APPROXIMATE NONLINEAR FILTERING FOR A TWO-DIMENSIONAL DIFFUSION WITH ONE-DIMENSIONAL OBSERVATIONS IN A LOW NOISE CHANNEL*

PAULA MILHEIRO DE OLIVEIRA† AND JEAN PICARD‡

Abstract. The asymptotic behavior of a nonlinear continuous time filtering problem is studied when the variance of the observation noise tends to 0. We suppose that the signal is a two-dimensional process from which only one of the components is noisy and that a one-dimensional function of this signal, depending only on the unnoisy component, is observed in a low noise channel. An approximate filter is considered in order to solve this problem. Under some detectability assumptions, we prove that the filtering error converges to 0, and an upper bound for the convergence rate is given. 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.

Key words. stochastic differential models, nonlinear filtering, approximate filters

AMS subject classifications. 93E11, 60G35, 60F99

PII. S0363012902363920

1. Introduction. Due to its vast application in engineering, the problem of filtering a random signal X_t from noisy observations of a function $h(X_t)$ of this signal has been considered by several authors. In particular, the case of small observation noise has been widely studied, and several articles are devoted to the research of approximate filters which are asymptotically efficient when the observation noise vanishes. Among them, one notices a first group in which a one-dimensional system is observed through an injective observation function h (see [4, 5, 7, 1]); in this case, the filtering error is small when the observation noise is small, and one can find efficient suboptimal finite-dimensional filters. The multidimensional case appears later with [8, 9], but an assumption of injectivity of h is again required; in particular, the extended Kalman filter is studied in [9]. See also previous work by Krener [6] for systems with linear observations. When h is not injective, the process $\{X_t\}$ cannot always be restored from the observation of $\{h(X_t)\}$, so the filtering error is not always small; such a case is studied in [3]. However, there are some classes of problems in which $\{X_t\}$ can be restored from $\{h(X_t)\}$; in these cases, the filtering error is small, and one again looks for efficient suboptimal filters. For instance, $\{X_t\}$ is sometimes obtained from $\{h(X_t)\}$ and its quadratic variation; see [2, 10, 11, 13]. Here, we are interested in another case in which $h(X_t)$ is differentiable with respect to the time t, and $\{X_t\}$ is obtained from $\{h(X_t)\}$ and its derivative. As opposed to [9], the existence of a Lipschitz inverse of h is not assumed in this paper, as the dimension of the measurements that we consider is lower than that of the state. More precisely, we consider the framework of [12], which we now describe.

We consider the two-dimensional process $X_t = (x_t^{(1)}, x_t^{(2)})$ given by the Itô equa-

---

†Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias, P–4200–465 Porto, Portugal (poliv@fe.up.pt).

‡Laboratoire de Mathématiques Appliquées (CNRS–UMR 6620), Université Blaise Pascal, F–63177 Aubière Cedex, France (Jean.Picard@math.univ-bpclermont.fr).

tion

$$(1.1) \quad \begin{cases} dx_t^{(1)} &= f_1(x_t^{(1)}, x_t^{(2)}) \, dt, \\[2mm] dx_t^{(2)} &= f_2(x_t^{(1)}, x_t^{(2)}) \, dt + \sigma(x_t^{(1)}, x_t^{(2)}) \, dw_t, \end{cases}$$

with initial condition $X_0 = (x_0^{(1)}, x_0^{(2)})$, and we are concerned by the problem of estimating the signal X_t when the observation process is modelled by the equation

$$(1.2) \qquad\qquad dy_t = h(x_t^{(1)}) \, dt + \varepsilon \, d\bar{w}_t,$$

where $\{w_t\}$ and $\{\bar{w}_t\}$ are standard independent real-valued Wiener processes and ε is a small nonnegative parameter. In particular, if $f_1(x_1, x_2) = x_2$, then $x_t^{(1)}$ is the position of some moving body on $\mathbf{R}$, $x_t^{(2)}$ is its speed, the body is submitted to a dynamical force described by f_2 and to a random force described by σ, and one has a noisy observation of the position. This class of problems arises in practice in tracking RADAR applications, for instance, as well as in control and communications engineering. The use of the method of proof introduced in [7] and later extended to [9] in the class of systems (1.1)–(1.2) is not covered by previous work.

If $\varepsilon = 0$ and if the functions h and $x_2 \mapsto f_1(x_1, x_2)$ are injective, then the signal X_t can (at least theoretically) be exactly restored from the observation; we are here interested by the asymptotic case $\varepsilon \to 0$, and we look for a good approximation of the optimal filter

$$\hat{X}_t = (\hat{x}_t^{(1)}, \hat{x}_t^{(2)}) = E\big[X_t \mid y_s, 0 \le s \le t\big].$$

This approximation should be finite-dimensional (a solution of a finite-dimensional equation driven by y_t).

The same problem has been dealt with in [12] (with σ constant) by means of a formal asymptotic expansion of the optimal filter in a stationary situation. Our aim is to work out a rigorous mathematical study of the filter proposed by [12], namely the solution $M_t = (m_t^{(1)}, m_t^{(2)})$ of

$$(1.3) \qquad\qquad dM_t = f(M_t)dt + R_t[dy_t - h(m_t^{(1)})dt],$$

$$(1.4) \qquad\qquad R_t \stackrel{def}{=} \begin{bmatrix} \sqrt{\dfrac{2\sigma(M_t)F_{12}(M_t)}{h'(m_t^{(1)})\,\varepsilon}} \\[4mm] \dfrac{\sigma(M_t)}{\varepsilon} \end{bmatrix},$$

with $F_{12} = \partial f_1/\partial x_2$ and with initial condition $M_0 = E[X_0]$. This filter does in fact correspond to the extended Kalman filter with stationary gain if one neglects the contribution of the derivatives of f other than $\partial f_1/\partial x_2$. The stability of this filter is not evident and requires some assumptions. When it is stable, we prove in this work that

$$(1.5) \qquad x_t^{(1)} - m_t^{(1)} = \mathcal{O}(\varepsilon^{3/4}), \quad x_t^{(2)} - m_t^{(2)} = \mathcal{O}(\varepsilon^{1/4}),$$

and

$$(1.6) \qquad \hat{x}_t^{(1)} - m_t^{(1)} = \mathcal{O}(\varepsilon), \quad \hat{x}_t^{(2)} - m_t^{(2)} = \mathcal{O}(\sqrt{\varepsilon}).$$

We also verify that (1.6) can be improved when σ is constant, h is linear, and f_1 is linear with respect to x_2. (This case will be referred to as the almost linear case.) The proofs follow the method of [9].

The contents are organized as follows. In section 2, we introduce the assumptions which will be needed in what follows, and we study the filtering error as ε converges to zero; more precisely, we obtain the rate (1.5). In section 3, the error between the approximate filter and the optimal filter is studied, and we prove (1.6). Section 4 is devoted to the almost linear case. Results of numerical simulations that illustrate the performance of this approach are included in section 5.

Notation. The following notation is used:

$$f = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 0 \\ \sigma \end{bmatrix}, \quad H = \begin{bmatrix} h' & 0 \end{bmatrix};$$

$F = \begin{matrix} \breve{F}_{11} & \breve{F}_{12} \\ F_{21} & F_{22} \end{matrix}$ and $\Sigma' = \begin{matrix} \breve{0} & \breve{0} \\ \Sigma'_{21} & \Sigma'_{22} \end{matrix}$ are the Jacobian matrices of f and Σ; $\nabla_0 \Phi = \frac{\partial \Phi}{\partial X_0}$ is either a 2×2 matrix (if Φ is $\mathbf{R}^2$-valued) or a line-vector (if Φ is real-valued); see section 3. The symbol $*$ is used for the transposition of matrices.

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

DEFINITION 1.1. *Consider a real- or vector-valued stochastic process $\{\xi_t\}$. If β is real and $p \geq 1$, we will write that*

$$\xi_t = \mathcal{O}(\varepsilon^\beta) \quad in \quad L^p$$

when, for some $q \geq 0$, $\alpha > 0$, and some positive constants C_1, C_2, c_3,

$$E\big[\|\xi_t\|^p\big]^{1/p} \leq \frac{C_1}{\varepsilon^q} e^{-c_3 t/\varepsilon^\alpha} + C_2 \varepsilon^\beta$$

for $t \geq 0$ and ε small. In this situation, the process $\{\xi_t\}$ is usually said to converge to zero with rate of order ε^β, in a time scale of order ε^α.

