

# NONLINEAR FILTERING PROBLEM WITH CONTAMINATION <sup>1</sup>

BY M.L. KLEPTSINA, R. SH. LIPTSER AND A.P. SEREBROVSKI

*Moscow State University of Railway Engineering,*

*Tel Aviv University*

*and*

*Institute for Information Transmission Sciences, Moscow*

## Abstract

We give an approximation for optimal filtering estimates of diffusion type signal and observation under contamination affecting their drifts. The method of analysis is based on the averaging principle and convergence in the total variation for distributions of signal and observation. <sup>2</sup>

**1. Introduction. Formulation of main result.** There are only a few filtering models for which the ‘filtering equation’ obeys a closed form like the Kalman or the Kushner-Zakai filters [8], [17] (see also [16]). A lot of effort has therefore been put into developing approximation techniques for a wide class of filtering models, having more complicated structure than those for which Kalman’s or Kushner-Zakai’s filters are applicable. One approach in this direction is to consider, instead of the original model, a model where the underlying processes are replaced by simple ones that make it possible to construct nearly optimal filters. In the present paper we will consider the filtering problem for continuous random process  $(X_t^\varepsilon, Y_t^\varepsilon)_{t \geq 0}$  with  $X_t^\varepsilon, Y_t^\varepsilon \in \mathbf{R}$ , where  $X_t^\varepsilon$  represents an unobservable signal and  $Y_t^\varepsilon$  is the corresponding observation, and where  $\varepsilon \leq \varepsilon^\circ$  is a small parameter. Suppose the probabilistic structure of  $(X_t^\varepsilon, Y_t^\varepsilon)_{t \geq 0}$  is too complicated to find the optimal (in the mean square sense) filtering estimate but as  $\varepsilon \rightarrow 0$   $(X_t^\varepsilon, Y_t^\varepsilon)_{t \geq 0}$  converges (in some sense) to the limit  $(\bar{X}_t, \bar{Y}_t)_{t \geq 0}$  which is also a continuous random process having a simpler description than the prelimit one, for example, it is a Markov diffusion process or, more specifically, a Gaussian diffusion. Assume some function of the signal  $u(X_t^\varepsilon)$  has to be filtered by observations  $Y_s^\varepsilon, s \leq t$ . Following Kushner [9] and Kushner and Runggaldier [10], as a filtering estimate one can take

$$(1.1) \quad \bar{\pi}_t(Y^\varepsilon) = E[u(\bar{X}_t) | \bar{Y}_s, s \leq t]_{\bar{Y}_s \equiv Y_s^\varepsilon}.$$

---

<sup>1</sup>Supported by International Science Foundation Researches and in part by AT&T Bell Lab; for the second author also partially supported by ONR Grant N00014-96-10413

<sup>2</sup>AMS Subject Classification, 60G35.

Key words and phrases. Diffusions, Averaging Principle, Convergence of Total Variation.

Such a choice of filtering estimate is warranted in [9] by the following arguments. For both the continuous function  $u = u(x)$  and the functional  $\bar{\pi}_t(Y) = E[u(\bar{X}_t)|\bar{Y}_s, s \leq t]_{\bar{Y}_s=Y_s}$  ( $Y_s, s \geq 0$  is a continuous function)

$$(1.2) \quad \lim_{\varepsilon \rightarrow 0} E(u(X_t^\varepsilon) - \bar{\pi}_t(Y^\varepsilon))^2 = E(u(\bar{X}_t) - \bar{\pi}_t(\bar{Y}))^2$$

at least for bounded  $u$  as in [9] or under uniform integrability conditions, which allows one to expect that  $\bar{\pi}_t(Y^\varepsilon)$  is a nearly optimal filtering estimate.

The goal of this paper is to investigate asymptotic properties (with  $\varepsilon \rightarrow 0$ ) of the filtering estimate  $\bar{\pi}_t(Y^\varepsilon)$  from the point of view of asymptotic optimality. We compare  $\bar{\pi}_t(Y^\varepsilon)$  with the optimal (in the mean square sense) filtering estimate  $\pi_t^\varepsilon(Y^\varepsilon) = E[u(X_t^\varepsilon)|Y_s^\varepsilon, s \leq t]$ , and so the following definition is natural.

**DEFINITION.** *Any filtering estimate  $\hat{\pi}_t^\varepsilon$  (with  $\sup_{\varepsilon \leq \varepsilon^\circ} E[\hat{\pi}_t^\varepsilon]^2 < \infty$ ) for  $u(X_t^\varepsilon)$  (with  $\sup_{\varepsilon \leq \varepsilon^\circ} E[u(X_t^\varepsilon)]^2 < \infty$ ) given observations  $Y_s^\varepsilon, s \leq t$  is asymptotically optimal in the mean square sense if*

$$(1.3) \quad \begin{aligned} & \lim_{\varepsilon \rightarrow 0} E(u(X_t^\varepsilon) - \hat{\pi}_t^\varepsilon)^2 \quad \text{exists and is finite,} \\ & \lim_{\varepsilon \rightarrow 0} E(\pi_t^\varepsilon(Y^\varepsilon) - \hat{\pi}_t^\varepsilon)^2 = 0 \end{aligned}$$

or, what is equivalent,

$$\lim_{\varepsilon \rightarrow 0} E(u(X_t^\varepsilon) - \pi_t^\varepsilon(Y^\varepsilon))^2 = \lim_{\varepsilon \rightarrow 0} E(u(X_t^\varepsilon) - \hat{\pi}_t^\varepsilon)^2.$$

Assume (1.2) for  $\bar{\pi}_t(Y^\varepsilon)$ . Then the asymptotic optimality for it, under  $\sup_{\varepsilon \leq \varepsilon^\circ} E[\bar{\pi}_t(Y^\varepsilon)]^2 < \infty$ , is equivalent to

$$(1.4) \quad \lim_{\varepsilon \rightarrow 0} E(u(X_t^\varepsilon) - \bar{\pi}_t(Y^\varepsilon))^2 = E(u(\bar{X}_t) - \bar{\pi}_t(\bar{Y}))^2.$$

In [9], it was mentioned that  $\bar{\pi}_t(Y^\varepsilon)$  is not necessarily asymptotically optimal in sense (1.4) although might be asymptotically optimal in some restricted class of filtering estimates. Also, an example of  $(X_t^\varepsilon, Y_t^\varepsilon)$  is known (see e.g. [13]) for which there exists a non optimal (!) filtering estimate  $\tilde{\pi}_t^\varepsilon$  such that

$$\lim_{\varepsilon \rightarrow 0} E(u(X_t^\varepsilon) - \tilde{\pi}_t^\varepsilon)^2 < \lim_{\varepsilon \rightarrow 0} E(u(X_t^\varepsilon) - \bar{\pi}_t(Y^\varepsilon))^2.$$

It is clear that the problem of the asymptotic optimality for filtering estimates is related to the convergence of the conditional expectation. This problem is effectively

studied by Goggin from the theoretical [5] and filtering [4] points of view, where conditions are given under which  $\pi_t^\varepsilon(Y^\varepsilon)$  converges in law to  $\bar{\pi}_t(\bar{Y})$ :

$$(1.5) \quad \pi_t^\varepsilon(Y^\varepsilon) \xrightarrow{law} \bar{\pi}_t(\bar{Y}), \quad \varepsilon \rightarrow 0.$$

To compare Kushner's and Goggin's results consider the filtering problem on a fixed time interval, say,  $[0, T]$ . Let  $Q_T^\varepsilon$  and  $\bar{Q}_T$  be the distributions of the processes  $(X_t^\varepsilon, Y_t^\varepsilon)_{t \leq T}$  and  $(\bar{X}_t, \bar{Y}_t)_{t \leq T}$  respectively. In Kushner [9], the validity of (1.2) is proved by applying the weak convergence:

$$(1.6) \quad Q_T^\varepsilon \xrightarrow{w} \bar{Q}_T, \quad \varepsilon \rightarrow 0,$$

while in Goggin [5] it was shown that (1.6) does not necessarily imply (1.5), furthermore (1.5) holds if parallel to (1.6) a likelihood, involved in the Bayes formula for  $\pi^\varepsilon(Y^\varepsilon)$ , converges in the law to the corresponding likelihood from the Bayes formula for  $\bar{\pi}(\bar{Y})$ .

Thus, asymptotic optimality of the filtering estimate  $\bar{\pi}_t(Y^\varepsilon)$  requires a different type of convergence than (1.6). Here we use the convergence of  $Q_T^\varepsilon$  to  $\bar{Q}_T$  in the total variation norm:

$$(1.7) \quad \lim_{\varepsilon \rightarrow 0} \|Q_T^\varepsilon - \bar{Q}_T\| = 0$$

and show that it implies (1.2), (1.4) and (1.5) as well.

