The Annals of Probability
1984, Vol. 12, No. 3, 890–902

ASYMPTOTIC BAYESIAN ESTIMATION OF A FIRST ORDER EQUATION WITH SMALL DIFFUSION

BY OMAR HIJAB

The Ohio State University

In this paper a finite-dimensional diffusion is observed in the presence of an additive Brownian motion. A large deviations result is obtained for the conditional probability distribution of the diffusion given the observations as the noise variances go to zero.

0. Introduction. Consider a diffusion process $t \to x^\varepsilon(t)$ evolving in $\mathbb{R}^n$ and governed by a generator of the form

$$A^\varepsilon = f + (\varepsilon/2)\ (g_1^2 + \cdots + g_m^2) \tag{0.1}$$

corresponding to a given set of vector fields $f, g_1, \cdots, g_m$ on $\mathbb{R}^n$. It is of interest to study the asymptotic behaviour of the probability distributions P_x^ε on $\Omega^n \equiv C([0, T]; \mathbb{R}^n)$ of the diffusions $t \to x^\varepsilon(t)$ as $\varepsilon \downarrow 0$. It turns out that the asymptotic properties of P_x^ε depend strongly on properties of the associated control system

$$\dot{x} = f(x) + g_1(x)u_1 + \cdots + g_m(x)u_m. \tag{0.2}$$

Indeed it turns out that in some sense

$$P_x^\varepsilon(dx(\cdot)) \sim \exp\left(-\frac{1}{2\varepsilon}\int_0^T u(t)^2\,dt\right) dx(\cdot)$$

as $\varepsilon \downarrow 0$. More precisely, suppose that the diffusions $t \to x^\varepsilon(t)$ satisfy $x^\varepsilon(0) = x^0$ almost surely and suppose that for each u in $L^2([0, T]; \mathbb{R}^m)$ there is a well-defined solution x_u of (0.2) in Ω^n in satisfying $x_u(0) = x^0$. Then the asymptotic behaviour of P_x^ε is given by the following estimates: For any open set G in Ω^n and closed set C in Ω^n, one has

$$\begin{aligned} &\liminf_{\varepsilon\downarrow 0} \varepsilon \log P_x^\varepsilon(G) \ge -\inf\left\{\frac{1}{2}\int_0^T u^2\,dt \mid x_u \text{ in } G\right\} \\ &\limsup_{\varepsilon\downarrow 0} \varepsilon \log P_x^\varepsilon(C) \le -\inf\left\{\frac{1}{2}\int_0^T u^2\,dt \mid x_u \text{ in } C\right\}. \end{aligned} \tag{0.3}$$

In 1966 Varadhan set down a general framework [1] for dealing with the asymptotic behaviour of families of measures and certain associated expectations, and in particular derived analogous estimates for processes with independent increments [1]. Subsequently he derived these estimates for the case of drift-free nondegenerate diffusions (i.e. $f = 0$ and $g_1(x), \cdots, g_m(x)$ span $\mathbb{R}^n$ for all x) [2].

Received May 1983.
AMS 1980 *subject classifications.* 58G32, 60F10, 93E10.
Key words and phrases. Large deviations, Bayesian estimation, diffusion.

Later Glass [3] and Wentzell and Freidlin [4] established these estimates for nondegenerate diffusions with drift.

In 1978 Azencott [5] established these estimates in general; his results imply that if $f, g_1, \cdots, g_m$ are C^2 and the above stated conditions hold, then estimates (0.3) are valid.