2. Estimation of $X_t - M_t$. The following assumptions will be used throughout this article. The last one depends on a parameter $\delta \geq 1$.

(H1) X_0 is a random variable, the moments of which are finite.

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

(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

$$\frac{1}{\delta} \leq \sigma(x) \leq \delta, \quad \frac{1}{\delta} \leq h'(x_1) \leq \delta, \quad \frac{1}{\delta} \leq F_{12}(x) \leq \delta$$

for any $x = (x_1, x_2)$.

Remark 2.1. In order to reduce the notation in (H6.δ), system (1.1)–(1.2) has been rescaled. Indeed, if we assume instead that one has

$$\frac{1}{\delta} \leq \frac{\sigma(x)}{\bar{\sigma}} \leq \delta, \quad \frac{1}{\delta} \leq \frac{h'(x_1)}{\bar{H}} \leq \delta, \quad \frac{1}{\delta} \leq \frac{F_{12}(x)}{\bar{F}} \leq \delta$$

for any $x = (x_1, x_2)$ and for some positive $\bar{\sigma}$, $\bar{H}$, and $\bar{F}$ and if we replace the processes $x_t^{(1)}$, $x_t^{(2)}$, and y_t by $x_t^{(1)}/(\bar{\sigma}\bar{F})$, $x_t^{(2)}/\bar{\sigma}$, and $y_t/(\bar{\sigma}\bar{F}\bar{H})$, then the functions f_1, f_2, σ, and h are replaced, respectively, by

$$f_1(\bar{\sigma}\bar{F}x_1, \bar{\sigma}x_2) \,/\, (\bar{\sigma}\bar{F}), \quad f_2(\bar{\sigma}\bar{F}x_1, \bar{\sigma}x_2) \,/\, \bar{\sigma},$$

$$\sigma(\bar{\sigma}\bar{F}x_1, \bar{\sigma}x_2) \,/\, \bar{\sigma}, \quad h(\bar{\sigma}\bar{F}x) \,/\, (\bar{\sigma}\bar{F}\bar{H}),$$

and ε is replaced by $\varepsilon/(\bar{\sigma}\bar{F}\bar{H})$. We can apply the filter (1.3) to this new system, and we obtain $m_t^{(1)}/(\bar{\sigma}\bar{F})$ and $m_t^{(2)}/\bar{\sigma}$. This shows that the problem can be reduced to the case $\bar{\sigma} = \bar{F} = \bar{H} = 1$.

Assumption (H6.δ) says that the system does not contain too much nonlinearity; when it is not satisfied, there may be a small positive probability for the filter to lose the signal (see [10] for a similar problem). This is a rather restrictive condition, so we discuss at the end of the section the general case in which it does not hold.

We consider the system (1.1)–(1.2) and the filter (1.3). We let $\mathcal{F}_t$ be the filtration generated by $(X_0, w_t, \bar{w}_t)$ and $\mathcal{Y}_t$ the filtration generated by (y_t).

THEOREM 2.1. *Assume* (H1)–(H5). *For* $1 < \delta < 2^{1/5}$, *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* $p \geq 1$.

Consider a change of basis defined by a matrix T and its inverse T^{-1}, where

$$T \overset{def}{=} \begin{bmatrix} \sqrt{2/\varepsilon} & -1 \\ 0 & 1 \end{bmatrix}, \quad T^{-1} = \begin{bmatrix} \sqrt{\varepsilon/2} & \sqrt{\varepsilon/2} \\ 0 & 1 \end{bmatrix}.$$

Then consider the process

$$(2.1) \qquad Z_t \overset{def}{=} T(X_t - M_t).$$

We are going to check that Z_t is the solution of a linear stochastic differential equation; the study of the exponential stability of this equation will enable the estimation of both components of Z_t, and the theorem will immediately follow.

An equation for Z_t. From (1.1)–(1.3), we have

$$d(X_t - M_t) = \big(f(X_t) - f(M_t)\big)dt - R_t\big(h(x_t^{(1)}) - h(m_t^{(1)})\big)dt$$

$$+ \begin{bmatrix} 0 & -\sqrt{\dfrac{2\varepsilon\,\sigma(M_t)F_{12}(M_t)}{h'(m_t^{(1)})}} \\ \sigma(X_t) & -\sigma(M_t) \end{bmatrix} \begin{bmatrix} dw_t \\ d\bar{w}_t \end{bmatrix}.$$

In this equation, we introduce the Taylor expansions for the functions f and h,

$$f(X_t) - f(M_t) = F(\xi_t, \mu_t)(X_t - M_t)$$

and

$$h(x_t^{(1)}) - h(m_t^{(1)}) = h'(\eta_t)(x_t^{(1)} - m_t^{(1)}),$$

where $\{\xi_t\}$, $\{\mu_t\}$, and $\{\eta_t\}$ are $\mathbf{R}^2$- and $\mathbf{R}$-valued processes depending on $\{X_t\}$ and $\{M_t\}$, and

$$F(\xi_t, \mu_t) \stackrel{def}{=} \begin{bmatrix} F_{11}(\xi_t) & F_{12}(\xi_t) \\ F_{21}(\mu_t) & F_{22}(\mu_t) \end{bmatrix}.$$

We obtain a linear equation for $X_t - M_t$. By applying the transformation (2.1), we deduce for Z_t an equation of the type

$$(2.2) \qquad dZ_t = A_t\, Z_t\, dt + U_t \begin{bmatrix} dw_t \\ d\bar{w}_t \end{bmatrix}.$$

The precise computation shows that

$$A_t = T\big(F(\xi_t, \mu_t) - R_t H(\eta_t)\big) T^{-1} = \frac{\bar{A}_t}{\sqrt{2\varepsilon}} + \widetilde{A}_t\,,$$

with

$$\bar{A}_t^{(11)} = -2h'(\eta_t)\sqrt{\frac{F_{12}(M_t)\sigma(M_t)}{h'(m_t^{(1)})}} + h'(\eta_t)\sigma(M_t), \quad \bar{A}_t^{(12)} = \bar{A}_t^{(11)} + 2F_{12}(\xi_t),$$

$$\bar{A}_t^{(21)} = \bar{A}_t^{(22)} = -h'(\eta_t)\sigma(M_t),$$

and where $\widetilde{A}_t$ is a 2×2 matrix-valued process which is uniformly bounded as ε converges to 0; similarly, the matrix-valued process U_t is also uniformly bounded.

Stability of A_t. If $\delta = 1$, then $h' = F_{12} = \sigma = 1$, so $\bar{A}_t$ is the constant matrix

$$\bar{A}_t = \begin{bmatrix} -1 & 1 \\ -1 & -1 \end{bmatrix},$$

and

$$\bar{A}_t + \bar{A}_t^* = -2\,I.$$

In the general case $\delta > 1$, the coefficients of $\bar{A}_t + \bar{A}_t^*$ can be controlled so that this matrix is uniformly close to $-2I$ if δ is close to 1; in particular, for $1 < \delta < 2^{1/5}$, there exists $0 < \alpha < \alpha' < \sqrt{2}$ such that

$$\bar{A}_t + \bar{A}_t^* \le -\alpha'\sqrt{2}\,I$$

and, therefore,

$$(2.3) \qquad A_t + A_t^* \le -\frac{\alpha}{\sqrt{\varepsilon}}I$$

if ε is small.