To formulate the main result introduce the uniform integrability conditions: for any  $t \leq T$  and for some  $\delta > 0$

$$(1.8) \quad \begin{aligned} \sup_{\varepsilon \leq \varepsilon_0} E|u(X_t^\varepsilon)|^{2+\delta} &< \infty, \\ E|u(\bar{X}_t)|^{2+\delta} &< \infty, \\ \sup_{\varepsilon \leq \varepsilon_0} E\left[E(|u(\bar{X}_t)|^{2+\delta} | \bar{Y}_s, s \leq t) |_{\bar{Y} \equiv Y^\varepsilon}\right] &< \infty, \end{aligned}$$

**THEOREM 1.1.** *Assume (1.7). Then for any measurable function  $u$  satisfying (1.8), the filtering estimate  $\bar{\pi}_t(Y^\varepsilon)$ ,  $t \leq T$  is asymptotically optimal in the mean square sense and (1.2), (1.4) hold as well.*

Let  $(X_t^\varepsilon, Y_t^\varepsilon)_{t \leq T}$  and  $(\bar{X}_t, \bar{Y}_t)_{t \leq T}$  be defined on the same probability space. In this case, the distribution  $\bar{Q}_T^\varepsilon$  of the process  $(\bar{X}_t, Y_t^\varepsilon)_{t \leq T}$  is well defined. Suppose (1.7) fails while

$$(1.9) \quad P - \limsup_{\varepsilon \rightarrow 0} \sup_{t \leq T} |X_t^\varepsilon - \bar{X}_t| = 0.$$

THEOREM 1.2. *Assume (1.9) and*

$$(1.10) \quad \lim_{\varepsilon \rightarrow 0} \|\bar{Q}_T^\varepsilon - \bar{Q}_T\| = 0.$$

*Then for any continuous function  $u$  satisfying (1.8), the filtering estimate  $\bar{\pi}_t(Y^\varepsilon), t \leq T$  is asymptotically optimal in the mean square sense and (1.2), (1.4) hold as well.*

COROLLARY. *Under the assumptions of Theorems 1.1 and 1.2, (1.5) holds.*

Theorems 1.1 and 1.2 apply to two filtering models.

MODEL 1.  $(X_t^\varepsilon, Y_t^\varepsilon)_{t \geq 0}$  is defined by Itô's stochastic equations with respect to independent Wiener processes  $(W_t^x)_{t \geq 0}$  and  $(W_t^y)_{t \geq 0}$ :

$$(1.11) \quad \begin{aligned} dX_t^\varepsilon &= a(X_t^\varepsilon, Y_t^\varepsilon, \eta_{t/\varepsilon})dt + dW_t^x \\ dY_t^\varepsilon &= A(X_t^\varepsilon, Y_t^\varepsilon, \eta_{t/\varepsilon})dt + dW_t^y \end{aligned}$$

subject to the initial point  $(X_0, Y_0)$  which is a random vector independent of  $\varepsilon$ , where  $\eta_{t/\varepsilon}$  is a contamination (random or deterministic), affecting drifts. Functions  $a = a(x, y, z)$ ,  $A = A(x, y, z)$  are continuous and Lipschitz continuous in  $(x, y)$  uniformly in  $z$  and  $a(0, 0, z), A(0, 0, z)$  are bounded. In the case in which  $(\eta_t)_{t \geq 0}$  is a random process, it is independent of  $(W_t^x)_{t \geq 0}, (W_t^y)_{t \geq 0}$ , and  $(X_0, Y_0)$ .  $(\eta_t)_{t \geq 0}$  is assumed to obey the following ergodic properties: for any bounded and continuous function  $f$  and every  $t > 0$  there exists a constant  $c_f$ , depending on  $f$ , such that

$$(1.12) \quad P - \lim_{\varepsilon \rightarrow 0} \int_0^t f(\eta_{s/\varepsilon})ds = tc_f$$

and

$$(1.13) \quad P - \lim_{\varepsilon \rightarrow 0} E(f(\eta_{t/\varepsilon}) | X_s^\varepsilon, Y_s^\varepsilon, s \leq t) = c_f.$$

MODEL 2.  $(X_t^\varepsilon, Y_t^\varepsilon)_{t \geq 0}$  is defined by Itô's stochastic equations similar to (1.11):

$$(1.14) \quad \begin{aligned} dX_t^\varepsilon &= a(X_t^\varepsilon, \eta_{t/\varepsilon})dt + dW_t^x \\ dY_t^\varepsilon &= A(X_t^\varepsilon, Y_t^\varepsilon)dt + dW_t^y \end{aligned}$$

subject to the initial point  $(X_0, Y_0)$ . Function  $a = a(x, z)$  is continuous and Lipschitz continuous in  $x$  uniformly in  $z$ ,  $a(0, z)$  is bounded. Function  $A = A(x, y)$  is Lipschitz continuous. The process  $(\eta_t)$  satisfies only the ergodic property (1.12).

For both models, the ergodic property (1.12) is inherited by  $(X_t^\varepsilon, Y_t^\varepsilon)$  (see [7] and also [3]): for each  $T > 0$

$$(1.15) \quad P - \lim_{\varepsilon \rightarrow 0} (\sup_{t \leq T} |X_t^\varepsilon - \bar{X}_t| + \sup_{t \leq T} |Y_t^\varepsilon - \bar{Y}_t|) = 0,$$

where  $(\bar{X}_t, \bar{Y}_t)$  is a Markovian diffusion defined by Ito's equations: for Model 1

$$d\bar{X}_t = \bar{a}(\bar{X}_t, \bar{Y}_t)dt + dW_t^x, \quad d\bar{Y}_t = \bar{A}(\bar{X}_t, \bar{Y}_t)dt + dW_t^y$$

and for Model 2

$$d\bar{X}_t = \bar{a}(\bar{X}_t)dt + dW_t^x, \quad d\bar{Y}_t = A(\bar{X}_t, \bar{Y}_t)dt + dW_t^y$$

subject to the same initial points  $\bar{X}_0 = X_0$  and  $\bar{Y}_0 = Y_0$ , where for  $x, y \in \mathbf{R}$

$$(1.16) \quad \begin{aligned} \bar{a}(x, y) &= P - \lim_{\varepsilon \rightarrow 0} \int_0^1 a(x, y, \eta_{s/\varepsilon}) ds, \\ \bar{A}(x, y) &= P - \lim_{\varepsilon \rightarrow 0} \int_0^1 A(x, y, \eta_{s/\varepsilon}) ds \end{aligned}$$

and

$$(1.17) \quad \bar{a}(x) = P - \lim_{\varepsilon \rightarrow 0} \int_0^1 a(x, \eta_{s/\varepsilon}) ds.$$

Theorems 1.1 and 1.2 imply

**THEOREM 1.3.** *Let  $(X_t^\varepsilon, Y_t^\varepsilon)_{t \geq 0}$  be defined corresponding to Model 1 or 2. Let  $u$  satisfy (1.8). In the case of Model 2  $u$  is assumed to be continuous. Then  $\bar{\pi}_t(Y^\varepsilon)$ ,  $t \geq 0$  is an asymptotically optimal (in the mean square sense) filtering estimate and (1.2), (1.4), and (1.5) hold as well.*

it is clear that the ergodic property (1.12) holds for deterministic periodic functions  $\eta_t$ , stationary processes, ergodic Markov processes, etc., while the ergodic property (1.13) is not trivial to verify. Nevertheless, we examine it for a homogeneous Markov process  $\eta_t$ , having the unique invariant measure  $\mu$  such that its transition probability  $\lambda_{y,t} = \lambda(y, t, dz)$  converges in the total variation norm to  $\mu$ :

$$(1.18) \quad \lim_{t \rightarrow \infty} \|\lambda_{y,t} - \mu\| = 0, \quad \forall y \in \mathbf{R}.$$

Also in this case we have

$$\begin{aligned} \bar{a}(x, y) &= \int_{\mathbf{R}} a(x, y, z) \mu(dz) \\ \bar{A}(x, y) &= \int_{\mathbf{R}} A(x, y, z) \mu(dz) \\ \bar{a}(x) &= \int_{\mathbf{R}} Ra(x, z) \mu(dz). \end{aligned}$$

We show in Section 5 that (1.18) implies (1.12) and (1.13) and give two examples of Markov processes for which (1.18) takes place. Also in this Section an example is given which shows that if (1.13) fails then asymptotic optimality for  $\bar{\pi}_t(Y^\varepsilon)$  fails too.

Proofs of Theorem 1.1 and 1.2 are given in Section 2. Investigation of Models 1 and 2 appears in Sections 3 and 4 respectively.

**2. Proof of Theorems 1.1 and 1.2.** The statements of Theorems 1.1 and 1.2 are derived from

LEMMA 2.1 *Under the assumptions of Theorem 1.1 and 1.2 for any  $t \leq T$*

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \int (u(X_t) - \bar{\pi}_t(Y)) dQ_T^\varepsilon &= \int (u(X_t) - \bar{\pi}_t(Y)) d\bar{Q}_T \\ \liminf_{\varepsilon \rightarrow 0} \int (u(X_t) - \pi_t^\varepsilon(Y))^2 dQ_T^\varepsilon &\geq \int (u(X_t) - \bar{\pi}_t(Y))^2 d\bar{Q}_T. \end{aligned}$$

In fact, the first statement is nothing but (1.2). Therefore the asymptotic optimality of  $\bar{\pi}_t(Y^\varepsilon)$  is equivalent to (1.4). On the other hand, since for any  $\varepsilon$ ,  $\pi_t^\varepsilon(Y^\varepsilon)$  is an optimal (in the mean square sense) filtering estimate we get

$$(2.1) \quad \int (u(X_t) - \pi_t^\varepsilon(Y)) dQ_T^\varepsilon \leq \int (u(X_t) - \bar{\pi}_t(Y))^2 d\bar{Q}_T.$$

Thus, (1.4) is implied by the second statement of Lemma 2.1 and (2.1).