Suppose now that the diffusions $t \to x^\varepsilon(t)$ are observed in the presence of an independent Brownian motion $t \to b(t) \in \mathbb{R}^p$,

$$y^\varepsilon(t) = \int_0^t h(x^\varepsilon(s))\,ds + \sqrt{\varepsilon}\, b(t), \quad 0 \le t \le T, \tag{0.4}$$

where $h\colon \mathbb{R}^n \to \mathbb{R}^p$ is a given map. Then thinking of $t \to x^\varepsilon(t)$ and $t \to y^\varepsilon(t)$ as Ω^n-valued and Ω^p-valued ($\Omega^p \equiv C_0([0, T]; \mathbb{R}^p)$) random variables, one can write down the conditional distribution $P^\varepsilon_{x|y}$ of $t \to x^\varepsilon(t)$ given $t \to y^\varepsilon(t)$. Defined abstractly, $P^\varepsilon_{x|y}$ is a distribution on Ω^n measurably parametrized by y in Ω^p, and determined, of course, only up to P^ε_y-null sets in Ω^p, where P^ε_y is the distribution of $t \to y^\varepsilon(t)$ on Ω^p.

The main result of this paper is the establishment of the analogous estimates for $P^\varepsilon_{x|y}$. More precisely, we have the following theorem.

THEOREM A. *Assume that h is C^3 and that $h, f(h), g_i(h), g_i^2(h), i = 1, \cdots, m$, are all bounded on $\mathbb{R}^n$. Then for all $\varepsilon > 0$ there is one and only one version $P^\varepsilon_{x|y}$ of the conditional distribution that depends continuously on y in Ω^p, in the sense that for all Φ in $C_b(\Omega^n)$, the map $y \mapsto E^{P^\varepsilon_{x|y}}(\Phi)$ is in $C_b(\Omega^p)$; this version satisfies*

$$\begin{aligned} &\liminf_{\varepsilon\downarrow 0} \varepsilon \log P^\varepsilon_{x|y}(G) \ge -\inf\{I(u; y) \mid x_u \text{ in } G\} \\ &\limsup_{\varepsilon\downarrow 0} \varepsilon \log P^\varepsilon_{x|y}(C) \le -\inf\{I(u; y) \mid x_u \text{ in } C\} \end{aligned} \tag{0.5}$$

for all G open in Ω^n, C closed in Ω^n, and y in Ω^p, where

$$I(u; y) = J(u; y) - \inf\{J(u; y) \mid u \text{ in } L^2([0, T]; \mathbb{R}^m)\}$$

and $J\colon L^2([0, T]; \mathbb{R}^m) \times \Omega^p \to \mathbb{R}$ is given by

$$J(u; y) = \int_0^T \left[\frac{1}{2} u^2 + \frac{1}{2} h(x_u)^2\right] dt - \int_0^T h(x_u)\, dy.$$

In particular we note that for $h = 0$ this theorem reduces to estimates (0.3). The plan of the paper is as follows. In section one we establish and state precisely estimates (0.3) in the form that we need them. For completeness we include an appendix in which a proof of (0.3) is provided. The proof presented is standard except for the part dealing with the second estimate (0.3). Here instead of appealing to the Markov property of $t \to x^\varepsilon(t)$ to deduce the fact that $t \to x^\varepsilon(t)$ can be approximated sufficiently well by its "discretized version", we instead appeal to the compactness of the map $u \mapsto x_u$ defined above; this leads to a quick proof of the second estimate (0.3).

In section two we prove Theorem A and in section three we present an application. Other applications, analogous to some in [10], can be derived. Some applications appear in a conference proceedings note [9].

1. Large deviations. Throughout Ω^n will denote $C([0, T]; \mathbb{R}^n)$; Ω^m and Ω^p denote the corresponding spaces of paths starting at the origin. The topology of Ω^n is that of uniform convergence on $[0, T]$. We suppose that we are given

(i) (possibly time-varying) vector fields $f, g_1, \cdots, g_m$ in $C^{0,2}([0, T] \times \mathbb{R}^n; \mathbb{R}^n)$.

(Warning: No assumptions are made on the growth of $f, g_1, \cdots, g_m$ at infinity or any of their derivatives.) If g is any vector field and φ is a differentiable function, let $g(\varphi)(x)$ denote the directional derivative of φ in the direction of g at the point x. Thus g is a first-order differential operator taking φ to $g(\varphi)$. If $g(\varphi)$ is differentiable then $g^2(\varphi) = g(g(\varphi))$ makes sense; thus (0.1) defines a second-order (possibly time-varying) differential operator A^ε acting on the space $C_0^\infty(\mathbb{R}^n)$ of all smooth compactly supported functions on $\mathbb{R}^n$.