End of the proof of Theorem 2.1. Our goal is now to deduce an estimate of Z_t in L^{2p} for the p integer. From Itô's formula and (2.2), the process $\|Z_t\|^2 = Z_t^* Z_t$ is the solution of

$$d\|Z_t\|^2 = Z_t^*(A_t + A_t^*)Z_t\, dt + \mathrm{trace}(U_t^* U_t)\, dt + 2\, Z_t^* U_t \begin{bmatrix} dw_t \\ d\bar{w}_t \end{bmatrix}.$$

We deduce that the moment of order p of $\|Z_t\|^2$ is finite and that

$$\frac{d}{dt}E[\|Z_t\|^{2p}] = p\,E[\|Z_t\|^{2p-2}Z_t^*(A_t + A_t^*)Z_t] + p\,E[\|Z_t\|^{2p-2}\mathrm{trace}(U_t^*U_t)]$$

$$+2\,p\,(p-1)\,E[\|Z_t\|^{2p-4}\,\|U_t^*Z_t\|^2]\,.$$

From (2.3), one has

$$Z_t^*(A_t + A_t^*)Z_t \le -\frac{\alpha}{\sqrt{\varepsilon}}\,\|Z_t\|^2\,.$$

As a consequence of the Cauchy–Schwarz inequality, one has

$$\|U_t^*Z_t\|^2 \le \mathrm{trace}(U_t^*U_t)\,\|Z_t\|^2\,.$$

Thus we obtain the inequality

$$\frac{d}{dt}E[\|Z_t\|^{2p}] \le -p\,\frac{\alpha}{\sqrt{\varepsilon}}E[\|Z_t\|^{2p}] + p\,(2p-1)E[\|Z_t\|^{2p-2}\mathrm{trace}(U_t^*U_t)]$$

$$\le -p\,\frac{\alpha}{\sqrt{\varepsilon}}E[\|Z_t\|^{2p}] + C_pE[\|Z_t\|^{2p-2}]\,.$$

Moreover, there exists C_p' such that

$$C_p\|Z_t\|^{2p-2} \le p\,\frac{\alpha}{2\sqrt{\varepsilon}}\|Z_t\|^{2p} + C_p'\varepsilon^{(p-1)/2},$$

and so

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

By solving this differential inequality, one obtains that, for some $C_p'' > 0$,

$$(2.4) \qquad E[\|Z_t\|^{2p}] \le C_p''\varepsilon^{p/2} + C_p''\,E[\|Z_0\|^{2p}]\,e^{-\alpha\,p\,t/(2\sqrt{\varepsilon})}\,.$$

Thus Z_t is $\mathcal{O}(\varepsilon^{1/4})$, and the order of magnitude of the components of $X_t - M_t$ follows from (2.1) and the form of T^{-1}.   □

We remark in (2.4) that the time scale of the estimation is of order $\sqrt{\varepsilon}$; one can compare it with the time scale ε obtained when the observation function is injective (see, for instance, [7]). This means that here it takes more time to estimate the signal, and this is not surprising since the second component of the signal is not well observed. There are also other systems where the time scale is not the same for the different components of the signal (see [10]).

In Theorem 2.1, we need the assumption (H6.δ), which is a restriction to the nonlinearity of the system; otherwise, it is difficult to ensure that the filter does not lose the signal. (This problem also occurs in [10].) Actually, we have chosen the filter (1.3) because it gives a good approximation of $\hat{X}_t$ (see the next section), but it is not the most stable one. If in (1.4) we replace the processes $\sigma(M_t)$, $F_{12}(M_t)$, and $h'(m_t^{(1)})$ by constant numbers $\bar{\sigma}$, $\bar{F}$, and $\bar{H}$, then we obtain a filter with constant gain; we can again work out the previous estimations and prove that the result of Theorem 2.1 holds for this filter without (H6.δ) as soon as

$$\frac{\max F_{12}}{\bar{F}} < 2\frac{\min h'}{\bar{H}}\,.$$

Thus we have two filters—a filter which is stable and tracks the signal under rather weak assumptions and the filter (1.3) which seems more fragile but gives (under good stability assumptions) a better approximation of the optimal filter.

3. Estimation of $\hat{X}_t - M_t$. The main result contained in this section is Theorem 3.1, which states the rate of convergence of the approximate filter considered in this paper toward the optimal filter. In order to give a proof of this theorem, a sequence of steps is needed: a change of probability measure, the differentiation with respect to the initial condition, and an integration by parts formula. A similar method of proof is adopted in [9]. As in Theorem 2.1, we may have a problem of stability in the general nonlinear case.

THEOREM 3.1. *Consider a finite time interval $[0, \tau]$. Assume (H1)–(H6.δ) and the following:*

  (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.

If δ in (H6.δ) is close enough to 1, in the sense that $1 < \delta < 2^{2/9}$, then the filter M_t given by (1.3) satisfies

$$\hat{x}_t^{(1)} - m_t^{(1)} = \mathcal{O}(\varepsilon), \qquad \hat{x}_t^{(2)} - m_t^{(2)} = \mathcal{O}(\sqrt{\varepsilon})$$

in L^2.

The rest of this section is devoted to the proof of this theorem.

Consider the matrix

$$P_t \stackrel{def}{=} \begin{bmatrix} \dfrac{1}{h'(m_t^{(1)})}\sqrt{\dfrac{2\sigma(M_t)F_{12}(M_t)}{h'(m_t^{(1)})}}\,\varepsilon^{3/2} & \dfrac{\sigma(M_t)}{h'(m_t^{(1)})}\varepsilon \\[2em] \dfrac{\sigma(M_t)}{h'(m_t^{(1)})}\varepsilon & \sigma(M_t)\sqrt{\dfrac{2\sigma(M_t)}{h'(m_t^{(1)})F_{12}(M_t)}}\,\varepsilon^{1/2} \end{bmatrix},$$

which depends only on M_t. Notice that P_t is the solution of the stationary Riccati equation

$$(3.1) \qquad -\frac{1}{\varepsilon^2}P_t H^*(M_t)H(M_t)P_t + \widetilde{F}(M_t)P_t + P_t\widetilde{F}^*(M_t) + \Sigma(M_t)\Sigma^*(M_t) = 0$$

with

$$\widetilde{F}(M_t) = \begin{bmatrix} 0 & F_{12}(M_t) \\ 0 & 0 \end{bmatrix}$$

and that the process R_t of (1.4) is

$$(3.2) \qquad\qquad\qquad R_t = \frac{P_t}{\varepsilon^2}H^*(M_t).$$

We will also need the inverse of P_t, namely,

$$P_t^{-1} = \begin{bmatrix} h'(m_t^{(1)})\sqrt{\dfrac{2h'(m_t^{(1)})}{\sigma(M_t)F_{12}(M_t)}}\,\varepsilon^{-3/2} & -\dfrac{h'(m_t^{(1)})}{\sigma(M_t)}\varepsilon^{-1} \\[2em] -\dfrac{h'(m_t^{(1)})}{\sigma(M_t)}\varepsilon^{-1} & \dfrac{1}{\sigma(M_t)}\sqrt{\dfrac{2h'(m_t^{(1)})F_{12}(M_t)}{\sigma(M_t)}}\,\varepsilon^{-1/2} \end{bmatrix}.$$

Change of probability measure. Our random variables can be viewed as functions of the initial condition X_0 and of the Wiener processes w and $\bar{w}$. We are going to

make a change of variables; in view of the Girsanov theorem, this can be viewed as a change of probability measure; however, all the estimations will be made under the original probability P. Thus consider the new probability measure which is given on $\mathcal{F}_t$ by

$$\left.\frac{d\dot{P}}{dP}\right|_{\mathcal{F}_t} = L_t^{-1},$$

where

$$L_t^{-1} = \exp\left\{-\frac{1}{\varepsilon}\int_0^t h(x_s^{(1)})\,d\bar{w}_s - \frac{1}{2\varepsilon^2}\int_0^t h^2(x_s^{(1)})\,ds\right\}.$$

The probability $\dot{P}$ is the so-called reference probability, and one checks easily from the Girsanov theorem that y_t/ε and w_t are standard independent Wiener processes under $\dot{P}$. Let us define now the probability measure $\widetilde{P}$ on $\mathcal{F}_t$ by

$$\left.\frac{d\widetilde{P}}{d\dot{P}}\right|_{\mathcal{F}_t} = \Lambda_t^{-1},$$

where

$$\Lambda_t^{-1} = \exp\left\{\int_0^t \Sigma^*(M_s)P_s^{-1}(X_s - M_s)\,dw_s - \frac{1}{2}\int_0^t (\Sigma^*(M_s)P_s^{-1}(X_s - M_s))^2\,ds\right\}.$$