PROOF OF LEMMA 2.1. We consider separately two cases:  $u$  is bounded function;  $u$  satisfies (1.8).

Assume there exists a constant, say,  $\ell$  such that

$$(2.2) \quad |u| \leq \ell.$$

Hence, bounded versions of functionals  $\pi_t^\varepsilon(\cdot)$ ,  $\bar{\pi}_t(\cdot)$  can be chosen (say,  $|\pi_t^\varepsilon(\cdot)|, |\bar{\pi}_t(\cdot)| \leq \ell$ ).

Assume (1.7). Then

$$\begin{aligned} & \left| \int (u(X_t) - \bar{\pi}_t(Y))^2 dQ_T^\varepsilon - \int (u(X_t) - \bar{\pi}_t(Y))^2 d\bar{Q}_T \right| \\ & \leq 4\ell^2 \|Q_T^\varepsilon - \bar{Q}_T\| \rightarrow 0, \quad \varepsilon \rightarrow 0. \end{aligned}$$

Assume (1.9) and (1.10). Then by the triangle inequality

$$(2.3) \quad \begin{aligned} & \left| \int (u(X_t) - \bar{\pi}_t(Y))^2 dQ_T^\varepsilon - \int (u(X_t) - \bar{\pi}_t(Y))^2 d\bar{Q}_T \right| \\ & \leq \left| \int (u(X_t) - \bar{\pi}_t(Y))^2 dQ_T^\varepsilon - \int (u(X_t) - \bar{\pi}_t(Y))^2 d\bar{Q}_T^\varepsilon \right| \\ & + \left| \int (u(X_t) - \bar{\pi}_t(Y))^2 d\bar{Q}_T^\varepsilon - \int (u(X_t) - \bar{\pi}_t(Y))^2 d\bar{Q}_T \right| \end{aligned}$$

The last term on the right hand side of (2.3) is bounded above by  $4\ell^2 \|\bar{Q}_T^\varepsilon - \bar{Q}_T\|$  and goes to zero as  $\varepsilon \rightarrow 0$  by virtue of (1.10). The second term is nothing but  $|E(u(X_t^\varepsilon) - \bar{\pi}_t(Y^\varepsilon))^2 - E(u(\bar{X}_t - \bar{\pi}_t(Y^\varepsilon))^2)|$ . It can be bounded above by  $4\ell E|u(X_t^\varepsilon) - u(\bar{X}_t)|$ . So it converges to zero as  $\varepsilon \rightarrow 0$  by virtue of (1.9) and the continuity of the function  $u$  which is required by Theorem 1.2.

Thus, the first statement of Lemma 2.1 is proved. Now we prove the second statement. Write

$$(2.4) \quad \begin{aligned} \int (u(X_t) - \pi_t^\varepsilon(Y))^2 dQ_T^\varepsilon &= \int (u(X_t) - \pi_t^\varepsilon(Y))^2 d\bar{Q}_T \\ &+ \int (u(X_t) - \pi_t^\varepsilon(Y))^2 d(Q_T^\varepsilon - \bar{Q}_T). \end{aligned}$$

Since  $\bar{\pi}_t(\bar{Y})$  is the optimal (in the mean square sense) filtering estimate for  $\bar{X}_t$  given observations  $\bar{Y}_s, s \leq t$  we get

$$(2.5) \quad \int (u(X_t) - \pi_t^\varepsilon(Y))^2 d\bar{Q}_T \geq \int (u(X_t) - \bar{\pi}_t(Y))^2 d\bar{Q}_T.$$

Also it is clear  $\int (u(X_t) - \pi_t^\varepsilon(Y))^2 d(Q_T^\varepsilon - \bar{Q}_T) \geq -4\ell^2 \|Q_T^\varepsilon - \bar{Q}_T\|$ . Hence

$$\int (u(X_t) - \pi_t^\varepsilon(Y))^2 dQ_T^\varepsilon \geq \int (u(X_t) - \bar{\pi}_t(Y))^2 d\bar{Q}_T - 4\ell^2 \|Q_T^\varepsilon - \bar{Q}_T\|$$

and so, under (1.7), the second statement holds.

To check the second statement under (1.9) and (1.10) with a continuous function  $u$ , write

$$\begin{aligned} \int (u(X_t) - \pi_t^\varepsilon(Y))^2 dQ_T^\varepsilon &= \int (u(X_t) - \pi_t^\varepsilon(Y))^2 d\bar{Q}_T \\ &\quad + \int (u(X_t) - \pi_t^\varepsilon(Y))^2 d(Q_T^\varepsilon - \bar{Q}_T^\varepsilon) \\ (2.6) \quad &\quad + \int (u(X_t) - \pi_t^\varepsilon(Y))^2 d(\bar{Q}_T^\varepsilon - \bar{Q}_T). \end{aligned}$$

Due to (2.5), the first term on the right hand side of (2.6) is bounded below by  $\int (u(X_t) - \bar{\pi}_t(Y))^2 d\bar{Q}_T$ ; the second is equal to  $E(u(X_t^\varepsilon) - \pi_t^\varepsilon(Y^\varepsilon))^2 - E(u(\bar{X}_t) - \pi_t^\varepsilon(Y^\varepsilon))^2$  and so it can be bounded below by  $-4\ell E|u(X_t^\varepsilon) - u(\bar{X}_t)|$ ; the third obeys the lower bound  $-4\ell^2 \|\bar{Q}_T^\varepsilon - \bar{Q}_T\|$ . Hence,

$$\begin{aligned} \int (u(X_t) - \pi_t^\varepsilon(Y))^2 dQ_T^\varepsilon &\geq \int (u(X_t) - \bar{\pi}_t(Y))^2 d\bar{Q}_T \\ &\quad - 4\ell E|u(X_t^\varepsilon) - u(\bar{X}_t)| \\ &\quad - 4\ell^2 \|\bar{Q}_T^\varepsilon - \bar{Q}_T\| \\ &\rightarrow \int (u(X_t) - \bar{\pi}_t(Y))^2 d\bar{Q}_T, \quad \varepsilon \rightarrow 0. \end{aligned}$$

Thus, for bounded functions  $u$  the lemma is proved.

Assume  $u$  is an unbounded function for which (1.8) holds. For  $n \geq 1$ , put  $u_n(x) = g_n(u(x))$ , where  $g_n(x) = \begin{cases} x, & |x| \leq n \\ n \operatorname{sign} x, & |x| > n. \end{cases}$  Also put  $\pi_t^{\varepsilon, n}(Y^\varepsilon) = E(u_n(X_t^\varepsilon) | \mathcal{F}_t^{Y^\varepsilon})$  and  $\bar{\pi}_t^n(\bar{Y}) = E(u_n(\bar{X}_t) | \mathcal{F}_t^{\bar{Y}})$ . By the statements of the lemma proved above for bounded  $u$  we find for every fixed  $n$

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \int [u_n(X_t) - \bar{\pi}_t^n(Y)]^2 dQ_T^\varepsilon &= \int [u_n(X_t) - \bar{\pi}_t^n(Y)]^2 d\bar{Q}_T \\ (2.7) \quad \liminf_{\varepsilon \rightarrow 0} \int [u_n(X_t) - \pi_t^{\varepsilon, n}(Y)]^2 dQ_T^\varepsilon &\geq \int [u_n(X_t) - \bar{\pi}_t^n(Y)]^2 d\bar{Q}_T. \end{aligned}$$

Evidently, the statements of the lemma hold if

$$(2.8) \quad \begin{aligned} \lim_n \limsup_{\varepsilon \rightarrow 0} E[u(X_t^\varepsilon) - u_n(X_t^\varepsilon)]^2 &= 0 \\ \lim_n E[u(\bar{X}_t) - u_n(\bar{X}_t)]^2 &= 0 \end{aligned}$$

and

$$(2.9) \quad \begin{aligned} \lim_n \limsup_{\varepsilon \rightarrow 0} E[\bar{\pi}_t(Y^\varepsilon) - \bar{\pi}_t^n(Y^\varepsilon)]^2 &= 0 \\ \lim_n \limsup_{\varepsilon \rightarrow 0} E[\pi_t^\varepsilon(Y^\varepsilon) - \pi_t^{\varepsilon, n}(Y^\varepsilon)]^2 &= 0. \end{aligned}$$