Let $b(t)$: $\Omega^m \to \mathbb{R}^m$ be given by $b(t, \omega) = \omega(t)$ and impose Wiener measure W on Ω^m. Then $t \to b(t) = (b_1(t), \cdots, b_m(t))$ is an $\mathbb{R}^m$-valued Brownian motion.

One way to construct diffusions on $\mathbb{R}^n$ governed by A^ε is to pick a point x^0 in $\mathbb{R}^n$ and to let $t \to x^\varepsilon(t)$ be the unique process $\Omega^m \to \Omega^n$ satisfying

$$(1.1) \quad \varphi(x^\varepsilon(t)) - \varphi(x^\varepsilon(s)) - \int_s^t A^\varepsilon(\varphi)(x^\varepsilon(r))\, dr = \sqrt{\varepsilon} \int_s^t g(\varphi)(x^\varepsilon(r))\, db(r)$$

for all φ in $C_0^\infty(\mathbb{R}^n)$ and $0 \le s \le t \le T$, $x^\varepsilon(0) = x^0$ almost surely on Ω^m. Here and elsewhere, $g(\varphi)\, db$ is short for $g_1(\varphi)\, db_1 + \cdots + g_m(\varphi)\, db_m$ where $g_i(\varphi)$ is defined above. Similarly, $u^2 = u_1^2 + \cdots + u_m^2$, $yg(h) = y_1 g(h_1) + \cdots + y_p g(h_p)$, etc. Using the standard existence and uniqueness theorem for stochastic differential equations coupled with Ito's differential rule, it is straightforward to verify that there is a unique such process defined up to an explosion time $\zeta^\varepsilon \le \infty$ characterized by the fact that $t \to x^\varepsilon(t)$ leaves every compact subset of $\mathbb{R}^n$ as $t \uparrow \zeta^\varepsilon$, almost surely on $\zeta^\varepsilon < \infty$.

As is well-known [7], the merit of the above definition is that it makes sense on any manifold X. Indeed the Whitney embedding theorem allows one to embed any such X into some $\mathbb{R}^N$ and by extending the given vector fields $f, g_1, \cdots, g_m$ on X to all of $\mathbb{R}^N$ one can derive the result described above on any manifold. What follows is stated in such a way as to lend itself easily to modification from $\mathbb{R}^n$ to the general manifold case.

If $\zeta^\varepsilon \ge T$ almost surely then the probability distribution P_x^ε of $t \to x^\varepsilon(t)$ exists on Ω^n and is the unique solution to the martingale problem corresponding to A^ε and starting at x^0 at time 0. In other words,

$$(1.2) \qquad E^{P_x^\varepsilon}(\varphi(x(t)) - \varphi(x(s)) - \int_s^t A^\varepsilon(\varphi)(x^\varepsilon(r))\, dr \mid \mathscr{F}_s) = 0$$

for all φ in $C_0^\infty(\mathbb{R}^n)$ and $0 \le s \le t \le T$. Here $x(t)$: $\Omega^n \to \mathbb{R}^n$ is the canonical map and $\mathscr{F}_s$ is the σ-algebra on Ω^n generated by the maps $x(r)$, $0 \le r \le s$.

Conversely, if one assumes that

(ii) for each $\varepsilon > 0$ there is a probability measure P_x^ε on Ω^n satisfying (1.2) and $P_x^\varepsilon(x(0) = x^0) = 1$,

then there is a unique such measure and the solution $t \to x^\varepsilon(t)$ of (1.1) explodes after time T and P_x^ε is then the distribution of $t \to x^\varepsilon(t)$ on Ω^n.

In what follows we shall assume (ii) and

(iii) to each u in $L^2([0, T]; \mathbb{R}^m)$ there is a path x_u in Ω^n satisfying (0.2) and $x_u(0) = x^0$. In other words the solution of (0.2) starting at x^0 explodes after time T, for all u in L^2.

