The Annals of Applied Probability
1997, Vol. 7, No. 4, 917–934

NONLINEAR FILTERING PROBLEM WITH CONTAMINATION[1]

BY M. L. KLEPTSINA, R. SH. LIPTSER[2] AND A. P. SEREBROVSKI

Moscow State University of Railway Engineering, Tel Aviv University and Institute for Information Transmission Sciences, Moscow

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

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] and [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 a continuous random process $(X_t^\varepsilon, Y_t^\varepsilon)_{t\geq 0}$ with $X_t^\varepsilon,\ Y_t^\varepsilon \in \mathrm{R}$, where X_t^ε represents an unobservable signal and Y_t^ε 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 $\varepsilon \to 0$ $(X_t^\varepsilon, Y_t^\varepsilon)_{t\geq 0}$ converges (in some sense) to the limit $(\overline{X}_t, \overline{Y}_t)_{t\geq 0}$, 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

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

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 $\overline{\pi}_t(Y) = E[u(\overline{X}_t)|\overline{Y}_s,\ s \leq t]_{\overline{Y}_s \equiv Y_s}$ ($Y_s,\ s \geq 0$ is a continuous function),

$$\lim_{\varepsilon\to 0} E\big(u(X_t^\varepsilon) - \overline{\pi}_t(Y^\varepsilon)\big)^2 = E\big(u(\overline{X}_t) - \overline{\pi}_t(\overline{Y})\big)^2, \tag{1.2}$$

Received November 1994; revised April 1997.

[1]Research supported by International Science Foundation Researches and in part by AT&T Bell Lab.

[2]Research partially supported by ONR Grant N00014-96-10413.

AMS 1991 *subject classification*. 60G35.

Key words and phrases. Diffusions, averaging principle, convergence of total variation.

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

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

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

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

or, equivalently,

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

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

$$\lim_{\varepsilon \to 0} E\big(u(X_t^\varepsilon) - \pi_t^\varepsilon(Y^\varepsilon)\big)^2 = E\big(u(\overline{X}_t) - \overline{\pi}_t(\overline{Y})\big)^2. \tag{1.4}$$

In [9], it was mentioned that $\overline{\pi}_t(Y^\varepsilon)$ is not necessarily asymptotically optimal in sense (1.4) although it 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 nonoptimal (!) filtering estimate $\widetilde{\pi}_t^\varepsilon$ such that

$$\lim_{\varepsilon \to 0} E\big(u(X_t^\varepsilon) - \widetilde{\pi}_t^\varepsilon\big)^2 < \lim_{\varepsilon \to 0} E\big(u(X_t^\varepsilon) - \overline{\pi}_t(Y^\varepsilon)\big)^2.$$

It is clear that the problem of the asymptotic optimality of 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 $\overline{\pi}_t(\overline{Y})$:

$$\pi_t^\varepsilon(Y^\varepsilon) \overset{\text{law}}{\to} \pi_t(\overline{Y}), \qquad \varepsilon \to 0. \tag{1.5}$$

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

$$Q_T^\varepsilon \overset{w}{\to} \overline{Q}_T, \qquad \varepsilon \to 0, \tag{1.6}$$

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 law to the corresponding likelihood from the Bayes formula for $\overline{\pi}(\overline{Y})$.

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

$$\lim_{\varepsilon\to 0} \|Q^\varepsilon_T - \overline{Q}_T\| = 0 \tag{1.7}$$

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 $t \le T$ and for some $\delta > 0$,

$$\begin{aligned} \sup_{\varepsilon\le\varepsilon_0} E\big|u(X^\varepsilon_t)\big|^{2+\delta} &< \infty, \\ E\big|u(\overline{X}_t)\big|^{2+\delta} &< \infty, \\ \sup_{\varepsilon\le\varepsilon_0} E\Big[E\big(|u(\overline{X}_t)|^{2+\delta}\big|\overline{Y}_s, s\le t\big)_{|_{\overline{Y}\equiv Y^\varepsilon}}\Big] &< \infty, \end{aligned} \tag{1.8}$$