An obvious estimate  $[u(x) - u_n(x)]^2 \leq u^2(x)I(|u(x)| > n) \leq \frac{1}{n^\delta}|u(x)|^{2+\delta}$  is used for checking (2.8) and (2.9). Namely,

$$\begin{aligned} E[u(X_t^\varepsilon) - u_n(X_t^\varepsilon)]^2 &\leq \frac{1}{n^\delta} E|u(X_t^\varepsilon)|^{2+\delta} \\ E[u(\bar{X}_t) - u_n(\bar{X}_t)]^2 &\leq \frac{1}{n^\delta} E|u(\bar{X}_t)|^{2+\delta} \end{aligned}$$

and

$$\begin{aligned} E[\bar{\pi}_t(Y^\varepsilon) - \bar{\pi}_t^n(Y^\varepsilon)]^2 &= E[E(u(\bar{X}_t) - u_n(\bar{X}_t)|\bar{Y}_s, s \leq t)_{\bar{Y} \equiv Y^\varepsilon}]^2 \\ &\leq \frac{1}{n^\delta} E[E(|u(\bar{X}_t)|^{2+\delta}|\bar{Y}_s, s \leq t)_{\bar{Y} \equiv Y^\varepsilon}] \\ E[\pi_t^\varepsilon(Y^\varepsilon) - \pi_t^{\varepsilon, n}(Y^\varepsilon)]^2 &\leq E[E(u(X_t^\varepsilon) - u_n(X_t^\varepsilon)|Y_s^\varepsilon, s \leq t)]^2 \\ &\leq \frac{1}{n^\delta} E|u(X_t^\varepsilon)|^{2+\delta}, \end{aligned}$$

that is (2.8), (2.9) hold due to (1.8).

**PROOF OF THE COROLLARY:** It is sufficient to show the convergence of the characteristic functions:

$$\lim_{\varepsilon \rightarrow 0} Ee^{iv\pi_t^\varepsilon(Y^\varepsilon)} = Ee^{iv\bar{\pi}_t(\bar{Y})}, \quad v \in \mathbf{R}, \quad \mathbf{i} = \sqrt{-1}.$$

To this end, we use the following estimate

$$(2.10) \quad \begin{aligned} |Ee^{iv\pi_t^\varepsilon(Y^\varepsilon)} - Ee^{iv\bar{\pi}_t(\bar{Y})}| &\leq E|e^{iv\pi_t^\varepsilon(Y^\varepsilon)} - e^{iv\bar{\pi}_t(Y^\varepsilon)}| \\ &\quad + |E(e^{iv\bar{\pi}_t(Y^\varepsilon)} - e^{iv\bar{\pi}_t(\bar{Y})})| \end{aligned}$$

and note that the first term on the right hand side of (2.10) is bounded above by  $\sqrt{E(\pi_t^\varepsilon(Y^\varepsilon) - \bar{\pi}_t(Y^\varepsilon))^2}$  while the second, under the assumptions of Theorem 1.1, by  $\|Q_T^\varepsilon - \bar{Q}_T\|$  and, under the assumptions of Theorem 1.2, by  $\|\bar{Q}_T^\varepsilon - \bar{Q}_T\|$ .

**3. Model 1.** It is sufficient to show that (1.12), (1.13) imply (1.7) for every  $T > 0$ . Let  $(\Omega, \mathcal{F}, \mathbf{F}^\varepsilon = (\mathcal{F}_t^\varepsilon)_{t \geq 0}, P)$  be a stochastic basis with the filtration  $\mathbf{F}^\varepsilon$  generated by paths of  $(X_t^\varepsilon, Y_t^\varepsilon)_{t \geq 0}$ . Denote by  $a_s^\varepsilon(x, y; \{X^\varepsilon, Y^\varepsilon\})$  the  $\mathbf{F}^\varepsilon$ -predictable projection of  $a(x, y, \eta_{s/\varepsilon})$ . It is well known (see e.g. [11] or [15]) that for each fixed  $s$   $P$ -a.s.

$$(3.1) \quad \begin{aligned} a_s^\varepsilon(x, y; \{X^\varepsilon, Y^\varepsilon\}) &= E(a(x, y, \eta_{s/\varepsilon}) | \mathcal{F}_s^\varepsilon) \\ a_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon; \{X^\varepsilon, Y^\varepsilon\}) &= E(a(X_s^\varepsilon, Y_s^\varepsilon, \eta_{s/\varepsilon}) | \mathcal{F}_s^\varepsilon). \end{aligned}$$

Condition (1.13) implies that for fixed  $s, x, y$ , as  $\varepsilon \rightarrow 0$ ,  $a_s^\varepsilon(x, y; \{X^\varepsilon, Y^\varepsilon\})$  converges in probability to  $\bar{a}(x, y)$  defined in (1.16). Due to the assumptions on  $a(x, y, z)$ , for fixed  $x, y$  a uniformly bounded in  $s \leq T$  and  $\varepsilon \leq \varepsilon^\circ$  version of  $a_s^\varepsilon(x, y; \{X^\varepsilon, Y^\varepsilon\})$  can be chosen. Thus the above convergence implies

$$(3.2) \quad P - \lim_{\varepsilon \rightarrow 0} \int_0^T [a_s^\varepsilon(x, y; \{X^\varepsilon, Y^\varepsilon\}) - \bar{a}(x, y)]^2 ds = 0.$$

Next we show that (3.2) remains true upon replacing  $x, y$  by  $X_s^\varepsilon, Y_s^\varepsilon$ , that is, for each  $T > 0$

$$(3.3) \quad P - \lim_{\varepsilon \rightarrow 0} \int_0^T [a_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon; \{X^\varepsilon, Y^\varepsilon\}) - \bar{a}(X_s^\varepsilon, Y_s^\varepsilon)]^2 ds = 0.$$

Denote by  $[x]$  the greatest integer function and define the following random processes:  $X_t^{\varepsilon, n} = X_{[nt]/n}^\varepsilon$  and  $X_t^{\varepsilon, n, m} = [mX_t^{\varepsilon, n}]/m$ . Put  $X_T^{\varepsilon, *} = \sup_{t \leq T} |X_t^\varepsilon|$ . In the same way define  $Y_t^{\varepsilon, n}, Y_t^{\varepsilon, n, m}$ , and  $Y_T^{\varepsilon, *}$ . We show that for every  $C > 0$ , and  $m, n \geq 1$  on the set  $\{X_T^{\varepsilon, *} + Y_T^{\varepsilon, *} \leq C\}$

$$(3.4) \quad \int_0^T [a_s^\varepsilon(X_s^{\varepsilon, n, m}, Y_s^{\varepsilon, n, m}; \{X^\varepsilon, Y^\varepsilon\}) - \bar{a}(X_s^{\varepsilon, n, m}, Y_s^{\varepsilon, n, m})]^2 ds \rightarrow 0,$$

in probability as  $\varepsilon \rightarrow 0$ . The proof of (3.4) is based on the fact that the processes  $X_t^{\varepsilon, n, m}$  and  $Y_t^{\varepsilon, n, m}$  on the set  $\{X_T^{\varepsilon, *} + Y_T^{\varepsilon, *} \leq C\}$  have a finite number of paths, which is independent of  $\varepsilon$ , and so (3.4) holds if for any continuous functions  $x_s, y_s, 0 \leq s \leq T$  and  $x_s^{n, m}, y_s^{n, m}, x_{j/n}^m$ , and  $y_{j/n}^m$  defined similarly to  $X_t^{\varepsilon, n, m}, Y_t^{\varepsilon, n, m}, X_{j/n}^{\varepsilon, m}$ , and  $Y_{j/n}^{\varepsilon, m}$

$$(3.5) \quad P - \lim_{\varepsilon \rightarrow 0} \int_0^T [a_s^\varepsilon(x_s^{n,m}, y_s^{n,m}; \{X^\varepsilon, Y^\varepsilon\}) - \bar{a}(x_s^{n,m}, y_s^{n,m})]^2 ds = 0.$$

In turn the validity of (3.5) follows from a chain of upper bounds:

$$\begin{aligned} & \int_0^T [a_s^\varepsilon(x_s^{n,m}, y_s^{n,m}; \{X^\varepsilon, Y^\varepsilon\}) - \bar{a}(x_s^{n,m}, y_s^{n,m})]^2 ds \\ & \leq \sum_{j=1}^{[nT]} \int_{\frac{j-1}{n}}^{\frac{j}{n}} [a_s^\varepsilon(x_s^{n,m}, y_s^{n,m}; \{X^\varepsilon, Y^\varepsilon\}) - \bar{a}(x_s^{n,m}, y_s^{n,m})]^2 ds \\ & \leq 2 \sum_{j=1}^{[nT]} \int_0^T [a_s^\varepsilon(x_{j/n}^m, y_{j/n}^m; \{X^\varepsilon, Y^\varepsilon\}) - \bar{a}(x_{j/n}^m, y_{j/n}^m)]^2 ds. \end{aligned}$$