Then the processes

$$\widetilde{w}_t = w_t - \int_0^t \Sigma^*(M_s)P_s^{-1}(X_s - M_s)\,ds$$

and y_t/ε are standard independent Wiener processes under $\widetilde{P}$. On the other hand, one has

$$(3.3) \qquad dX_t = f(X_t)\,dt + \Sigma(X_t)\Sigma^*(M_t)P_t^{-1}(X_t - M_t)\,dt + \Sigma(X_t)\,d\widetilde{w}_t$$

and

$$
\begin{aligned}
\log(L_t\Lambda_t) = {}& \frac{1}{\varepsilon^2}\int_0^t h(x_s^{(1)})\,dy_s - \frac{1}{2\varepsilon^2}\int_0^t h^2(x_s^{(1)})\,ds - \int_0^t \Sigma^*(M_s)P_s^{-1}(X_s - M_s)\,d\widetilde{w}_s \\
& - \frac{1}{2}\int_0^t (\Sigma^*(M_s)P_s^{-1}(X_s - M_s))^2\,ds.
\end{aligned}
$$
(3.4)

Differentiation with respect to the initial condition and an estimation. The random variables involved in our computation can now be viewed as functions of X_0, $\{\widetilde{w}_t\}$, and $\{y_t\}$; let us denote by ∇_0 the differentiation with respect to the initial condition X_0 (computed in L^p). In particular, we can see on (3.3) and (3.4) that the processes X_t and $\log(L_t\Lambda_t)$ are differentiable, and we obtain matrix- and vector-valued processes, respectively. Our aim is to estimate the process

$$(3.5) \qquad V_t \stackrel{def}{=} (\nabla_0\log(L_t\Lambda_t)(\nabla_0 X_t)^{-1} + (X_t - M_t)^*P_t^{-1})U$$

with

$$U \overset{\text{def}}{=} \begin{bmatrix} 1 & 1 \\ 0 & \sqrt{2/\varepsilon} \end{bmatrix}, \qquad U^{-1} = \begin{bmatrix} 1 & -\sqrt{\varepsilon/2} \\ 0 & \sqrt{\varepsilon/2} \end{bmatrix}.$$

Then an integration by parts will enable us to conclude.

By applying the operator ∇_0 to (3.4), one gets

$$
\begin{aligned}
\nabla_0 \log(L_t \Lambda_t) &= \frac{1}{\varepsilon^2} \int_0^t h'(x_s^{(1)}) \nabla_0 x_s^{(1)} \, (dy_s - h(x_s^{(1)})ds) \\
&\quad - \int_0^t \Sigma^*(M_s) P_s^{-1} \nabla_0 X_s \, (d\widetilde{w}_s + \Sigma^*(M_s) P_s^{-1}(X_s - M_s)ds) \\
&= \frac{1}{\varepsilon} \int_0^t h'(x_s^{(1)}) \nabla_0 x_s^{(1)} \, d\bar{w}_s - \int_0^t \Sigma^*(M_s) P_s^{-1} \nabla_0 X_s \, dw_s \, .
\end{aligned}
$$
(3.6)

We can also differentiate (3.3), and, if Σ' is the Jacobian matrix of Σ, we obtain

$$d(\nabla_0 X_t) = [F(X_t) + \Sigma(X_t)\Sigma^*(M_t) P_t^{-1}] \nabla_0 X_t \, dt + \Sigma'(X_t) \nabla_0 X_t \, dw_t.$$

The matrix $\nabla_0 X_t$ is invertible, and Itô's calculus shows that

$$
\begin{aligned}
d(\nabla_0 X_t)^{-1} &= -(\nabla_0 X_t)^{-1}[F(X_t) + \Sigma(X_t)\Sigma^*(M_t) P_t^{-1} - {\Sigma'}^2(X_t)] \, dt \\
&\quad - (\nabla_0 X_t)^{-1} \Sigma'(X_t) dw_t \, .
\end{aligned}
$$
(3.7)

From this equation and (3.6), one can write that

$$
\begin{aligned}
d\left(\nabla_0 \log(L_t \Lambda_t)(\nabla_0 X_t)^{-1}\right) &= \frac{1}{\varepsilon} H(X_t) \, d\bar{w}_t - \Sigma^*(M_t) P_t^{-1} \, dw_t \\
&\quad - \nabla_0 \log(L_t \Lambda_t)(\nabla_0 X_t)^{-1} \Sigma'(X_t) \, dw_t \\
&\quad - \nabla_0 \log(L_t \Lambda_t)(\nabla_0 X_t)^{-1} \\
&\qquad . [F(X_t) + \Sigma(X_t)\Sigma^*(M_t) P_t^{-1} - {\Sigma'}^2(X_t)] dt \\
&\quad + \Sigma^*(M_t) P_t^{-1} \Sigma'(X_t) dt
\end{aligned}
$$
(3.8)

since one has $h'(x_t^{(1)}) \nabla_0 x_t^{(1)} (\nabla_0 X_t)^{-1} = H(X_t)$.

On the other hand, from the equations of X_t and M_t ((1.1) and (1.3), respectively), one has

$$d(X_t - M_t) = [f(X_t) - f(M_t)] \, dt - R_t[h(x_t^{(1)}) - h(m_t^{(1)})] \, dt - R_t \varepsilon \, d\bar{w}_t + \Sigma(X_t) \, dw_t \, .$$

By writing the differential of P_t^{-1} in the form

$$dP_t^{-1} = J_t^{(1)} \, dt + J_t^{(2)} \, d\bar{w}_t,$$

we obtain

$$
\begin{aligned}
d\left((X_t - M_t)^* P_t^{-1}\right) &= [f^*(X_t) - f^*(M_t) - R_t^*(h(x_t^{(1)}) - h(m_t^{(1)}))] P_t^{-1} \, dt \\
&\quad + \Sigma^*(X_t) P_t^{-1} \, dw_t - \varepsilon \, R_t^* P_t^{-1} \, d\bar{w}_t \\
&\quad + (X_t - M_t)^* [J_t^{(1)} \, dt + J_t^{(2)} \, d\bar{w}_t] - \varepsilon \, R_t^* J_t^{(2)} \, dt \, .
\end{aligned}
$$
(3.9)

One can write the Taylor expansions for f and h,

$$f(X_t) - f(M_t) = F(M_t)(X_t - M_t) + \phi_t,$$
$$h(x_t^{(1)}) - h(m_t^{(1)}) = H(M_t)(X_t - M_t) + \gamma_t,$$

with

$$\|\phi_t\| \leq C\|X_t - M_t\|^2, \quad |\gamma_t| \leq C|x_t^{(1)} - m_t^{(1)}|^2.$$

By using these expansions together with the consequence of (3.2),

$$H^*(M_t)R_t^*P_t^{-1} = \frac{1}{\varepsilon^2}H^*(M_t)H(M_t),$$

in (3.9), we obtain

$$d((X_t - M_t)^*P_t^{-1}) = (X_t - M_t)^*\left(F^*(M_t)P_t^{-1} - \frac{1}{\varepsilon^2}H^*(M_t)H(M_t)\right)dt$$
$$+\Sigma^*(X_t)P_t^{-1}dw_t - \varepsilon R_t^*P_t^{-1}d\bar{w}_t$$
$$+(X_t - M_t)^*[J_t^{(1)}dt + J_t^{(2)}d\bar{w}_t] - \varepsilon R_t^*J_t^{(2)}dt$$
$$+(\phi_t^* - \gamma_t R_t^*)P_t^{-1}dt.$$

By adding this equation to (3.8), we obtain that the process V_t of (3.5) satisfies

$$d(V_tU^{-1}) = -V_tU^{-1}[F(X_t) + \Sigma(X_t)\Sigma^*(M_t)P_t^{-1} - \Sigma'^2(X_t)]dt - V_tU^{-1}\Sigma'(X_t)dw_t$$
$$+\frac{1}{\varepsilon}H(X_t)d\bar{w}_t - \Sigma^*(M_t)P_t^{-1}dw_t + \Sigma^*(M_t)P_t^{-1}\Sigma'(X_t)dt$$
$$(3.10) \qquad +(X_t - M_t)^*S_tdt + (X_t - M_t)^*P_t^{-1}\Sigma'(X_t)dw_t + \Sigma^*(X_t)P_t^{-1}dw_t$$
$$-\varepsilon R_t^*P_t^{-1}d\bar{w}_t + (X_t - M_t)^*[J_t^{(1)}dt + J_t^{(2)}d\bar{w}_t] - \varepsilon R_t^*J_t^{(2)}dt$$
$$+(\phi_t^* - \gamma_t R_t^*)P_t^{-1}dt,$$

where S_t is the matrix given by

$$(3.11) \qquad \begin{aligned} S_t &\stackrel{def}{=} -\frac{1}{\varepsilon^2}H^*(M_t)H(M_t) + F^*(M_t)P_t^{-1} + P_t^{-1}F(X_t) \\ &\quad +P_t^{-1}\Sigma(X_t)\Sigma^*(M_t)P_t^{-1} - P_t^{-1}\Sigma'^2(X_t). \end{aligned}$$

