $T \stackrel{def}{=} \begin{bmatrix} \sqrt{2/\varepsilon} & -1 \\ 0 & 1 \end{bmatrix}$. The process $Z_t \stackrel{def}{=} T(X_t - M_t)$ is solution of

$$dZ_t = A_t\, Z_t\, dt + U \begin{bmatrix} dw_t \\ d\bar{w}_t \end{bmatrix} \qquad (6)$$

where $A_t = \dfrac{1}{\sqrt{2\varepsilon}} \begin{bmatrix} -1 & 1 \\ -1 & -1 \end{bmatrix} + \widetilde{A}_t$, $U = \begin{bmatrix} -1 & -1 \\ 1 & -1 \end{bmatrix}$ and $\widetilde{A}_t$ is a 2×2 matrix valued process which is uniformly bounded as ε converges to 0.

From Itô's Formula and (6), the process $\|Z_t\|^2$ is solution of $d(\|Z_t\|^2) = Z_t^*(A_t + A_t^*)Z_t dt + \mathrm{trace}(U^*U)dt + 2Z_t^*U \begin{bmatrix} dw_t \\ d\bar{w}_t \end{bmatrix}$ and one concludes, after some manipulation, that, for ε small enough, there is some positive constant α such that

$$\frac{d}{dt} E[\,\|Z_t\|^{2p}\,] \leq -\frac{\alpha}{2\sqrt{\varepsilon}} p\, E[\,\|Z_t\|^{2p}\,] + C_p' \varepsilon^{(p-1)/2}. \quad \diamond$$

Theorem 2.2 *Assume* **(H1)** *to* **(H4)** *in model* (5) *and*

(H5) *The law of X_0 has a C^1 positive density p_0 with respect to the Lebesgue measure and $\nabla p_0(X_0) \,/\, p_0(X_0)$ is in L^2.*

Then the filter M_t given by equation (3) satisfies

$$\hat{x}_t^{(1)} - m_t^{(1)} = \mathcal{O}(\varepsilon^{5/4}), \quad \hat{x}_t^{(2)} - m_t^{(2)} = \mathcal{O}(\varepsilon^{3/4})$$

in L^2.

Proof A sequence of steps are needed: a change of probability measure, the differentiation with respect to the initial condition and an integration by parts.

We suppose as in Theorem 2.1 that $\sigma = F_{12} = h' = 1$.

The first step is to define a new probability measure $\tilde{P}$ on $\mathcal{F}_t$ by $\left.\frac{d\tilde{P}}{dP}\right|_{\mathcal{F}_t} = \exp\left\{\int_0^t \Sigma^* \check{P}^{-1}(X_s - M_s)dw_s \quad -\frac{1}{2}\int_0^t(\Sigma^*\check{P}^{-1}(X_s - M_s))^2 ds - \frac{1}{\varepsilon}\int_0^t x_s^{(1)} d\bar{w}_s - \frac{1}{2\varepsilon^2}\int_0^t x_s^{(1)} ds\right\}$ with $\check{P} = \begin{bmatrix} \sqrt{2}\varepsilon^{3/2} & \varepsilon \\ \varepsilon & \sqrt{2}\varepsilon^{1/2} \end{bmatrix}$. Then use the properties of the derivatives of the processes involved with respect to X_0, together with (1) and (3) to estimate the process $V_t = \Big(\nabla_0 \log(L_t\Lambda_t)(\nabla_0 X_t)^{-1} + (X_t - M_t)^*\check{P}^{-1}\Big)\begin{bmatrix} 1 & 1 \\ 0 & \sqrt{2/\varepsilon} \end{bmatrix}$ and conclude that $V_t = \mathcal{O}(\varepsilon^{-1/4})$ in L^2.

The result follows from applying an integration by parts formula to evaluate the order of magnitude of $E\left[\nabla_0 \log(L_t\Lambda_t)(\nabla_0 X_t)^{-1}\begin{bmatrix} 1 & 1 \\ 0 & \sqrt{2/\varepsilon} \end{bmatrix} \Big| \mathcal{Y}_t\right]$ and establish that $(\hat{X}_t - M_t)^*\check{P}^{-1}\begin{bmatrix} 1 & 1 \\ 0 & \sqrt{2/\varepsilon} \end{bmatrix} = \mathcal{O}(\varepsilon^{-1/4})$. ⋄