Assumptions (ii) and (iii) holds, for instance, if $f, g_1, \cdots, g_m$ grow at most linearly at infinity and the first partial derivatives of $g_1, \cdots, g_m$ are bounded on $\mathbb{R}^n$ [8]. Under assumptions (i), (ii), (iii) estimates (0.3) hold. To understand these estimates from a more general perspective consider the following definition [1]:

DEFINITION. Let Ω be a completely regular space and let P^ε, $\varepsilon > 0$, be a family of probability measures on Ω. We say that $\{P^\varepsilon\}$ *admits large deviation* if there is a function I of Ω satisfying

(LD1) $0 \le I \le \infty$.
(LD2) I is lower semicontinuous on Ω.
(LD3) $\{\omega \mid I(\omega) \le M\}$ is a compact subset of Ω for all $M < \infty$.
(LD4) For any open set G in Ω, $\liminf_{\varepsilon \downarrow 0} \varepsilon \log P^\varepsilon(G) \ge -\inf\{I(\omega) \mid \omega \text{ in } G\}$.
(LD5) For any closed set C in Ω,

$$\limsup_{\varepsilon \downarrow 0} \varepsilon \log P^\varepsilon(C) \le -\inf\{I(\omega) \mid \omega \text{ in } C\}.$$

The function I is then referred to as the corresponding "I-functional".

Estimates (0.3) then state that (LD4) and (LD5) hold for the probability distributions $\{P_x^\varepsilon\}$ of the diffusions $t \to x^\varepsilon(t)$, where I is given by

$$I(\omega) = \inf\left\{\frac{1}{2}\int_0^T u^2\, dt \mid x_u = \omega\right\} \tag{1.3}$$

for all ω in Ω^n, with the understanding that the infimum of an empty set of real numbers is $+\infty$. Actually more is true.

THEOREM B. *The probability distributions* $\{P_x^\varepsilon\}$ *corresponding to* A^ε *admit large deviation with I-functional given by* (1.3).

A proof is provided in the appendix.

A consequence of the above definition is the following proposition which is a summary of results appearing in Section 3 of [1].

PROPOSITION C. *Let* $\{P^\varepsilon\}$ *admit large deviation with corresponding I-functional I and let* Φ_ε *be a bounded continuous function on* Ω *such that* Φ_ε *converges* *uniformly to* Φ *as* $\varepsilon \downarrow 0$. *Let* Q^ε *be given by*

$$dQ^\varepsilon = \exp(-(1/\varepsilon)\Phi_\varepsilon)\ dP^\varepsilon.$$

Then $\{Q^\varepsilon\}$ *satisfies*

$$\liminf_{\varepsilon\downarrow 0} \varepsilon \log Q^\varepsilon(G) \geq -\inf\{\Phi(\omega) + I(\omega) \mid \omega \text{ in } G\},$$

$$\limsup_{\varepsilon\downarrow 0} \varepsilon \log Q^\varepsilon(C) \leq -\inf\{\Phi(\omega) + I(\omega) \mid \omega \text{ in } C\},$$

for G *open and* C *closed in* Ω.

We note that for the results of this proposition to hold, all that is required is that the tail estimate

$$\lim_{R\to\infty} \limsup_{\varepsilon\downarrow 0} \varepsilon \log E^{P^\varepsilon}(1_{[-\Phi_\varepsilon \geq R]} \times \exp(-\Phi_\varepsilon/\varepsilon)) = -\infty$$

holds [1].

2. Nonlinear filtering. Let $h\colon \mathbb{R}^n \to \mathbb{R}^p$ be a locally bounded measurable map and let $t \to b(t)$ denote an $\mathbb{R}^p$-valued Brownian motion independent of given processes $t \to x^\varepsilon(t)$ on $\mathbb{R}^n$. Let $t \to y^\varepsilon(t)$ be given by (0.4). In this section we study the conditional distribution $P^\varepsilon_{x|y}$ on Ω^n of $t \to x^\varepsilon(t)$ given $t \to y^\varepsilon(t)$. We use Bayes' rule to compute $P^\varepsilon_{x|y}$.