Consider also the matrix-valued process

$$(3.12) \qquad A_t \stackrel{def}{=} -U^{-1}[F(X_t) + \Sigma(X_t)\Sigma^*(M_t)P_t^{-1} - \Sigma'^2(X_t)]U.$$

Then (3.10) can be written in the form

$$(3.13) \qquad \begin{aligned} &dV_t = V_tA_tdt - V_tU^{-1}\Sigma'(X_t)Udw_t + J_t^{(3)}dt + J_t^{(4)}dw_t + J_t^{(5)}d\bar{w}_t, \\ &V_0 = (X_0 - M_0)^*P_0^{-1}U, \end{aligned}$$

where

$$\begin{aligned} J_t^{(3)} &= \Sigma^*(M_t)P_t^{-1}\Sigma'(X_t)U + (X_t - M_t)^*S_tU + (X_t - M_t)^*J_t^{(1)}U \\ &\quad -\varepsilon R_t^*J_t^{(2)}U + (\phi_t^* - \gamma_t R_t^*)P_t^{-1}U, \\ J_t^{(4)} &= (\Sigma^*(X_t) - \Sigma^*(M_t))P_t^{-1}U + (X_t - M_t)^*P_t^{-1}\Sigma'(X_t)U, \\ J_t^{(5)} &= \frac{1}{\varepsilon}H(X_t)U - \varepsilon R_t^*P_t^{-1}U + (X_t - M_t)^*J_t^{(2)}U \\ &= \frac{1}{\varepsilon}(H(X_t) - H(M_t))U + (X_t - M_t)^*J_t^{(2)}U \end{aligned}$$

(apply (3.2) for the last line). We deduce that $E[\|V_0\|^2]$ is of order ε^{-3} and that

$$\text{(3.14)} \quad \begin{aligned} \frac{d}{dt}E[\|V_t\|^2] &= E[V_t(A_t + A_t^*)V_t^*] + 2E[J_t^{(3)}V_t^*] \\ &\quad + E[\|V_tU^{-1}\Sigma'(X_t)U + J_t^{(4)}\|^2] + E[\|J_t^{(5)}\|^2]. \end{aligned}$$

We have to estimate the terms of the right-hand side.

By computing the matrix A_t, we obtain that

$$A_t = \frac{\bar{A}_t}{\sqrt{2\varepsilon}} + \widetilde{A}_t$$

with

$$\bar{A}_t^{(11)} = -\bar{A}_t^{(21)} = -h'(m_t^{(1)})\sigma(X_t),$$

$$\bar{A}_t^{(12)} = -2\,F_{12}(X_t) - h'(m_t^{(1)})\sigma(X_t) + 2\,\sigma(X_t)\sqrt{\frac{h'(m_t^{(1)})F_{12}(M_t)}{\sigma(M_t)}},$$

$$\bar{A}_t^{(22)} = h'(m_t^{(1)})\sigma(X_t) - 2\,\sigma(X_t)\sqrt{\frac{h'(m_t^{(1)})F_{12}(M_t)}{\sigma(M_t)}},$$

and $\widetilde{A}_t$ is uniformly bounded. As in the proof of Theorem 2.1, we see that, if $\delta = 1$, then the matrix $\bar{A}_t$ is simply

$$\bar{A}_t = \begin{bmatrix} -1 & -1 \\ 1 & -1 \end{bmatrix},$$

which satisfies

$$\bar{A}_t + \bar{A}_t^* = -2\,I.$$

Thus, for $0 < \alpha < \sqrt{2}$, when δ is close enough to 1, that is, $1 < \delta < 2^{2/9}$, and when ε is small enough, we have

$$\text{(3.15)} \quad A_t + A_t^* \le -\frac{\alpha}{\sqrt{\varepsilon}}I.$$

We also notice that

$$2J_t^{(3)}V_t^* \le \frac{\alpha}{3\sqrt{\varepsilon}}\|V_t\|^2 + C\sqrt{\varepsilon}\|J_t^{(3)}\|^2$$

and that

$$\|V_tU^{-1}\Sigma'(X_t)U + J_t^{(4)}\|^2 \le C\|V_t\|^2 + 2\|J_t^{(4)}\|^2$$

because $U^{-1}\Sigma'(X_t)U$ is bounded. Thus (3.14) implies that, for small ε,

$$\text{(3.16)} \quad \begin{aligned} \frac{d}{dt}E[\|V_t\|^2] &\le -\frac{\alpha}{3\sqrt{\varepsilon}}E[\|V_t\|^2] + C\sqrt{\varepsilon}\,E[\|J_t^{(3)}\|^2] \\ &\quad + 2\,E[\|J_t^{(4)}\|^2] + E[\|J_t^{(5)}\|^2]. \end{aligned}$$

Let us first estimate $J_t^{(3)}$. We deduce from the Riccati equation (3.1) satisfied by P_t that the process S_t defined in (3.11) satisfies

$$S_t = \big(F^*(M_t) - \widetilde{F}^*(M_t)\big)P_t^{-1} + P_t^{-1}\big(F(X_t) - \widetilde{F}(M_t)\big)$$
$$+ P_t^{-1}\big(\Sigma(X_t) - \Sigma(M_t)\big)\Sigma^*(M_t)P_t^{-1} - P_t^{-1}\Sigma'^2(X_t).$$

By computing this matrix and applying Theorem 2.1, we check that

$$S_t = \left[\begin{array}{cc} \mathcal{O}(\varepsilon^{-7/4}) & \mathcal{O}(\varepsilon^{-5/4}) \\ \mathcal{O}(\varepsilon^{-5/4}) & \mathcal{O}(\varepsilon^{-3/4}) \end{array}\right]$$

in the spaces L^p. Thus

$$(X_t - M_t)^* S_t U = \mathcal{O}(\varepsilon^{-1}).$$

The term $\Sigma^*(M_t)P_t^{-1}\Sigma'(X_t)U$ is easily shown to have the same order of magnitude. On the other hand, by looking at the equation of M_t and by applying Itô's formula, we can prove that, for any C^2 function ρ with bounded derivatives, one has

$$d\rho(M_t) = \mathcal{O}(\varepsilon^{-1/4})dt + \mathcal{O}(1)d\bar{w}_t.$$

By applying this result to the functions involved in P_t^{-1}, it appears that

$$J_t^{(1)} = \left[\begin{array}{cc} \mathcal{O}(\varepsilon^{-7/4}) & \mathcal{O}(\varepsilon^{-5/4}) \\ \mathcal{O}(\varepsilon^{-5/4}) & \mathcal{O}(\varepsilon^{-3/4}) \end{array}\right], \quad J_t^{(2)} = \left[\begin{array}{cc} \mathcal{O}(\varepsilon^{-3/2}) & \mathcal{O}(\varepsilon^{-1}) \\ \mathcal{O}(\varepsilon^{-1}) & \mathcal{O}(\varepsilon^{-1/2}) \end{array}\right].$$

We deduce that the terms of $J_t^{(3)}$ involving $J_t^{(1)}$ and $J_t^{(2)}$ are also of order ε^{-1}. Finally, ϕ_t and γ_t are, respectively, of order $\varepsilon^{1/2}$ and $\varepsilon^{3/2}$, and so the last term is of order ε^{-1}, and we deduce that

$$J_t^{(3)} = \mathcal{O}(\varepsilon^{-1}).$$

We can also estimate $J_t^{(4)}$ and $J_t^{(5)}$ and check that they are of order $\varepsilon^{-3/4}$. Thus (3.16) enables us to conclude that

$$V_t = \mathcal{O}(1/\sqrt{\varepsilon})$$

in L^2. We can take the conditional expectation with respect to $\mathcal{Y}_t$ in this estimation because the conditional expectation is a contraction in L^2; thus $E[V_t|\mathcal{Y}_t]$ is $\mathcal{O}(1/\sqrt{\varepsilon})$ in L^2, and, therefore, we obtain from the definition (3.5) that

$$(3.17) \qquad (\hat{X}_t - M_t)^* P_t^{-1} U = -E[\nabla_0 \log(L_t\Lambda_t)(\nabla_0 X_t)^{-1}U|\mathcal{Y}_t] + \mathcal{O}(1/\sqrt{\varepsilon}).$$