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

Let $(X^\varepsilon_t, Y^\varepsilon_t)_{t\le T}$ and $(\overline{X}_t, \overline{Y}_t)_{t\le T}$ be defined on the same probability space. In this case, the distribution $\overline{Q}^\varepsilon_T$ of the process $(\overline{X}_t, Y^\varepsilon_t)_{t\le T}$ is well defined. Suppose (1.7) fails while

$$P - \lim_{\varepsilon\to 0} \sup_{t\le T} |X^\varepsilon_t - \overline{X}_t| = 0. \tag{1.9}$$

THEOREM 1.2. *Assume (1.9) and*

$$\lim_{\varepsilon\to 0} \|\overline{Q}^\varepsilon_T - \overline{Q}_T\| = 0. \tag{1.10}$$

Then for any continuous function u satisfying (1.8), the filtering estimate $\overline{\pi}_t(Y^\varepsilon)$, $t \le T$, is asymptotically optimal in the mean square sense and (1.2) and (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. The process $(X^\varepsilon_t, Y^\varepsilon_t)_{t\ge 0}$ is defined by Itô's stochastic equations with respect to independent Wiener processes $(W^x_t)_{t\ge 0}$ and $(W^y_t)_{t\ge 0}$:

$$\begin{aligned} dX^\varepsilon_t &= a(X^\varepsilon_t, Y^\varepsilon_t, \eta_{t/\varepsilon})\,dt + dW^x_t, \\ dY^\varepsilon_t &= A(X^\varepsilon_t, Y^\varepsilon_t, \eta_{t/\varepsilon})\,dt + dW^y_t, \end{aligned} \tag{1.11}$$

subject to the initial point (X_0, Y_0) which is a random vector independent of ε, where $\eta_{t/\varepsilon}$ is a contamination (random or deterministic), affecting drifts. The 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). The process $(\eta_t)_{t\geq 0}$ 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

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

and

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

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

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

subject to the initial point (X_0, Y_0). The function $a = a(x, z)$ is continuous and Lipschitz continuous in x uniformly in z; $a(0, z)$ is bounded. The function $A = A(x, y)$ is Lipschitz continuous. The process (η_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 $T > 0$,

$$P - \lim_{\varepsilon\to 0}\Big(\sup_{t\leq T} |X_t^\varepsilon - \overline{X}_t| + \sup_{t\leq T} |Y_t^\varepsilon - \overline{Y}_t|\Big) = 0, \tag{1.15}$$

where $(\overline{X}_t, \overline{Y}_t)$ is a Markovian diffusion defined by Itô's equations. For Model 1,

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

and for Model 2,

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

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

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

and

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

Theorems 1.1 and 1.2 imply Theorem 1.3.

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 $\overline{\pi}_t(Y^\varepsilon)$, $t \geq 0$, 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 η_t, stationary processes, ergodic Markov processes, and so on, while the ergodic property (1.13) is not trivial to verify. Nevertheless, we examine it for a homogeneous Markov process η_t, having the unique invariant measure μ such that its transition probability $\lambda_{y,t} = \lambda(y, t, dz)$ converges in the total variation norm to μ:

$$\lim_{t\to\infty} \|\lambda_{y,\,t} - \mu\| = 0 \qquad \forall\ y \in \mathrm{R}. \tag{1.18}$$

Also in this case we have

$$\overline{a}(x, y) = \int_{\mathrm{R}} a(x, y, z)\mu(dz),$$

$$\overline{A}(x, y) = \int_{\mathrm{R}} A(x, y, z)\mu(dz),$$

$$\overline{a}(x) = \int_{\mathrm{R}} Ra(x, z)\mu(dz).$$

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 $\overline{\pi}_t(Y^\varepsilon)$ 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 the following lemma.

LEMMA 2.1. *Under the assumptions of Theorem 1.1 and 1.2 for any $t \leq T$, we have the following:*

$$\lim_{\varepsilon\to 0} \int \big(u(X_t) - \overline{\pi}_t(Y)\big)^2\, dQ_T^\varepsilon = \int \big(u(X_t) - \overline{\pi}_t(Y)\big)^2\, d\overline{Q}_T,$$

$$\liminf_{\varepsilon\to 0} \int \big(u(X_t) - \pi_t^\varepsilon(Y)\big)^2\, dQ_T^\varepsilon \geq \int \big(u(X_t) - \overline{\pi}_t(Y)\big)^2\, d\overline{Q}_T.$$