3 Extensions

The previous results can be extended to the more general case modeled by (1-2) (see [5] for proofs and details). Besides **(H1)** and **(H2)** in Section 2 the following assumptions are needed:

(H3') the function h is C^3 with bounded derivatives, and h' is positive;

(H4') the function f is C^3 with bounded partial derivatives and $F_{12} = \partial f_1/\partial x_2$ is positive;

(H5') the function σ is C^2 with bounded partial derivatives;

(H6.δ) one has
$$\frac{1}{\delta} \le \frac{\sigma(x)}{\bar{\sigma}} \le \delta,\ \frac{1}{\delta} \le \frac{h'(x_1)}{\bar{H}} \le \delta,\ \frac{1}{\delta} \le \frac{F_{12}(x)}{\bar{F}} \le \delta$$
for any $x = (x_1, x_2)$, and for some positive $\bar{\sigma}$, $\bar{H}$ and $\bar{F}$.

(H7) The law of X_0 has a C^1 positive density p_0 with respect to the Lebesgue measure and $\nabla p_0(X_0)\ /\ p_0(X_0)$ is in L^2.

Theorem 3.1 *Assume* **(H1)** *to* **(H5')**. *There exists some universal* $\delta > 1$ *such that, if* **(H6.δ)** *holds, then one has* $x_t^{(1)} - m_t^{(1)} = \mathcal{O}(\varepsilon^{3/4})$, $\quad x_t^{(2)} - m_t^{(2)} = \mathcal{O}(\varepsilon^{1/4})$ *in* L^p *for any* $p \ge 1$.

Theorem 3.2 *Assume* **(H1)** *to* **(H7)**. *If* δ *in* **(H6.δ)** *is close enough to 1, then the filter* M_t *in (3) satisfies* $\hat{x}_t^{(1)} - m_t^{(1)} = \mathcal{O}(\varepsilon)$, $\quad \hat{x}_t^{(2)} - m_t^{(2)} = \mathcal{O}(\sqrt{\varepsilon})$ *in* L^2.

References

[1] A. Bensoussan, On some approximation techniques in non linear filtering theory, *Stochastic Differential Systems, Stochastic Control Theory and Applications*, IMA Vol. in Math. and Appl., **10**, Springer, 1988.

[2] W.H. Fleming and E. Pardoux, Piecewise monotone filtering with small observation noise, *SIAM J. Control Optim.,* **27**(5) (1989), 1156–1181.

[3] A. Gégout-Petit, Approximate filter for the conditional law of a partially observed process in nonlinear filtering, *SIAM J. Control Optim.,* **36**(4) (1998), 1423–1447.

[4] R. Katzur, B.Z. Bobrovsky, and Z. Schuss, Asymptotic analysis of the optimal filtering problem for one-dimensional diffusions measured in a low noise channel, *SIAM J. Appl. Math.,* **44** (1984), Part I: 591–604, Part II: 1176–1191.

[5] P. Milheiro de Oliveira and J. Picard, Approximate Nonlinear Filtering for a Two-dimensional Diffusion with One-dimensional Observations in a Low Noise Channel, submitted to *SIAM J. Control and Optim.*

[6] J. Picard, Nonlinear filtering of one-dimensional diffusions in the case of a high signal-to-noise ratio, *SIAM J. Appl. Math.,* **46**(6) (1986), 1098–1125.

[7] J. Picard, Nonlinear filtering and smoothing with high signal-to-noise ratio, *Stochastic Processes in Physics and Engeneering (Bielefeld 1986)*, Reidel, 1988.

[8] J. Picard, Efficiency of the extended Kalman filter for non linear systems with small noise, *SIAM J. Appl. Math.,* **51**(3) (1991), 843-855.

[9] J. Picard, Estimation of the quadratic variation of nearly observed semimartingales with application to filtering, *SIAM J. Control Optim.,* **31**(2) (1993), 494–517.

[10] M.C. Roubaud, Filtrage linéaire par morceaux avec petit bruit d'observation, *Appl. Math. Optim.,* **32** (1995), 163–194.

[11] I. Yaesh, B.Z. Bobrovsky and Z. Schuss, Asymptotic analysis of the optimal filtering problem for two dimensional diffusions measured in a low noise channel, *SIAM J. Appl. Math.,* **50**(4) (1990), 1134–1155.

[12] Q. Zhang, Nonlinear filtering and control of a switching diffusion with small observation noise, *SIAM J. Control Optim.,* **36**(5) (1998), 1638–1668.