Application of an integration by parts formula. The estimation of the right-hand side of (3.17) can be completed by means of an integration by parts formula. It is proved in Lemma 3.4.2 of [9] that, if $G = G(X_0, \widetilde{w}, y)$ is a functional defined on the probability space which is differentiable with respect to the initial condition (in the spaces L^p) and if ∇_0^i is the differentiation with respect to the ith component of X_0, then

$$(3.18) \qquad E[G\nabla_0^i \log(L_t\Lambda_t) + G(p_0^{-1}\,\partial p_0/\partial x_i)(X_0) + \nabla_0^i G|\mathcal{Y}_t] = 0.$$

We can write (3.7) in the form

$$\begin{aligned}
d(\nabla_0 X_t)^{-1} = &-(\nabla_0 X_t)^{-1}(F(X_t) + \Sigma(X_t)\Sigma^*(M_t)P_t^{-1} - \Sigma'^2(X_t) \\
&+\Sigma'(X_t)(\Sigma^*(M_t)P_t^{-1}(X_t - M_t)))dt \\
&-(\nabla_0 X_t)^{-1}\Sigma'(X_t)d\widetilde{w}_t
\end{aligned}$$

(3.19)

with $(\nabla_0 X_0)^{-1} = I$. This equation can be differentiated with respect to X_0, and so we can apply the integration by parts formula (3.18) to the coefficients of the matrix $(\nabla_0 X_t)^{-1}$. Denote by $(\nabla_0 X_t)_i^{-1}$ its ith line. Then

$$E[(\nabla_0 X_t)_i^{-1}\nabla_0^i \log(L_t \Lambda_t) + (\nabla_0 X_t)_i^{-1}(p_0^{-1}\,\partial p_0/\partial x_i)(X_0) + \nabla_0^i(\nabla_0 X_t)_i^{-1}|\mathcal{Y}_t] = 0.$$

By summing on i and multiplying by U, we have

(3.20)

$$E\left[\nabla_0 \log(L_t \Lambda_t)(\nabla_0 X_t)^{-1}U + (p_0^{-1}p_0')(X_0)(\nabla_0 X_t)^{-1}U \right.$$

$$\left. + \sum_i \nabla_0^i(\nabla_0 X_t)_i^{-1}U \,\middle|\, \mathcal{Y}_t \right] = 0.$$

The first term of (3.20) is exactly the term that we want to estimate in (3.17).

For the second term of (3.20), if

$$\Psi_t \stackrel{def}{=} (p_0^{-1}p_0')(X_0)\,(\nabla_0 X_t)^{-1}U,$$

we have from (3.7) and (3.12) that

$$\Psi_0 = (p_0^{-1}p_0')(X_0)U, \quad d\Psi_t = \Psi_t A_t\,dt - \Psi_t U^{-1}\Sigma'(X_t)U\,dw_t.$$

We proceed as in the study of (3.13). The stability of the matrix A_t, which has been obtained in (3.15), and the boundedness of $U^{-1}\Sigma'(X_t)U$ imply that $(\nabla_0 X_t)^{-1}$ is exponentially small in L^2, and so the second term is negligible.

Let us study the third term of (3.20). If

$$\Phi_t^i = \nabla_0^i(\nabla_0 X_t)_i^{-1}U,$$

then by differentiating (3.19) and transforming $\widetilde{w}$ back into w, we get

$$\begin{aligned}
d\Phi_t^i = &\Phi_t^i A_t dt - \Phi_t^i U^{-1}\Sigma'(X_t)U\,dw_t - (\nabla_0 X_t)_i^{-1}\nabla_0^i\rho(X_t, M_t)U\,dt \\
&-\Sigma^*(M_t)P_t^{-1}\nabla_0^i X_t(\nabla_0 X_t)_i^{-1}\Sigma'(X_t)U\,dt - (\nabla_0 X_t)_i^{-1}\nabla_0^i\big(\Sigma'(X_t)\big)U\,dw_t
\end{aligned}$$

with

$$\rho(X_t, M_t) \stackrel{def}{=} F(X_t) + \Sigma(X_t)\Sigma^*(M_t)P_t^{-1} - \Sigma'^2(X_t).$$

By summing on i and using

$$\sum_i (\nabla_0 X_t)_i^{-1}\nabla_0^i\rho(X_t, M_t) = \sum_{i,j} \nabla_0^i X_t^j(\nabla_0 X_t)_i^{-1}\frac{\partial\rho}{\partial x_j}(X_t, M_t) = \sum_j \frac{\partial\rho_j}{\partial x_j}(X_t, M_t),$$

where ρ_j is the jth line of ρ, we obtain that $\Phi_t = \sum \Phi_t^i$ is the solution of

$$\Phi_0 = 0, \quad d\Phi_t = \Phi_t A_t \, dt - \Phi_t U^{-1} \Sigma'(X_t) U \, dw_t - \sum_j \frac{\partial \rho_j}{\partial x_j}(X_t, M_t) U \, dt$$

(3.21)

$$- \Sigma^*(M_t) P_t^{-1} \Sigma'(X_t) U \, dt - \frac{\partial \sigma'}{\partial x_2}(X_t) U \, dw_t,$$

where σ' is the Jacobian of σ. A computation shows that

$$\frac{\partial \rho_j}{\partial x_j}(X_t, M_t) = \begin{bmatrix} \mathcal{O}(1) & \mathcal{O}(1) \\ \mathcal{O}(\varepsilon^{-1}) & \mathcal{O}(\varepsilon^{-1/2}) \end{bmatrix}.$$

The multiplication on the right by U yields a process of order ε^{-1}; the term $\Sigma^*(M_t)$ $.P_t^{-1} \Sigma'(X_t) U$ is also $\mathcal{O}(\varepsilon^{-1})$, and the term involving the second derivative of σ is $\mathcal{O}(\varepsilon^{-1/2})$. By proceeding again as in the study of (3.13), we deduce that Φ_t is of order $\varepsilon^{-1/2}$.

Thus (3.17), (3.20), and the estimation of Ψ_t and Φ_t yield

$$(\hat{X}_t - M_t)^* P_t^{-1} U = \mathcal{O}(1/\sqrt{\varepsilon}).$$

We multiply on the right by the matrix $U^{-1} P_t$, the coefficients of which are of order $\varepsilon^{3/2}$ for the first column and ε for the second column, and we deduce the order of $\hat{X}_t - M_t$ which was claimed in the theorem.   □

4. An almost linear case. It is interesting to consider a particular case in which σ, h', and F_{12} are constant so that the system (1.1)–(1.2) is

(4.1)
$$\begin{cases} dx_t^{(1)} &= \left(f_1^0(x_t^{(1)}) + F_{12} x_t^{(2)}\right) dt, \\ dx_t^{(2)} &= f_2(x_t^{(1)}, x_t^{(2)}) \, dt + \sigma \, dw_t, \\ dy_t &= h' x_t^{(1)} \, dt + \varepsilon \, d\bar{w}_t. \end{cases}$$

In particular, (H6.δ) holds with $\delta = 1$. Then it is possible to improve the upper bounds given in Theorem 3.1. The time interval that we consider may be infinite. The result is stated in the following proposition.

PROPOSITION 4.1. *Assuming that* (H1)–(H7) *hold for* (4.1), *the filter* M_t *given by* (1.3) *verifies*

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

in L^2.