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

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

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

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

Assume there exists a constant, say l, such that

$$|u| \leq l. \tag{2.2}$$

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

Assume (1.7). Then

$$\begin{aligned}\left|\int \big(u(X_t) - \overline{\pi}_t(Y)\big)^2 \, dQ_T^\varepsilon - \int \big(u(X_t) - \overline{\pi}_t(Y)\big)^2 \, d\overline{Q}_T\right| \\ \leq 4l^2 ||Q_T^\varepsilon - \overline{Q}_T|| \to 0, \qquad \varepsilon \to 0.\end{aligned}$$

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

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

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

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

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

$$\int \big(u(X_t) - \pi_t^\varepsilon(Y)\big)^2 \, d\overline{Q}_T \geq \int \big(u(X_t) - \overline{\pi}_t(Y)\big)^2 \, d\overline{Q}_T. \tag{2.5}$$

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

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

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

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

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

$$\begin{aligned}\int \big(u(X_t)-\pi_t^\varepsilon(Y)\big)^2\,dQ_T^\varepsilon \ge {}&\int \big(u(X_t)-\overline{\pi}_t(Y)\big)^2\,d\overline{Q}_T\\ &-4lE|u(X_t^\varepsilon)-u(\overline{X}_t)|-4l^2\|\overline{Q}_T^\varepsilon-\overline{Q}_T\|\\ \to{}&\int \big(u(X_t)-\overline{\pi}_t(Y)\big)^2 d\overline{Q}_T,\qquad \varepsilon\to 0.\end{aligned}$$

Thus, for bounded functions u the lemma is proved.

Assume u is an unbounded function for which (1.8) holds. For $n\ge 1$, put $u_n(x)=g_n(u(x))$, where

$$g_n(x)=\begin{cases}x, & |x|\le n,\\ n\,\mathrm{sign}\,x, & |x|>n.\end{cases}$$

Also put $\pi_t^{\varepsilon,n}(Y^\varepsilon)=E(u_n(X_t^\varepsilon)|\mathscr{F}_t^{Y^\varepsilon})$ and $\overline{\pi}_t^n(\overline{Y})=E(u_n(\overline{X}_t)|\mathscr{F}_t^{\overline{Y}})$. By the statements of the lemma proved above for bounded u we find, for every fixed n,

$$\begin{aligned}\lim_{\varepsilon\to 0}\int\big[u_n(X_t)-\overline{\pi}_t^n(Y)\big]^2\,dQ_T^\varepsilon &= \int\big[u_n(X_t)-\overline{\pi}_t^n(Y)\big]^2\,d\overline{Q}_T,\\ \liminf_{\varepsilon\to 0}\int\big[u_n(X_t)-\pi_t^{\varepsilon,n}(Y)\big]^2\,dQ_T^\varepsilon &\ge \int\big[u_n(X_t)-\overline{\pi}_t^n(Y)\big]^2\,d\overline{Q}_T.\end{aligned}\tag{2.7}$$

Evidently, the statements of the lemma hold if

$$\begin{aligned}\lim_n\limsup_{\varepsilon\to 0} E\big[u(X_t^\varepsilon)-u_n(X_t^\varepsilon)\big]^2&=0,\\ \lim_n E\big[u(\overline{X}_t)-u_n(\overline{X}_t)\big]^2&=0\end{aligned}\tag{2.8}$$

and

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

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

$$E\big[u(X^\varepsilon_t)-u_n(X^\varepsilon_t)\big]^2 \le \frac{1}{n^\delta}E|u(X^\varepsilon_t)|^{2+\delta},$$
$$E\big[u(\overline{X}_t)-u_n(\overline{X}_t)\big]^2 \le \frac{1}{n^\delta}E|u(\overline{X}_t)|^{2+\delta}$$