Let W^ε denote Wiener measure of variance ε on Ω^p, let P^ε_x denote the distribution of $t \to x^\varepsilon(t)$ on Ω^n, let P^ε_y denote the distribution of $t \to y^\varepsilon(t)$ on Ω^p, let $P^\varepsilon_{(x,y)}$ denote the distribution of $t \to (x^\varepsilon(t), y^\varepsilon(t))$ on $\Omega^n \times \Omega^p$, and let $P^\varepsilon_{x|y}$ denote the conditional distribution of $t \to x^\varepsilon(t)$ given $t \to y^\varepsilon(t)$. Let $y(t)\colon \Omega^p \to \mathbb{R}^p$ denote the canonical map.

For $0 \leq t \leq T$ set

$$\Lambda(t) = \frac{1}{2}\int_0^t h(x(s))^2\, ds - \int_0^t h(x(s))\, dy(s).$$

$\Lambda(t)$ is then a measurable function on $\Omega^n \times \Omega^p$ for each t. Using (0.4) and invoking Cameron-Martin it is easy to see that

$$dP^\varepsilon_{(x,y)} = e^{-(1/\varepsilon)\Lambda} d(P^\varepsilon_x \times W^\varepsilon)$$

where $\Lambda = \Lambda(T)$. Here and elsewhere, $h^2 = h_1^2 + \cdots + h_p^2$, $hdy = h_1dy_1 + \cdots + h_pdy_p$, etc.

Using Bayes' rule, the conditional distribution is given by

$$dP^\varepsilon_{x|y} = \frac{dP^\varepsilon_{(x,y)}}{d(P^\varepsilon_x \times P^\varepsilon_y)} \cdot dP^\varepsilon_x = \frac{dP^\varepsilon_{(x,y)}}{d(P^\varepsilon_x \times W^\varepsilon)} \cdot \frac{d(P^\varepsilon_x \times W^\varepsilon)}{d(P^\varepsilon_x \times P^\varepsilon_y)} \cdot dP^\varepsilon_x \tag{2.1}$$

$$= \exp(-\Lambda/\varepsilon) \cdot (E^{P^\varepsilon_x}(\exp(-\Lambda/\varepsilon)))^{-1} \cdot dP^\varepsilon_x.$$

If we set

$$dQ^\varepsilon_{x|y} \equiv \exp(-\Lambda/\varepsilon)\ dP^\varepsilon_x \tag{2.2}$$

then (2.1) becomes

$$dP^{\varepsilon}_{x|y} = dQ^{\varepsilon}_{x|y}/Q^{\varepsilon}_{x|y}(\Omega^n), \quad \text{a.s. } W^{\varepsilon}, \tag{2.1}$$

which is the Kallianpur-Striebel formula [6]. We refer to $Q^{\varepsilon}_{x|y}$ as the unnormalized conditional distribution. So far (2.1) holds for *any* processes $t \to x^{\varepsilon}(t)$.

We suppose that

(iv) h is C^3 and $f(h)$, $g_i(h)$, $g_i^2(h)$, $i = 1, \cdots, m$, and h are all bounded on $\mathbb{R}^n$.

Then for each $\varepsilon > 0$ let Φ_ε on Ω^n be given by

$$\Phi_\varepsilon(\omega; y) = -y(T)h(\omega(T)) + \int_0^T \left[yA^{\varepsilon}(h)(\omega) + \frac{1}{2} h(\omega)^2 - \frac{1}{2} y^2 g(h)(\omega)^2 \right] dt.$$