For fixed  $n, m$ , each summand in the last sum converges to zero in probability as  $\varepsilon \rightarrow 0$ . Thus, (3.4) holds. Consequently (3.3) takes place if

$$(3.6) \quad \lim_{C \rightarrow \infty} \limsup_{\varepsilon \rightarrow 0} P(X_T^{\varepsilon,*} + Y_T^{\varepsilon,*} \geq C) = 0$$

and for every  $C > 0$  on the set  $\{X_T^{\varepsilon,*} + Y_T^{\varepsilon,*} \leq C\}$

$$(3.7) \quad \begin{aligned} & \int_0^T [a_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon; \{X^\varepsilon, Y^\varepsilon\}) - a_s^\varepsilon(X_s^{\varepsilon,n,m}, Y_s^{\varepsilon,n,m}; \{X^\varepsilon, Y^\varepsilon\})]^2 ds \rightarrow 0 \\ & \int_0^T [\bar{a}(X_s^\varepsilon, Y_s^\varepsilon) - \bar{a}(X_s^{\varepsilon,n,m}, Y_s^{\varepsilon,n,m})]^2 ds \rightarrow 0 \end{aligned}$$

in probability as the limit  $\lim_{n,m} \limsup_{\varepsilon \rightarrow 0}$  is taken. It is clear that (3.5) is implied by (1.15). On the other hand, both functions  $a_s^\varepsilon(x, y; \{X^\varepsilon, Y^\varepsilon\})$  and  $\bar{a}(x, y)$  inherit the Lipschitz property in  $x, y$  with an absolute constant, which implies (3.7) under

$$P - \lim_{n,m} \limsup_{\varepsilon \rightarrow 0} \sup_{s \leq T} (|X_s^\varepsilon - X_s^{\varepsilon,n,m}| + |Y_s^\varepsilon - Y_s^{\varepsilon,n,m}|) = 0.$$

The last holds by virtue of (1.15) too.

Thus, the validity of (3.3) is proved. In the same way, for  $A_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon; \{X^\varepsilon, Y^\varepsilon\}) = E(A(X_s^\varepsilon, Y_s^\varepsilon, \eta_{s/\varepsilon}) | \mathcal{F}_s^\varepsilon)$  and  $\bar{A}(x, y)$ , defined in (1.16), we obtain

$$(3.8) \quad P - \lim_{\varepsilon \rightarrow 0} \int_0^T [A_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon; \{X^\varepsilon, Y^\varepsilon\}) - \bar{A}(X_s^\varepsilon, Y_s^\varepsilon)]^2 ds = 0.$$

For brevity of notations, hereafter  $a_s^\varepsilon$  and  $A_s^\varepsilon$  will be used to designate  $a_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon; \{X^\varepsilon, Y^\varepsilon\})$  and  $A_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon; \{X^\varepsilon, Y^\varepsilon\})$  respectively. It is well known (see e.g. Theorem 7.12 in [14]) that

$$(3.9) \quad \widetilde{W}_t^{x,\varepsilon} = X_t^\varepsilon - X_0 - \int_0^t a_s^\varepsilon ds, \quad \widetilde{W}_t^{y,\varepsilon} = Y_t^\varepsilon - Y_0 - \int_0^t A_s^\varepsilon ds$$

are independent Wiener processes with respect to the filtration  $\mathbf{F}^\varepsilon$ , that is, the process  $(X_t^\varepsilon, Y_t^\varepsilon)_{t \geq 0}$  is defined by the past-dependent Itô equations

$$(3.10) \quad \begin{aligned} X_t^\varepsilon &= X_0 + \int_0^t a_s^\varepsilon ds + \widetilde{W}_t^{x,\varepsilon} \\ Y_t^\varepsilon &= Y_0 + \int_0^t A_s^\varepsilon ds + \widetilde{W}_t^{y,\varepsilon}. \end{aligned}$$

Hence, by Theorems 7.19, 7.20 in [14], for every  $T > 0$  the measures  $Q_T^\varepsilon$  and  $\overline{Q}_T$  are equivalent to the distribution of a pair of independent Wiener processes on the time interval  $[0, T]$ , having initial points  $X_0$  and  $Y_0$  respectively. Thus, these measures are equivalent ( $Q_T^\varepsilon \sim \overline{Q}_T$ ) and the density  $\frac{d\overline{Q}_T}{dQ_T^\varepsilon}$  at a point “ $X^\varepsilon, Y^\varepsilon$ ” is given by the formula (see Theorems 7.19, 7.20 in [14])

$$(3.11) \quad \begin{aligned} \frac{d\overline{Q}_T}{dQ_T^\varepsilon}(X^\varepsilon, Y^\varepsilon) &= \exp \left( \int_0^T \{[\overline{a}(X_s^\varepsilon, Y_s^\varepsilon) - a_s^\varepsilon] dX_s^\varepsilon \right. \\ &\quad \left. + \int_0^T [\overline{A}(X_s^\varepsilon, Y_s^\varepsilon) - A_s^\varepsilon] dY_s^\varepsilon \right\} \\ &\quad - \frac{1}{2} \int_0^T \{[(\overline{a}(X_s^\varepsilon, Y_s^\varepsilon))^2 - (a_s^\varepsilon)^2]\} ds \\ &\quad - \frac{1}{2} \int_0^T \{[(\overline{A}(X_s^\varepsilon, Y_s^\varepsilon))^2 - (A_s^\varepsilon)^2]\} ds). \end{aligned}$$

Taking into account that  $X_t^\varepsilon$  and  $Y_t^\varepsilon$  satisfy (3.10), one can to rewrite the formula (3.11) in the form:

$$(3.12) \quad \frac{d\overline{Q}_T}{dQ_T^\varepsilon}(X^\varepsilon, Y^\varepsilon) = \exp(M_T^\varepsilon - \frac{1}{2} \langle M^\varepsilon \rangle_T)$$

with a continuous martingale  $(M_t^\varepsilon)_{t \leq T}$  and its predictable quadratic variation  $(\langle M^\varepsilon \rangle_t)_{t \leq T}$ , where

$$M_T^\varepsilon = \int_0^T [\bar{a}(X_s^\varepsilon, Y_s^\varepsilon) - a_s^\varepsilon] d\widetilde{W}_s^{x,\varepsilon} + \int_0^T [\bar{A}(X_s^\varepsilon, Y_s^\varepsilon) - A_s^\varepsilon] d\widetilde{W}_s^{y,\varepsilon},$$

and where

$$\langle M^\varepsilon \rangle_T = \int_0^T \{[\bar{a}(X_s^\varepsilon, Y_s^\varepsilon) - a_s^\varepsilon]^2 + [\bar{A}(X_s^\varepsilon, Y_s^\varepsilon) - A_s^\varepsilon]^2\} ds.$$

Due to (3.3) and (3.8),  $\langle M^\varepsilon \rangle_T$  converges to zero in probability as  $\varepsilon \rightarrow 0$  and so, by Problem 1.9.2 in [15] the same convergence holds for  $M_T^\varepsilon$ . Consequently, the density  $\frac{d\bar{Q}_T}{dQ_T^\varepsilon}(X^\varepsilon, Y^\varepsilon)$  converges to 1 in probability as  $\varepsilon \rightarrow 0$ .

We show that (1.7) holds. Since  $\|Q_T^\varepsilon - \bar{Q}_T\| = E|1 - \frac{d\bar{Q}_T}{dQ_T^\varepsilon}(X^\varepsilon, Y^\varepsilon)|$ , (1.7) is implied by Scheffe's theorem [1].

**4. Model 2.** Since (1.12) implies (1.15) and so (1.9) holds, it remains to show that (1.12) implies (1.10) too. Corresponding to Model 2 the process  $(\bar{X}_t, \bar{Y}_t)$  is defined by the Itô equations:

$$(4.1) \quad \begin{aligned} \bar{X}_t &= X_0 + \int_0^t \bar{a}(\bar{X}_s) ds + W_t^x \\ \bar{Y}_t &= Y_0 + \int_0^t A(\bar{X}_s, \bar{Y}_s) ds + W_t^y, \end{aligned}$$

where  $\bar{a}(x)$  is defined in (1.17). By virtue of the independence of  $(W_t^y)$  and  $(X_0, Y_0, (X_t^\varepsilon), (\bar{X}_t), (\eta_t))$ , without loss of generality, one can assume that  $X_0, Y_0, (X_t^\varepsilon), (\bar{X}_t), (\eta_t)$  are defined on a probability space  $(\Omega', \mathcal{F}', P')$ , while  $(W_t^y)$  is defined on its copy  $(\Omega'', \mathcal{F}'', P'')$  (notations  $E'$  and  $E''$  will be used for denoting expectations w.r.t.  $P'$  and  $P''$  respectively). So, the processes  $(Y_t^\varepsilon)$  and  $(\bar{Y}_t)$  are both defined on  $(\Omega' \times \Omega'', \mathcal{F}' \otimes \mathcal{F}'', P' \times P'')$ :

$$(4.2) \quad \begin{aligned} Y_s^\varepsilon(\omega', \omega'') &= Y_0(\omega') + \int_0^s A(X_u^\varepsilon(\omega'), Y_u^\varepsilon(\omega', \omega'')) du + W_s^y(\omega'') \\ \bar{Y}_s(\omega', \omega'') &= Y_0(\omega') + \int_0^s A(\bar{X}_u(\omega'), \bar{Y}_u(\omega', \omega'')) du + W_s^y(\omega''). \end{aligned}$$