and

$$\begin{aligned}E\big[\overline{\pi}_t(Y^\varepsilon)-\overline{\pi}^n_t(Y^\varepsilon)\big]^2 &= E\big[E\big(u(\overline{X}_t)-u_n(\overline{X}_t)\big|\overline{Y}_s,\ s\le t\big)_{\overline{Y}\equiv Y^\varepsilon}\big]^2\\ &\le \frac{1}{n^\delta}E\Big[E\Big(|u(\overline{X}_t)|^{2+\delta}\Big|\overline{Y}_s, s\le t\Big)_{\overline{Y}\equiv Y^\varepsilon}\Big],\\ E\big[\pi^\varepsilon_t(Y^\varepsilon)-\pi^{\varepsilon,n}_t(Y^\varepsilon)\big]^2 &\le E\Big[E\Big(u(X^\varepsilon_t)-u_n(X^\varepsilon_t)\Big|Y^\varepsilon_s, s\le t\Big)\Big]^2\\ &\le \frac{1}{n^\delta}E|u(X^\varepsilon_t)|^{2+\delta};\end{aligned}$$

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

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

$$\lim_{\varepsilon\to 0} E\exp(\mathrm{i}v\pi^\varepsilon_t(Y^\varepsilon)) = E\exp(\mathrm{i}v\overline{\pi}_t(\overline{Y})), \qquad v\in\mathrm{R},\ \mathrm{i}=\sqrt{-1}.$$

To this end, we use the following estimate:

$$\begin{aligned}&\big|E\exp(\mathrm{i}v\pi^\varepsilon_t(Y^\varepsilon)) - E\exp(\mathrm{i}v\overline{\pi}_t(\overline{Y}))\big|\\ &\quad\le E\big|\exp(\mathrm{i}v\pi^\varepsilon_t(Y^\varepsilon)) - \exp(\mathrm{i}v\overline{\pi}_t(Y^\varepsilon))\big|\\ &\qquad+\big|E\big(\exp(\mathrm{i}v\overline{\pi}_t(Y^\varepsilon)) - \exp(\mathrm{i}v\overline{\pi}_t(\overline{Y}))\big)\big|\end{aligned} \tag{2.10}$$

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

3. Model 1. It is sufficient to show that (1.12), (1.13) imply (1.7) for every $T>0$. Let $(\Omega,\mathscr{F},\mathrm{F}^\varepsilon=(\mathscr{F}^\varepsilon_t)_{t\ge 0},P)$ be a stochastic basis with the filtration F^ε generated by paths of $(X^\varepsilon_t,Y^\varepsilon_t)_{t\ge 0}$. Denote by $a^\varepsilon_s(x,y;\{X^\varepsilon,Y^\varepsilon\})$ the F^ε-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.,

$$\begin{aligned}a^\varepsilon_s(x,y;\{X^\varepsilon,Y^\varepsilon\}) &= E\big(a(x,y,\eta_{s/\varepsilon})|\mathscr{F}^\varepsilon_s\big),\\ a^\varepsilon_s(X^\varepsilon_s,Y^\varepsilon_s;\{X^\varepsilon,Y^\varepsilon\}) &= E\big(a(X^\varepsilon_s,Y^\varepsilon_s,\eta_{s/\varepsilon})|\mathscr{F}^\varepsilon_s\big).\end{aligned} \tag{3.1}$$

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

$$P - \lim_{\varepsilon \to 0} \int_0^T \left[a_s^\varepsilon(x, y; \{X^\varepsilon, Y^\varepsilon\}) - \overline{a}(x, y)\right]^2 ds = 0. \tag{3.2}$$

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

$$P - \lim_{\varepsilon \to 0} \int_0^T \left[a_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon; \{X^\varepsilon, Y^\varepsilon\}) - \overline{a}(X_s^\varepsilon, Y_s^\varepsilon)\right]^2 ds = 0. \tag{3.3}$$

Denote by $[x]$ the greatest integer function and define the following random processes: $X_t^{\varepsilon, n} = X_{[nt]/n}^\varepsilon$ and $X_t^{\varepsilon, n, m} = [m X_t^{\varepsilon, n}]/m$. Put $X_T^{\varepsilon *} = \sup_{t \le T} |X_t^\varepsilon|$. 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 $m, n \ge 1$ on the set $\{X_T^{\varepsilon, *} + Y_T^{\varepsilon, *} \le C\}$,

$$\int_0^T \left[a_s^\varepsilon(X_s^{\varepsilon, n, m}, Y_s^{\varepsilon, n, m}; \{X^\varepsilon, Y^\varepsilon\}) - \overline{a}(X_s^{\varepsilon, n, m}, Y_s^{\varepsilon, n, m})\right]^2 ds \to 0 \tag{3.4}$$