Proof. The proof closely follows the sequence of steps adopted in Theorem 3.1. The matrices $P_t = P$ and $R_t = R$ are now constant; the processes $J_t^{(1)}$, $J_t^{(2)}$, $J_t^{(4)}$, and $J_t^{(5)}$ are zero. The order of S_t is improved into

$$S_t = \begin{bmatrix} \mathcal{O}(\varepsilon^{-3/2}) & \mathcal{O}(\varepsilon^{-1}) \\ \mathcal{O}(\varepsilon^{-1}) & \mathcal{O}(\varepsilon^{-1/2}) \end{bmatrix},$$

and

$$|\phi_t^{(1)}| \leq C|x_t^{(1)} - m_t^{(1)}|^2 = \mathcal{O}(\varepsilon^{3/2}), \quad |\phi_t^{(2)}| \leq C\|X_t - M_t\|^2 = \mathcal{O}(\varepsilon^{1/2})$$

so that

$$J_t^{(3)} = (X_t - M_t)^* S_t U + \phi_t^* P^{-1} U$$

is of order $\varepsilon^{-3/4}$. Thus V_t is $\mathcal{O}(\varepsilon^{-1/4})$, and we obtain $\mathcal{O}(\varepsilon^{-1/4})$ in (3.17).

For the end of the proof, we see that

$$\rho(X_t, M_t) = F(X_t) + \Sigma\Sigma^* P^{-1},$$

and so

$$\frac{\partial \rho_j}{\partial x_j}(X_t, M_t) = \frac{\partial F_j}{\partial x_j}(X_t)$$

is bounded. Multiplication by U yields a process of order $\varepsilon^{-1/2}$, and so the process Φ_t of (3.21) is bounded for small ε. We can conclude that

$$(\hat{X}_t - M_t)^* P_t^{-1} U = \mathcal{O}(\varepsilon^{-1/4})$$

and deduce the proposition.   □

With more computational effort, it is possible to extend these results to the case in which the component $x^{(1)}$ is driven by low noise:

$$(4.2) \qquad \begin{cases} dx_t^{(1)} &= \left(f_1^0(x_t^{(1)}) + F_{12}x_t^{(2)}\right) dt + \varepsilon^\gamma \, dw_t^{(1)}, \\[2mm] dx_t^{(2)} &= f_2(x_t^{(1)}, x_t^{(2)}) \, dt + \sigma \, dw_t^{(2)}, \\[2mm] dy_t &= h' x_t^{(1)} \, dt + \varepsilon \, d\bar{w}_t \end{cases}$$

with M_t given by (1.5) and with the gain R_t given by (1.4), as before, if $\gamma > 1/2$, and with R_t given by

$$R_t \overset{def}{=} \left[\frac{\sqrt{\frac{2\sigma F_{12}}{h'} + 1} \; \frac{1}{\sqrt{\varepsilon}}}{\frac{\sigma}{\varepsilon}} \right]$$

if $\gamma = 1/2$.

Clearly, Theorem 2.1 extends to system (4.2) as soon as $\gamma \geq 1/2$. This results from the fact that, in the SDE of Z_t, the matrices involved in the martingale terms are still uniformly bounded as ε converges to 0, and the matrix A_t of (2.2) has the same stability property as before.

Regarding the extension of Proposition 4.1 to system (4.2), one can see that, assuming $\gamma \geq 3/4$, the estimation in Proposition 4.1 still holds. This happens because the matrix $\bar{A}_t$ in the decomposition of A_t remains the same. More effort is needed if one considers the cases $1/2 < \gamma < 3/4$ and $\gamma = 1/2$.

Another class of almost linear filtering problems when some of the observations and driving noises are small is considered by Krener [6]. Krener studied the multi-dimensional case, where nonlinearities depend only on state variables which can be estimated quickly and accurately; that is, the only nonlinearity allowed in (4.2) is that of the function f_2 with respect to $x_t^{(1)}$. Observations with at least two components, instead of one, are also assumed.
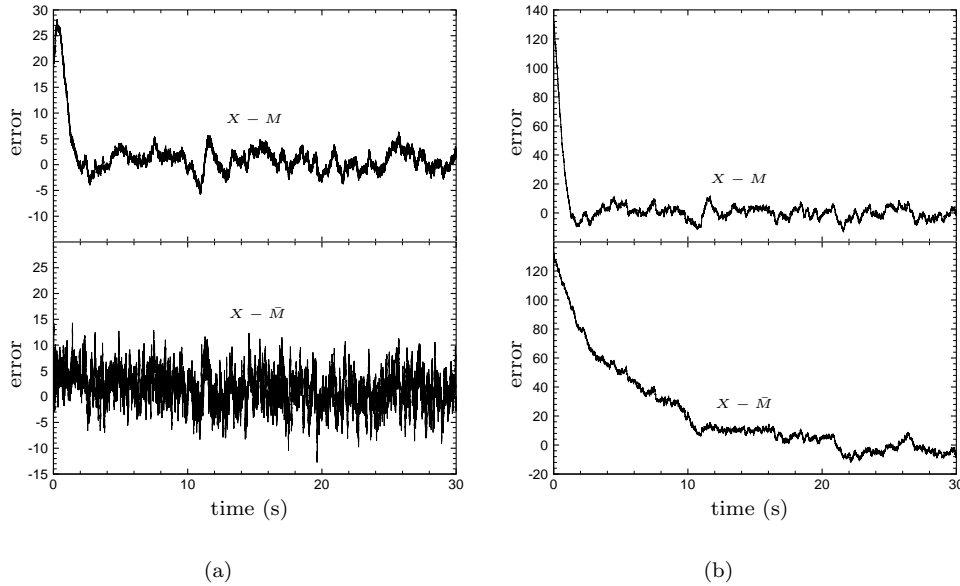
FIG. 5.1. *Estimation errors for the* (a) *first and* (b) *second components of X computed on a single trajectory.*

5. Numerical simulation results. Let us consider the following example illustrating the case of free fall of a body through the atmosphere:

$$\begin{cases} dx_t^{(1)} & = & x_t^{(2)}dt, \\ dx_t^{(2)} & = & (\rho_0 e^{-x_t^{(1)}/k}(x_t^{(2)})^2/(2\beta) - g)\,dt + \sigma dw_t \end{cases}$$

and

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

where $x_t^{(1)}$ is the position of the moving body and $x_t^{(2)}$ is its speed, ρ_0 being the reference air density, k the atmosphere thickness, β the ballistic coefficient of the body, g the acceleration due to gravity, and a the horizontal distance between the body and the measuring device ($\rho_0 = 3.4 \times 10^{-3}\,lb\,s^2/ft^4$, $k = 22 \times 10^3\,ft$, $\beta = 1.6 \times 10^3\,lb^2/ft^4$, $g = 32.2\,ft/s^2$, $\sigma = 5\,ft/s$, and $a = 10^4\,ft$). Figure 5.1 shows the estimation errors obtained from applying the two approximate filters (filter (1.5), noted M_t, and the constant gain filter mentioned at the end of section 2 with $\bar{H} = 0.02$, noted $\bar{M}_t$) to a single trajectory of the state with measurements taken each 0.001 s. The parameter ε is equal to 1 and

$$X_0 \sim \mathcal{N}\left(\begin{bmatrix} 3 \times 10^5 \\ -10^3 \end{bmatrix}, \begin{bmatrix} 900 & 0 \\ 0 & 2 \times 10^4 \end{bmatrix}\right).$$