Now define a random process

$$(4.3) \quad \begin{aligned} Z_t^\varepsilon(\omega, \omega') &= \exp \left( \int_0^t [A(\bar{X}_s(\omega'), Y_s^\varepsilon(\omega', \omega'')) - A(X_s^\varepsilon(\omega'), Y_s^\varepsilon(\omega', \omega''))] dW_s^y(\omega'') \right. \\ &\quad \left. - \frac{1}{2} \int_0^t [A(\bar{X}_s(\omega'), Y_s^\varepsilon(\omega', \omega'')) - A(X_s^\varepsilon(\omega'), Y_s^\varepsilon(\omega', \omega''))]^2 ds \right), \end{aligned}$$

and show that for every  $T > 0$

$$(4.4) \quad (E' \times E'')Z_T^\varepsilon(\omega', \omega'') = 1.$$

For fixed  $\omega'$  and  $T$ , the distributions  $R_T^\varepsilon(\omega')$ ,  $\bar{R}_T^\varepsilon(\omega')$  of the processes  $(Y_t^\varepsilon(\omega', \omega''))$ ,  $(\bar{Y}_t(\omega', \omega''))$  on the time interval  $[0, T]$ , are equivalent to the distribution of the Wiener process having initial point  $Y_0(\omega')$  (see Theorem 7.7 in [14]). Therefore, these distributions are equivalent and obey the density

$$\frac{d\bar{R}_T^\varepsilon}{dR_T^\varepsilon}(\omega'; Y^\varepsilon(\omega', \omega'')) = Z_T^\varepsilon(\omega', \omega''), \quad \forall \omega' \in \Omega'$$

which implies  $E''Z_T^\varepsilon(\omega', \omega'') = 1$ ,  $\omega' \in \Omega'$  and in turn (4.4).

Denote by  $\mathcal{G}_T^\varepsilon$  the  $\sigma$ -algebra generated by  $(\eta_{t/\varepsilon}, X_t^\varepsilon, \bar{X}_t, Y_t^\varepsilon, t \leq T)$ , and by  $P_T^\varepsilon$  the restriction of  $P' \times P''$  to  $\mathcal{G}_T^\varepsilon$ . Define a new probability measure  $\bar{P}_T^\varepsilon$  by

$$(4.5) \quad d\bar{P}_T^\varepsilon = Z_T^\varepsilon(\omega', \omega'')dP_T^\varepsilon.$$

The structure of the density  $Z_T^\varepsilon(\omega', \omega'')$  implies the following properties for the processes considered. All of them, defined on  $(\Omega', \mathcal{F}')$ , have the same distributions with respect to both measures  $P_T^\varepsilon$  and  $\bar{P}_T^\varepsilon$  while the distribution of  $(Y_t^\varepsilon(\omega', \omega''))_{0 \leq t \leq T}$  is changed. In fact, by Theorem 7.12 in [14]

$$\bar{W}_t^\varepsilon(\omega', \omega'') = W_t^y(\omega'') - \int_0^t [A(\bar{X}_s(\omega'), Y_s^\varepsilon(\omega', \omega'')) - A(X_s^\varepsilon(\omega'), Y_s^\varepsilon(\omega', \omega''))]ds$$

is a Wiener process with respect to  $\bar{P}_T^\varepsilon$ , which implies that the process  $(Y_t^\varepsilon(\omega', \omega''))_{t \leq T}$  obeys a new representation with respect to  $\bar{P}_T^\varepsilon$ :

$$Y_t^\varepsilon(\omega', \omega'') = Y_0(\omega') + \int_0^t A(\bar{X}_s(\omega'), Y_s^\varepsilon(\omega', \omega''))du + \bar{W}_t^\varepsilon(\omega', \omega'').$$

Comparing this Itô equation with the second equation from (4.2) for  $\bar{Y}_t(\omega', \omega'')$ , one can conclude that the distribution of  $(Y_t^\varepsilon(\omega', \omega''))_{0 \leq t \leq T}$  with respect to  $\bar{P}_T^\varepsilon$  coincides with the distribution of  $(\bar{Y}_t(\omega', \omega''))_{0 \leq t \leq T}$  with respect to  $P_T^\varepsilon$ . Therefore, repeating arguments from the proof of Theorem 7.1 in [14], we find that  $\bar{Q}_T \sim \bar{Q}_T^\varepsilon$  and

$$(4.6) \quad \frac{d\bar{Q}_T}{d\bar{Q}_T^\varepsilon}(\bar{X}, Y^\varepsilon) = (E' \times E'')(Z_T^\varepsilon(\omega', \omega'') | \bar{X}_t, Y_t^\varepsilon, t \leq T) \quad P' \times P'' - \text{a.s.}$$

We show that (1.10) takes place. From (4.6) and Jensen's inequality, it follows that

$$\begin{aligned}
\|\bar{Q}_T^\varepsilon - \bar{Q}_T\| &= (E' \times E'')|(E' \times E'')(1 - Z_T^\varepsilon(\omega', \omega''))|\bar{X}_t(\omega'), Y_t^\varepsilon(\omega', \omega''), t \leq T| \\
(4.7) \quad &\leq (E' \times E'')|1 - Z_T^\varepsilon(\omega', \omega'')|.
\end{aligned}$$

On the other hand, the representation for  $Z_T^\varepsilon(\omega', \omega'')$  can be rewritten in the form

$$Z_T^\varepsilon(\omega', \omega'') = \exp\left(M_T^\varepsilon - \frac{1}{2}\langle M^\varepsilon \rangle_T\right)$$

with a continuous martingale  $(M_t^\varepsilon)_{t \leq T}$  and its predictable quadratic variation  $(\langle M^\varepsilon \rangle_t)_{t \leq T}$ , where

$$\begin{aligned}
M_T^\varepsilon &= \int_0^T [A(\bar{X}_s(\omega'), Y_s^\varepsilon(\omega', \omega'')) - A(X_s^\varepsilon(\omega'), Y_s^\varepsilon(\omega', \omega''))] dW_s^y(\omega'') \\
\langle M^\varepsilon \rangle_T &= \int_0^T [A(\bar{X}_s, Y_s^\varepsilon(\omega', \omega'')) - A(X_s^\varepsilon(\omega'), Y_s^\varepsilon(\omega', \omega''))]^2 ds.
\end{aligned}$$

Due to the Lipschitz property of  $A(x, y)$  and (1.9),  $\langle M^\varepsilon \rangle_T \leq \text{const.} \sup_{s \leq T} |\bar{X}_s(\omega') - X_s^\varepsilon(\omega')| \rightarrow 0$  in probability as  $\varepsilon \rightarrow 0$ , which implies the same convergence for  $M_T^\varepsilon$  (see Problem 1.9.2 in [15]). Hence,  $Z_T^\varepsilon(\omega', \omega'') \rightarrow 1$  in probability as  $\varepsilon \rightarrow 0$  and so, by Scheffe's theorem [1] the right hand side of (4.7) converges to 0, i.e. (1.10) holds.

**5. Markov process as contamination. Examples.** In this Section, we consider Markov process  $(\eta_t)$  as a contamination, especially for Model 1.

**THEOREM 5.1.** *Let  $(\eta_t)$  be a Markov process with right continuous paths having limits from the left, satisfying (1.18). Then (1.12) and (1.13) hold.*

**PROOF.** Without loss of generality, one can assume that the initial point  $\eta_0$  is fixed:  $\eta_0 = y$ ; and the function  $f(z)$  from (1.12) and (1.13) satisfies the following conditions:

$$(5.1) \quad |f(z)| \leq 1, \quad \int_{\mathbf{R}} f(z) \mu(dz) = 0.$$

It is clear that, to check (1.12), it is sufficient to show that  $\lim_{\varepsilon \rightarrow 0} E\left(\int_0^t f(\eta_{s/\varepsilon}) ds\right)^2 = 0$ . By direct calculations we find

$$\begin{aligned}
&E\left(\int_0^t f(\eta_{s/\varepsilon}) ds\right)^2 \\
&= 2 \int_0^t \int_0^s E[E(f(\eta_{s/\varepsilon})|\eta_{s'/\varepsilon})f(\eta_{s'/\varepsilon})] ds' ds
\end{aligned}$$

$$\begin{aligned}
&\leq 2 \int_0^t \int_0^s E \left| E(f(\eta_{s/\varepsilon}) | \eta_{s'/\varepsilon}) f(\eta_{s'/\varepsilon}) \right| ds' ds \\
&= 2 \int_0^t \int_0^s \int_{\mathbf{R}} \left| \int_{\mathbf{R}} f(z) [\lambda(x, (s-s')/\varepsilon, dz) f(x)] \lambda(y, s'/\varepsilon, dx) \right| ds' ds \\
&= 2 \int_0^t \int_0^s \int_{\mathbf{R}} \left| \int_{\mathbf{R}} f(z) [\lambda(x, (s-s')/\varepsilon, dz) - \mu(dz)] f(x) \right| \lambda(y, s'/\varepsilon, dx) ds' ds \\
&\leq 2 \int_0^t \int_0^s \int_{\mathbf{R}} \|\lambda_{x, (s-s')/\varepsilon} - \mu\| \lambda(y, s'/\varepsilon, dx) ds' ds \\
&\leq 2 \int_0^t \int_0^s \int_{\mathbf{R}} \|\lambda_{x, (s-s')/\varepsilon} - \mu\| \mu(dx) ds' ds + 4 \int_0^t \int_0^s \|\lambda_{y, s'/\varepsilon} - \mu\| ds' ds. \\
&\rightarrow 0, \quad \varepsilon \rightarrow 0,
\end{aligned}$$

that is, (1.12) holds.