in probability as $\varepsilon \to 0$. 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 $\{X_T^{\varepsilon, *} + Y_T^{\varepsilon, *} \le C\}$ have a finite number of paths, which is independent of ε, and so (3.4) holds if for any continuous functions $x_s, y_s,\ 0 \le s \le T$ 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 $Y_{j/n}^{\varepsilon, m}$,

$$P - \lim_{\varepsilon \to 0} \int_0^T \left[a_s^\varepsilon(x_s^{n, m}, y_s^{n, m}; \{X^\varepsilon, Y^\varepsilon\}) - \overline{a}(x_s^{n, m}, y_s^{n, m})\right]^2 ds = 0. \tag{3.5}$$

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

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

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

$$\lim_{C \to \infty} \limsup_{\varepsilon \to 0} P(X_T^{\varepsilon, *} + Y_T^{\varepsilon, *} \ge C) = 0 \tag{3.6}$$

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

$$\begin{gathered}
\int_0^T \left[a_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon; \{X^\varepsilon, Y^\varepsilon\}) - a_s^\varepsilon(X_s^{\varepsilon, n, m}, Y_s^{\varepsilon, n, m}; \{X^\varepsilon, Y^\varepsilon\})\right]^2 ds \to 0, \\
\int_0^T \left[\overline{a}(X_s^\varepsilon, Y_s^\varepsilon) - \overline{a}(X_s^{\varepsilon, n, m}, Y_s^{\varepsilon, n, m})\right]^2 ds \to 0
\end{gathered} \tag{3.7}$$

in probability as the limit $\lim_{n,m} \limsup_{\varepsilon\to 0}$ 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 $\overline{a}(x, y)$ inherit the Lipschitz property in x, y with an absolute constant, which implies (3.7) under

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

The last holds by virtue of (1.15) too.

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

$$P - \lim_{\varepsilon\to 0} \int_0^T \left[A_s^\varepsilon(X_s^\varepsilon, Y_s^\varepsilon; \{X^\varepsilon, Y^\varepsilon\}) - \overline{A}(X_s^\varepsilon, Y_s^\varepsilon)\right]^2 ds = 0. \tag{3.8}$$

For brevity of notation, hereafter a_s^ε and A_s^ε 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

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

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

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

Hence, by Theorems 7.19 and 7.20 in [14], for every $T > 0$ the measures Q_T^ε 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 $d\overline{Q}_T/dQ_T^\varepsilon$ at a point "$X^\varepsilon, Y^\varepsilon$" is given by the formula (see Theorems 7.19 and 7.20 in [14])

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

Taking into account that X_t^ε and Y_t^ε satisfy (3.10), one can rewrite (3.11) in the form

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

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

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

and where

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

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

We show that (1.7) holds. Since $\|Q_T^\varepsilon - \overline{Q}_T\| = E|1 - (d\overline{Q}_T/dQ_T^\varepsilon)(X^\varepsilon, Y^\varepsilon)|$, (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 $(\overline{X}_t, \overline{Y}_t)$ is defined by the Itô equations

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

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

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

Now define a random process

$$
\begin{aligned}
Z_t^\varepsilon(\omega,\omega') = \exp\biggl(\int_0^t \bigl[&A(\overline{X}_s(\omega'), Y_s^\varepsilon(\omega',\omega''))\\
&- A(X_s^\varepsilon(\omega'), Y_s^\varepsilon(\omega',\omega''))\bigr]\,dW_s^y(\omega'')\\
&- \tfrac{1}{2}\int_0^t \bigl[A(\overline{X}_s(\omega'), Y_s^\varepsilon(\omega',\omega''))\\
&- A(X_s^\varepsilon(\omega'), Y_s^\varepsilon(\omega',\omega''))\bigr]^2\,ds\biggr)
\end{aligned} \tag{4.3}
$$

and show that for every $T>0$,

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

For fixed ω' and T, the distributions $R_T^\varepsilon(\omega')$, $\overline{R}_T^\varepsilon(\omega')$ of the processes $(Y_t^\varepsilon(\omega',\omega''))$, $(\overline{Y}_t(\omega',\omega''))$ 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