It illustrates the fact that the errors get small very quickly, and one notices that the constant gain filter needs more time than filter (1.5) to attain small errors in the second component.
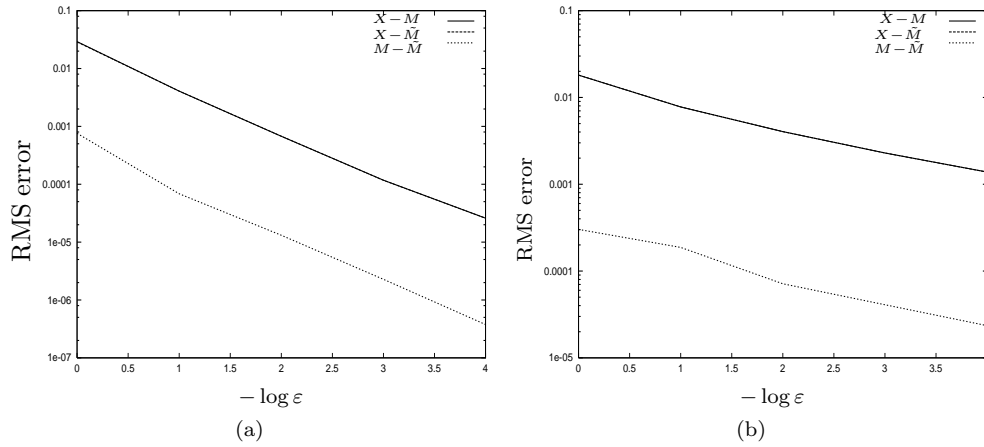
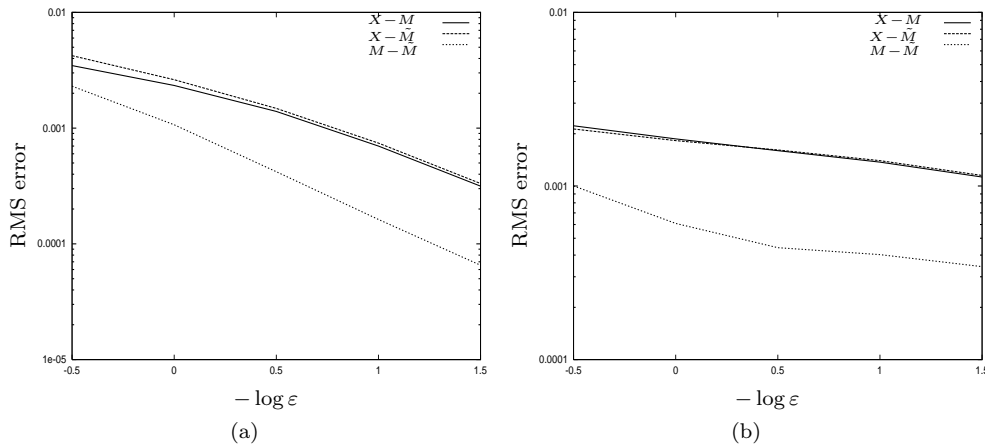FIG. 5.2. *Estimation errors for the* (a) *first and* (b) *second components of* X.

FIG. 5.3. *Estimation errors for the* (a) *first and* (b) *second components of* X ($G = 0.5$).

Figure 5.2 illustrates the asymptotic behavior of the estimation errors when system (1.1)–(1.2) with $f(x_1, x_2) = [x_2 \quad -1.5 \times 10^{-3}x_1^2]^*$, $\sigma = 2$, and $h(x_1) = \sqrt{x_1^2 + 10^8}$ is considered. Although f and h' fail to verify assumption (H3) and, in fact, $\inf h' = 0$, we will assume that the state remains in a bounded domain with high probability, thus assuming that $\inf h' > 1/\sqrt{120}$. The root mean square error between the two approximate filters (with $\bar{H} = 0.18$) was computed for $\varepsilon = 1, 10^{-1}, \dots, 10^{-4}$ over 200 simulations for both components in the time interval $[0, 5]$. The solid lines exhibit approximate slopes of -0.76 (first component) and -0.28 (second component) which agree with the results in section 2. The error associated with the constant gain filter and that associated to filter (1.5) are very similar.

Figures 5.3–5.5 illustrate the van der Pol oscillator example presented in [12, section 6]: $f(x_1, x_2) = [x_2 \quad -x_1 - x_2]^*$, $\sigma = 1$, and $h(x_1) = 0.606(1 - G)x_1 +$
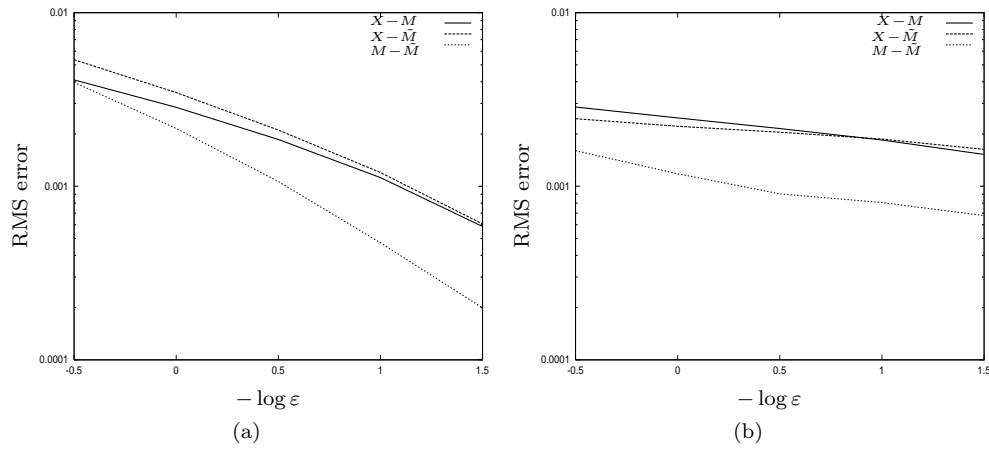
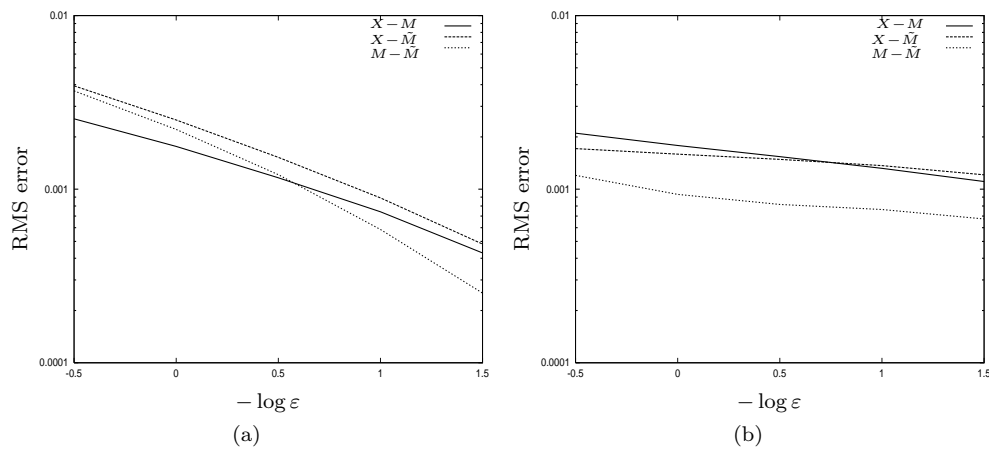FIG. 5.4. *Estimation errors for the* (a) *first and* (b) *second components of X* (G = 0.8).

FIG. 5.5. *Estimation errors for the* (a) *first and* (b) *second components of X* (G = 0.9).

Gx_1^3 with $G = 0.5$, 0.8, 0.9, respectively. The time interval $[0, 100]$ was considered. One can observe the increasing benefit of using filter (1.5) as the nonlinearity in the observations gets stronger. The results obtained by using the extended Kalman filter (EKF) are also shown in [12] for comparison.

REFERENCES

[1] A. BENSOUSSAN, *On some approximation techniques in nonlinear filtering*, in Stochastic Differential Systems, Stochastic Control Theory and Applications, IMA Vol. Math. Appl. 10, Springer-Verlag, New York, 1988, pp. 17–31.

[2] W. H. FLEMING AND É. PARDOUX, *Piecewise monotone filtering with small observation noise*, SIAM J. Control Optim., 27 (1989), pp. 1156–1181.

[3] A. GEGOUT-PETIT, *Approximate filter for the conditional law of a partially observed process in nonlinear filtering*, SIAM J. Control Optim., 36 (1998), pp. 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* I, SIAM J. Appl. Math., 44 (1984), pp. 591–604.

[5] R. KATZUR, B. Z. BOBROVSKY, AND Z. SCHUSS, *Asypmptotic analysis of the optimal filtering problem for one-dimensional diffusions measured in a low noise channel* II, SIAM J. Appl. Math., 44 (1984), pp. 1176–1191.

[6] A. J. KRENER, *The asymptotic approximation of nonlinear filters by linear filters*, in Theory and Applications of Nonlinear Control Systems (Stockholm, 1985), North–Holland, Amsterdam, 1986, pp. 359–378.

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

[8] J. PICARD, *Nonlinear filtering and smoothing with high signal-to-noise ratio*, in Stochastic Processes in Physics and Engineering (Bielefeld, 1986), D. Reidel, Dordrecht, The Netherlands, 1988, pp. 237–251.

[9] J. PICARD, *Efficiency of the extended Kalman filter for nonlinear systems with small noise*, SIAM J. Appl. Math., 51 (1991), pp. 843–885.

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

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

[12] 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 (1990), pp. 1134–1155.

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