To check the validity (1.13), introduce the filtration  $(\mathcal{F}_t^\eta)_{t \geq 0}$ , generated by the paths of the process  $(\eta_t)$  and satisfying the general conditions. Also letting

$$(5.2) \quad N_s^\varepsilon = E(f(\eta_{t/\varepsilon}) | \mathcal{F}_{s/\varepsilon}^\eta),$$

introduce a square integrable martingale  $(N_s^\varepsilon, \mathcal{F}_{s/\varepsilon}^\eta)_{s \leq t}$ ,  $t \leq T$ , and consider the filtering problem for  $N_s^\varepsilon$ ,  $s \leq t$ , given observation  $(X_{s'}^\varepsilon, Y_{s'}^\varepsilon, s' \leq s)$ . By Theorem 8.5 in [14] we find (recall  $\mathcal{F}_s^\varepsilon = \sigma\{X_{s'}^\varepsilon, Y_{s'}^\varepsilon, s' \leq s\}$ )

$$\begin{aligned}
(5.3) \quad E[N_t^\varepsilon | \mathcal{F}_s^\varepsilon] &= E[N_t^\varepsilon | \mathcal{F}_0^\varepsilon] \\
&\quad + \int_0^s E \left( E[N_t^\varepsilon | \mathcal{F}_{s'/\varepsilon}^\eta] [a(X_{s'}^\varepsilon, Y_{s'}^\varepsilon, \eta_{s'/\varepsilon}) - a_{s'}^\varepsilon] | \mathcal{F}_{s'}^\varepsilon \right) d\widetilde{W}_{s'}^{x, \varepsilon} \\
&\quad + \int_0^s E \left( E[N_t^\varepsilon | \mathcal{F}_{s'/\varepsilon}^\eta] [A(X_{s'}^\varepsilon, Y_{s'}^\varepsilon, \eta_{s'/\varepsilon}) - A_{s'}^\varepsilon] | \mathcal{F}_{s'}^\varepsilon \right) d\widetilde{W}_{s'}^{y, \varepsilon},
\end{aligned}$$

where  $a_s^\varepsilon \equiv a_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon; \{X^\varepsilon, Y^\varepsilon\})$  and  $A_s^\varepsilon \equiv A_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon; \{X^\varepsilon, Y^\varepsilon\})$  are defined in Section 3. Since (1.13) is nothing but  $P - \lim_{\varepsilon \rightarrow 0} E[N_t^\varepsilon | \mathcal{F}_t^\varepsilon] = 0$ , we show that all terms on the right hand side of (5.3) with  $s = t$  converge to zero in probability as  $\varepsilon \rightarrow 0$ . Due to the second property of the function  $f$  (see (5.1))

$$\begin{aligned}
E[N_t^\varepsilon | \mathcal{F}_0^\varepsilon] &= \int_{\mathbf{R}} f(z) \lambda(y, t/\varepsilon, dz) \\
&= \int_{\mathbf{R}} f(z) [\lambda(y, t/\varepsilon, dz) - \mu(dz)] \\
&\leq \|\lambda_{y, t/\varepsilon} - \mu\| \rightarrow 0, \quad \varepsilon \rightarrow 0.
\end{aligned}$$

For brevity of notation use  $\alpha_{s'}^\varepsilon$  and  $\widetilde{W}_{s'}^\varepsilon$  for denoting any of  $a(X_{s'}^\varepsilon, Y_{s'}^\varepsilon, \eta_{s'/\varepsilon}) - a_{s'}^\varepsilon$  or  $A(X_{s'}^\varepsilon, Y_{s'}^\varepsilon, \eta_{s'/\varepsilon}) - A_{s'}^\varepsilon$  and Wiener processes  $\widetilde{W}_{s'}^{x, \varepsilon}$  or  $\widetilde{W}_{s'}^{y, \varepsilon}$  respectively. Then, it

remains to show that  $\int_0^t E[N_t^\varepsilon | \mathcal{F}_{s'/\varepsilon}^\eta] \alpha_{s'}^\varepsilon d\widetilde{W}_{s'}$  converges to zero in probability as  $\varepsilon \rightarrow 0$ . The last, due to Problem 1.9.2 in [15], holds provided that

$$(5.4) \quad P - \lim_{\varepsilon \rightarrow 0} \int_0^t \left\{ E \left( E[N_t^\varepsilon | \mathcal{F}_{s'/\varepsilon}^\eta] \alpha_{s'}^\varepsilon | \mathcal{F}_{s'}^\varepsilon \right) \right\}^2 ds' = 0.$$

Under the assumptions on the functions  $a(x, y, z)$ ,  $A(x, y, z)$ , the random variable  $|\alpha_{s'}^\varepsilon|$  is bounded above by the  $\mathcal{F}_{s'}^\varepsilon$ -measurable random variable  $\beta_{s'}^\varepsilon = \text{const.} (1 + \sup_{s'' \leq s'} |X_{s''}^\varepsilon| + \sup_{s'' \leq s'} |Y_{s''}^\varepsilon|)$ . By virtue of (1.15),  $\beta_{s'}^\varepsilon$  converges in probability as  $\varepsilon \rightarrow 0$  to  $\beta_{s'} = \text{const.} (1 + \sup_{s'' \leq T} |\overline{X}_{s''}| + \sup_{s'' \leq T} |\overline{Y}_{s''}|)$ . Therefore, instead of (5.4) it suffices to check

$$P - \lim_{\varepsilon \rightarrow 0} \int_0^t \left\{ E \left( E[N_t^\varepsilon | \mathcal{F}_{s'/\varepsilon}^\eta] | \mathcal{F}_{s'}^\varepsilon \right) \right\}^2 ds' = 0;$$

or, what is equivalent by virtue of the Markov property

$$\begin{aligned} E \left( E[N_t^\varepsilon | \mathcal{F}_{s'/\varepsilon}^\eta] | \mathcal{F}_{s'}^\varepsilon \right) &= E[N_t^\varepsilon | \mathcal{F}_{s'/\varepsilon}^\eta] \\ &= E[N_t^\varepsilon | \eta_{s'/\varepsilon}] \\ &= E[f(\eta_{t/\varepsilon}) | \eta_{s'/\varepsilon}], \end{aligned}$$

to check only

$$(5.5) \quad \lim_{\varepsilon \rightarrow 0} E \int_0^t \left\{ E[f(\eta_{t/\varepsilon}) | \eta_{s'/\varepsilon}] \right\}^2 ds' = 0.$$

On the other hand, due to (5.1) and the Markovian property of the process  $(\eta_t)$ , we get

$$\begin{aligned} E[f(\eta_{t/\varepsilon}) | \eta_{s'/\varepsilon}] &= \int_{\mathbf{R}} f(z) \lambda_{\eta_{(t-s')/\varepsilon}, s'/\varepsilon}(dz) \\ &= \int_{\mathbf{R}} f(z) [\lambda_{\eta_{(t-s')/\varepsilon}, s'/\varepsilon}(dz) - \mu(dz)] \\ &\leq \|\lambda_{\eta_{(t-s')/\varepsilon}, s'/\varepsilon} - \mu\| \end{aligned}$$

and so

$$\begin{aligned} E \int_0^t \left\{ E[f(\eta_{t/\varepsilon}) | \eta_{s'/\varepsilon}] \right\}^2 ds' &\leq \int_0^t \int_{\mathbf{R}} \|\lambda_{z, (t-s')/\varepsilon} - \mu\|^2 \lambda_{y, s'/\varepsilon}(dz) ds' \\ &\leq 2 \int_0^t \int_{\mathbf{R}} \|\lambda_{z, (t-s')/\varepsilon} - \mu\| \mu(dz) ds' \\ &\quad + 4 \int_0^t \|\lambda_{y, s'/\varepsilon} - \mu\| ds' \\ &\rightarrow 0 \text{ as } \varepsilon \rightarrow 0 \end{aligned}$$

that is, (5.5) holds.

Now we give two examples of ergodic Markov processes for which the assumptions of Theorem 5.1 hold.