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

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

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

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

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

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

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

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

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

$$
\frac{d\overline{Q}_T}{d\overline{Q}_T^{\varepsilon}}(\overline{X}, Y^\varepsilon) = (E'\times E'')\Bigl(Z_T^\varepsilon(\omega',\omega'')\big|\overline{X}_t, Y_t^\varepsilon, t\le T\Bigr)\,P'\times P''\text{-a.s.} \tag{4.6}
$$

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

$$
\begin{aligned}
&\|\overline{Q}^{\varepsilon}_T - \overline{Q}_T\| \\
&= (E' \times E'')\big|(E' \times E'')\big(1 - Z^{\varepsilon}_T(\omega', \omega'')\big|\overline{X}_t(\omega'), Y^{\varepsilon}_t(\omega', \omega''), t \le T\big)\big| \\
&\le (E' \times E'')\big|1 - Z^{\varepsilon}_T(\omega', \omega'')\big|.
\end{aligned} \tag{4.7}
$$

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

$$
Z^{\varepsilon}_T(\omega', \omega'') = \exp\big(M^{\varepsilon}_T - \tfrac{1}{2}\langle M^{\varepsilon}\rangle_T\big)
$$

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

$$
M^{\varepsilon}_T = \int_0^T [A(\overline{X}_s(\omega'), Y^{\varepsilon}_s(\omega', \omega'')) - A(X^{\varepsilon}_s(\omega'), Y^{\varepsilon}_s(\omega', \omega''))]\, dW^y_s(\omega''),
$$

$$
\langle M^{\varepsilon}\rangle_T = \int_0^T \big[A(\overline{X}_s, Y^{\varepsilon}_s(\omega', \omega'')) - A(X^{\varepsilon}_s(\omega'), Y^{\varepsilon}_s(\omega', \omega''))\big]^2\, ds.
$$

Due to the Lipschitz property of $A(x, y)$ and (1.9),

$$
\langle M^{\varepsilon}\rangle_T \le \text{const.} \sup_{s\le T} |\overline{X}_s(\omega') - X^{\varepsilon}_s(\omega')| \to 0
$$

in probability as $\varepsilon \to 0$, which implies the same convergence for M^{ε}_T (see Problem 1.9.2 in [15]). Hence, $Z^{\varepsilon}_T(\omega', \omega'') \to 1$ in probability as $\varepsilon \to 0$ and so, by Scheffe's theorem [1] the right-hand side of (4.7) converges to 0; that is, (1.10) holds.

5. Markov process as contamination: examples.

In this section, we consider a Markov process (η_t) as a contamination, especially for Model 1.

THEOREM 5.1. *Let (η_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 η_0 is fixed: $\eta_0 = y$; and the function $f(z)$ from (1.12) and (1.13) satisfies the following conditions:

$$
|f(z)| \le 1, \qquad \int_{\mathbb{R}} f(z)\mu(dz) = 0. \tag{5.1}
$$

It is clear that, to check (1.12), it is sufficient to show that

$$
\lim_{\varepsilon\to 0} E\bigg(\int_0^t f\big(\eta_{s/\varepsilon}\big)\, ds\bigg)^2 = 0.
$$

By direct calculations we find

$$
\begin{aligned}
&E\Big(\int_0^t f(\eta_{s/\varepsilon})\,ds\Big)^2\\
&\quad = 2\int_0^t\int_0^s E\big[E(f(\eta_{s/\varepsilon})|\eta_{s'/\varepsilon})f(\eta_{s'/\varepsilon})\big]\,ds'\,ds\\
&\quad \le 2\int_0^t\int_0^s E\Big|E(f(\eta_{s/\varepsilon})|\eta_{s'/\varepsilon})f(\eta_{s'/\varepsilon})\Big|\,ds'\,ds\\
&\quad = 2\int_0^t\int_0^s\int_{\mathrm{R}}\Big|\int_{\mathrm{R}} f(z)[\lambda(x,(s-s')/\varepsilon,dz)f(x)\Big|\lambda(y,s'/\varepsilon,dx)\,ds'\,ds\\
&\quad = 2\int_0^t\int_0^s\int_{\mathrm{R}}\Big|\int_{\mathrm{R}} f(z)[\lambda(x,(s-s')/\varepsilon,dz)-\mu(dz)]f(x)\Big|\lambda(y,s'/\varepsilon,dx)\,ds'\,ds\\
&\quad \le 2\int_0^t\int_0^s\int_{\mathrm{R}} ||\lambda_{x,(s-s')/\varepsilon}-\mu||\lambda(y,s'/\varepsilon,dx)\,ds'\,ds\\
&\quad \le 2\int_0^t\int_0^s\int_{\mathrm{R}} ||\lambda_{x,(s-s')/\varepsilon}-\mu||\mu(dx)\,ds'\,ds + 4\int_0^t\int_0^s ||\lambda_{y,s'/\varepsilon}-\mu||\,ds'\,ds\\
&\quad \to 0, \qquad \text{as } \varepsilon\to 0;
\end{aligned}
$$