EXAMPLE 1. Let  $\eta_t$  be a diffusion Markov process defined by Ito's equation (with respect to the Wiener process  $W_t^\eta$ )

$$d\eta_t = a(\eta_t)dt + b(\eta_t)dW_t^\eta,$$

where the drift  $a(z)$  and the diffusion parameter  $b(z)$  are assumed to be continuously differentiable, having bounded derivatives, and to satisfy the following conditions: there exist constants  $\ell$  and  $L$  such that

$$(5.6) \quad 0 < \ell \leq b^2(z) \leq L, \quad \liminf_{|z| \rightarrow \infty} a(z) \text{sign } z < 0.$$

It is well known that the transition probability  $\lambda(y, t, dz)$  of  $(\eta_t)$  obeys a density  $p(y, t, z)$  (with respect to Lebesgue measure  $dz$ ) being a solution of the forward Fokker-Planck-Kolmogorov equation

$$(5.7) \quad \frac{\partial p(y, t, z)}{\partial t} = \mathcal{L}^* p(y, t, z),$$

where  $\mathcal{L}^*\{.\} = -\frac{\partial}{\partial z}(a(z)\{.\}) + \frac{1}{2}\frac{\partial^2}{\partial z^2}(b^2(z)\{.\})$  (with references for existence of the unique solution of (5.7) see [12]). Under assumptions (5.6) the invariant measure  $\mu(dz)$  exists and obeys a density  $m(z)$  (with respect to  $dz$ ) which is a solution of an ordinary differential equation  $\mathcal{L}^*m(z) = 0$ . It is well known (see [12], Ch.4, §9, Lemma 9.5; Ch.3, §8, Example 2 from [6]) that the solution of Cauchy's problem for the partial differential equation (5.7) is stabilized in the sense that for every fixed  $y, z$  we have  $\lim_{t \rightarrow \infty} p(y, t, z) = m(z)$ . Then, taking into account that  $\int_{\mathbf{R}} p(y, t, z) dz = 1$ ,  $\int_{\mathbf{R}} m(z) dz = 1$ , by virtue of Scheffe's theorem [1]

$$\lim_{t \rightarrow \infty} \int_{\mathbf{R}} |p(y, t, z) - m(z)| dz = 0,$$

which is nothing but (1.18).

EXAMPLE 2. Let the homogeneous Markov process  $\eta_t$  take values in the finite space  $\{\alpha_1, \dots, \alpha_N\}$  with  $\eta_0 = \alpha_j$ . Denote by  $\mathcal{P}$  its matrix of the transition intensities, and by  $p_j(t) = (p_{1,j}(t), \dots, p_{N,j}(t))$  the vector of transition probabilities:  $p_{i,j}(t)$  is the transition probability from  $\alpha_j$  to  $\alpha_i$  over time  $t$ . The vector  $p_j(t)$  is defined by the Fokker-Planck-Kolmogorov equation

$$(5.8) \quad \frac{dp_j(t)}{dt} = p_j(t)\mathcal{P}$$

subject to  $p_j(0)$  being the vector with zero components, excepting the component indexed by  $j$  which is one. Assume that 0 is a simple eigenvalue of the matrix  $\mathcal{P}$ . Then there exists a unique invariant distribution,  $p = (p_1, \dots, p_N)$  satisfying  $p\mathcal{P} = 0$ . Due to the uniqueness of the invariant distribution we have  $\lim_{t \rightarrow \infty} p_j(t) = p$ ,  $j = 1, \dots, N$  which implies (1.18) since

$$\|\lambda_{y,t} - \mu\| = \sum_{j=1}^N |p_j(t) - p|.$$

Also we give an example which shows that if condition (1.13) fails then the asymptotic optimality for Model 1 fails too.

**EXAMPLE 3.** Take deterministic contamination, say,  $\eta_t = \sin(t)$  and put

$$\begin{aligned} dX_t^\varepsilon &= dW_t^x, \\ dY_t^\varepsilon &= \sin(t/\varepsilon)X_t^\varepsilon dt + dW_t^y \end{aligned}$$

subject to  $X_0 = 0, Y_0 = 0$ . Evidently assumption (1.13) fails. In contrast, (1.12) remains true:  $\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t \sin(s) ds = 0$  and so (1.15) holds with

$$d\bar{X}_t = dW_t^x \quad \text{and} \quad d\bar{Y}_t = dW_t^y.$$

which implies  $\bar{\pi}_t(Y^\varepsilon) \equiv 0$ . Hence  $E(X_t^\varepsilon - \bar{\pi}_t(Y^\varepsilon))^2 \equiv t$ . On the other hand,  $\pi_t^\varepsilon(Y^\varepsilon)$ , being defined by the Kalman filter, has the mean square error  $P^\varepsilon(t) = E(X_t^\varepsilon - \pi_t^\varepsilon(Y^\varepsilon))^2$  which is given by the Riccati equation:

$$\dot{P}^\varepsilon(t) = 1 - \sin^2(t/\varepsilon)(P^\varepsilon(t))^2(t)$$

subject to  $P^\varepsilon(0) = 0$ . Due to the Bogolubov averaging principle [2],  $\lim_{\varepsilon \rightarrow 0} P^\varepsilon(t) = P^\circ(t)$ , where  $P^\circ(t)$  is a solution of Riccati's differential equation

$$\dot{P}^\circ(t) = 1 - c^2(P^\circ(t))^2(t)$$

subject to  $P^\circ(0) = 0$ , where  $c^2 = \lim_{\varepsilon \rightarrow 0} \int_0^1 \sin^2(s/\varepsilon) ds = \frac{1}{2}$ . Evidently, for any  $t > 0$   $P^\circ(t) < t$ , i.e., the asymptotic optimality for the filtering estimate  $\bar{\pi}_t(Y^\varepsilon)$  is lost.

**Acknowledgment.** The authors gratefully acknowledge the careful review by the anonymous referees, who also pointed out a mistake in the original version. Due to their comments the paper could be significantly improved.

## References

- [1] Billingsley, P. (1968). *Convergence of probability measures*, John Wiley & Sons Inc., New York-London-Sydney-Toronto.
- [2] Bogolubov, N. and Mitropolskii, A. (1961). *Asymptotics methods in Non-linear Mechanics*. Gordon and Breach, New York.
- [3] M.I. Freidlin and A.D. Wentzell *Random Perturbations of Dynamical Systems* Springer-Verlag. 1984
- [4] Goggin, E. (1992). Convergence of Filters with Application to the Kalman-Busy Case. *IEEE Transactions on Information Theory*. **38**, No 3 1091-1100.
- [5] Goggin, E. (1994). Convergence in Distribution of Conditional Expectations. *The Annals of Probability*. **22**, No 2 1097-1114.
- [6] Khasminskii, R.Z. (1980). *Stochastic stability of differential equations*. Sijthoff & Noordhoff.
- [7] Khasminskii, R.Z. (1993). The averaging principle for parabolic and elliptic differential equations and Markov processes with small diffusion. *Theory of Probability and its Applications*. **3** 1-21.
- [8] H.J. Kushner, H.J. (1967). Dynamical equations for optimal nonlinear filtering. *J. Differential equations*. **3** 179-190.
- [9] Kushner, H.J. (1990), *Weak convergence method and singularly perturbed stochastic control and filtering problems*. Birkhauser, New York.
- [10] Kushner, H.J. and Runggaldier, W.J. (1987). Filtering and control for wide bandwidth noise driven systems. *IEEE Transactions, AC32*. 123-133.
- [11] Jacod, J and A.N.Shiryaev, A.N. (1987). *Limit Theorems for Stochastic Processes*. Springer-Verlag, New York.
- [12] Ladyzhenskaya, A., Solonnikov, V.A. and Ural'tseva, N.N. (1967). *Linear and quasilinear parabolic equations*. Nauka. Moscow.
- [13] Liptser, R.Sh.and Runggaldier, W.J. (1991). On diffusion approximations for filtering. *Stochastic processes and their applications*. **38** 205-238.
- [14] Liptser, R.Sh. and Shiryaev, A.N. (1977, 1978). *Statistics of Random Processes* **1**, **2**. Springer-Verlag, New York.

- [15] Liptser, R.Sh. and Shiryaev, A.N. (1989). *Theory of Martingales*. Kluwer, Dordrecht.
- [16] Rozovskii, B.L. (1991). A simple proof of uniqueness for Kushner and Zakai equations. In book: *Stochastic Analysis*. 449-458.
- [17] Zakai, M. (1969). On the optimal filtering of diffusion processes. *Wahrsh. Verw. Geb.* **11** 205-238.

M. L. KLEPTSINA  
MOSCOW STATE UNIVERSITY OF RAILWAY ENGINEERING  
RUSSIA  
*e-mail*: asarin@ippi.ac.msk.su

R. SH. LIPTSER  
DEPARTMENT OF ENGINEERING  
TEL AVIV UNIVERSITY  
TEL AVIV 69978 ISRAEL  
*e-mail*: liptser@tau.ac.il

A.P.SEREBROVSKI  
INSTITUTE FOR INFORMATION TRANSMISSION SCIENCES  
MOSCOW 101447  
RUSSIA  
*e-mail*: serebrov@ippi.ac.msk.su