that is, (1.12) holds.

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

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

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

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

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 $P-\lim_{\varepsilon\to 0} E[N_t^\varepsilon|\mathscr{F}_t^\varepsilon] = 0$, we show that all terms on the right-hand side of (5.3) with $s = t$ converge to zero in probability as $\varepsilon \to 0$. Due to the second property of the function f [see (5.1)],

$$
\begin{aligned}
E\big[N_t^\varepsilon|\mathscr{F}_0^\varepsilon\big] &= \int_{\mathrm{R}} f(z)\lambda(y, t/\varepsilon, dz)\\
&= \int_{\mathrm{R}} f(z)[\lambda(y, t/\varepsilon, dz) - \mu(dz)]\\
&\le ||\lambda_{y,\,t/\varepsilon} - \mu|| \to 0, \qquad \text{as } \varepsilon \to 0.
\end{aligned}
$$

For brevity of notation, use $\alpha^\varepsilon_{s'}$ and $\widetilde{W}_{s'}$ for denoting any of $a(X^\varepsilon_{s'}, Y^\varepsilon_{s'}, \eta_{s'/\varepsilon}) - a^\varepsilon_{s'}$ or $A(X^\varepsilon_{s'}, Y^\varepsilon_{s}, \eta_{s'/\varepsilon}) - A^\varepsilon_{s'}$ and Wiener processes $\widetilde{W}^{x,\varepsilon}_{s'}$ or $\widetilde{W}^{y,\varepsilon}_{s'}$ respectively. Then, it remains to show that $\int_0^t E[N^\varepsilon_t|\mathscr{F}^\eta_{s'/\varepsilon}]\alpha^\varepsilon_{s'}\, d\widetilde{W}_{s'}$ converges to zero in probability as $\varepsilon \to 0$. The last, due to Problem 1.9.2 in [15], holds provided that

$$P-\lim_{\varepsilon\to 0}\int_0^t \left\{E\Big(E\big[N^\varepsilon_t|\mathscr{F}^\eta_{s'/\varepsilon}\big]\alpha^\varepsilon_{s'}|\mathscr{F}^\varepsilon_{s'}\Big)\right\}^2 ds' = 0. \tag{5.4}$$

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

$$P-\lim_{\varepsilon\to 0}\int_0^t \left\{E\Big(E\big[N^\varepsilon_t|\mathscr{F}^\eta_{s'/\varepsilon}\big]\big|\mathscr{F}^\varepsilon_{s'}\Big)\right\}^2 ds' = 0$$

or equivalently by virtue of the Markov property:

$$\begin{aligned} E\Big(E\big[N^\varepsilon_t|\mathscr{F}^\eta_{s'/\varepsilon}\big]\big|\mathscr{F}^\varepsilon_{s'}\Big) &= E\big[N^\varepsilon_t|\mathscr{F}^\eta_{s'/\varepsilon}\big] \\ &= E\big[N^\varepsilon_t|\eta_{s'/\varepsilon}\big] \\ &= E\big[f(\eta_{t/\varepsilon})|\eta_{s'/\varepsilon}\big], \end{aligned}$$

to check only

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

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

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

and so

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

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 η_t be a diffusion Markov process defined by Itô's equation (with respect to the Wiener process W_t^η):

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

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 l and L such that

$$0 < l \le b^2(z) \le L, \qquad \liminf_{|z|\to\infty} a(z)\operatorname{sign} z < 0. \tag{5.6}$$

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

$$\frac{\partial p(y,t,z)}{\partial t} = \mathscr{L}^\star p(y,t,z), \tag{5.7}$$

where

$$\mathscr{L}^\star\{\cdot\} = -\frac{\partial}{\partial z}(a(z)\{\cdot\}) + \frac{1}{2}\frac{\partial^2}{\partial z^2}(b^2(z)\{\cdot\})$$

(for 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 $\mathscr{L}^\star m(z) = 0$. It is well known (see Chapter 4, Section 9, Lemma 9.5; Chapter 3, Section 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 $\lim_{t\to\infty} p(y,t,z) = m(z)$. Then, taking into account that $\int_{\mathsf{R}} p(y,t,z)\,dz = 1$, $\int_{\mathsf{R}} m(z)dz = 1$, by virtue of Scheffe's theorem [1],

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

which is nothing but (1.18).

EXAMPLE 2. Let the homogeneous Markov process η_t take values in the finite space $\{\alpha_1,\dots,\alpha_N\}$ with $\eta_0 = a_j$. Denote by $\mathscr{P}$ 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 α_j to α_i over time t. The vector $p_j(t)$ is defined by the Fokker–Planck–Kolmogorov equation

$$\frac{dp_j(t)}{dt} = p_j(t)\mathscr{P}, \tag{5.8}$$

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 $\mathscr{P}$. Then there exists a unique invariant distribution, $p = (p_1,\dots,p_N)$

satisfying $p\mathscr{P} = 0$. Due to the uniqueness of the invariant distribution we have $\lim_{t\to\infty} p_j(t) = p$, $j = 1, \dots, N$, 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 a 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: $\lim_{t\to\infty} \frac{1}{t}\int_0^t \sin(s)\, ds = 0$ and so (1.15) holds with

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

which implies $\overline{\pi}_t(Y^\varepsilon) \equiv 0$. Hence $E(X_t^\varepsilon - \overline{\pi}_t(Y^\varepsilon))^2 \equiv t$. On the other hand, $\pi_t^\varepsilon(Y^\varepsilon)$, 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], $\lim_{\varepsilon\to 0} P^\varepsilon(t) = P^\circ(t)$, 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 $c^2 = \lim_{\varepsilon\to 0}\int_0^1 \sin^2(s/\varepsilon)\, ds = \frac{1}{2}$. Evidently, for any $t > 0$, $P^\circ(t) < t$; that is, the asymptotic optimality for the filtering estimate $\overline{\pi}_t(Y^\varepsilon)$ is lost.

Acknowledgment. The authors gratefully acknowledge the careful review by the anonymous referees, who also pointed out a mistake in the original version. Because of their comments, the paper was significantly improved.

REFERENCES

[1] BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

[2] BOGOLUBOV, N. and MITROPOLSKII, A. (1961). *Asymptotic Methods in Nonlinear Mechanics.* Gordon and Breach, New York.

[3] FREIDLIN, M. I. and WENTZELL, A. D. (1984). *Random Perturbations of Dynamical Systems.* Springer, New York.

[4] GOGGIN, E. (1992). Convergence of filters with application to the Kalman–Bucy case. *IEEE Trans. Inform. Theory* 38 1091–1100.

[5] GOGGIN, E. (1994). Convergence in distribution of conditional expectations. *Ann. Probab.* 22 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 Probab. Appl.* 3 1–21.

[8] 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.* Birkhäuser, Boston.

[10] KUSHNER, H. J. and RUNGGALDIER, W. J. (1987). Filtering and control for wide bandwidth noise driven systems. *IEEE Trans. Automat. Control* 32 123–133.

[11] JACOD, J. and SHIRYAEV, A. N. (1987). *Limit Theorems for Stochastic Processes.* Springer, 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 Process. Appl.* 38 205–238.

[14] LIPTSER, R. SH. and SHIRYAEV, A. N. (1977, 1978). *Statistics of Random Processes* 1, 2. Springer, 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 *Stochastic Analysis* 449–458.

[17] ZAKAI, M. (1969). On the optimal filtering of diffusion processes. *Z. Wahrsch. Verw. Gebiete* 11 205–238.

M. L. KLEPTSINA
MOSCOW STATE UNIVERSITY
OF RAILWAY ENGINEERING
MOSCOW
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