Then $\Phi_\varepsilon \to \Phi_0 = \Phi$ as $\varepsilon \downarrow 0$ uniformly on Ω^n, for each y in Ω^p. Referring to (2.2) and performing an integration by parts in the stochastic integral appearing in Λ and invoking Girsanov's theorem we see that

$$dP^{\varepsilon}_{x:y} \equiv \exp((1/\varepsilon)\Phi_\varepsilon)\, dQ^{\varepsilon}_{x|y} \quad \text{a.s. } W^{\varepsilon} \tag{2.3}$$

is governed by

$$A^{\varepsilon,y} = A^{\varepsilon} - yg_1(h)g_1 - \cdots - yg_m(h)g_m$$

in the sense that

$$E^{P^{\varepsilon}_{x:y}}(\varphi(x(t)) - \varphi(x(s)) - \int_s^t A^{\varepsilon,y}(\varphi)(x(r))\, dr \mid \mathscr{F}_s) = 0$$

for all $0 \le s \le t \le T$ and φ in $C_0^\infty(\mathbb{R}^n)$, almost surely on Ω^p.

We now wish to apply theorem B to $\{P^{\varepsilon}_{x:y}\}$. To do so we must check that assumptions (i), (ii), (iii) of section one hold for

$$f_y = f - yg_1(h)g_1 - \cdots - yg_m(h)g_m, \quad g_1, \cdots, g_m$$

for all y in Ω^p (given that they hold for $y = 0$). But for (i), (ii) this is immediate and for (iii) this is so because $g_1(h), \cdots, g_m(h)$ are bounded feedback terms. Thus theorem B applies to $\{P^{\varepsilon}_{x:y}\}$ and hence proposition C applies to $Q^{\varepsilon}_{x|y}$ via equation (2.3).

Thus let $x_{u:y}$ denote the unique path in Ω^n satisfying

$$\dot{x} = f_y(x) + g_1(x)u_1 + \cdots + g_m(x)u_m \quad \text{and} \quad x(0) = x^0.$$

Then according to proposition C,

$$\liminf_{\varepsilon\downarrow 0} \varepsilon \log Q^{\varepsilon}_{x|y}(G) \ge -\inf\left\{\frac{1}{2}\int_0^T u^2\, dt + \Phi(x_{u:y}, y) \mid x_{u:y} \in G\right\}.$$

Noting that $x_{u:y} = x_v$ where $v = u - yg(h)(x_v)$, we have

$$\frac{1}{2}\int_0^T u^2\, dt + \Phi(x_{u:y}, y) = \frac{1}{2}\int_0^T [v^2 + h(x_v)^2]\, dt - \int_0^T h(x_v)\, dy$$

and so the first of the inequalities (2.4) below follows. The second inequality follows similarly.

Now since $P^\varepsilon_{x:y}$ is governed by $A^{\varepsilon,y}$, it is straightforward to check that $y \to P^\varepsilon_{x:y}$ is continuous in the sense that $y \to E^{P^\varepsilon_{x:y}}(\Phi)$ is in $C_b(\Omega^p)$ for all Φ in $C_b(\Omega^n)$. Since $y \to \Phi_\varepsilon(\omega; y)$ is continuous in y uniformly in ω, it then follows that $Q^\varepsilon_{x|y}$, if defined by (2.3), is continuous in y. Thus $Q^\varepsilon_{x|y}$ has a version (that given by (2.3)), that is continuous in y. Since $P^\varepsilon_{x|y}$ satisfies (2.1) a.s. W^ε, we see that $P^\varepsilon_{x|y}$ also has a version that is continuous in y. Moreover such a version is uniquely determined since any two must agree a.s. W^ε and hence, being continuous, must agree on the support of W^ε which is all of Ω^p. We have shown the following:

THEOREM D. *There is one and only one continuous version $Q^\varepsilon_{x|y}$ of the unnormalized conditional distribution. This version satisfies, for all G open in Ω^n and C closed in Ω^n,*

$$\begin{aligned} &\liminf_{\varepsilon\downarrow 0} \varepsilon \log Q^\varepsilon_{x|y}(G) \geq -\inf\{J(u, y) \mid x_u \text{ in } G\} \\ &\limsup_{\varepsilon\downarrow 0} \varepsilon \log Q^\varepsilon_{x|y}(C) \leq -\inf\{J(u, y) \mid x_u \text{ in } C\} \end{aligned} \tag{2.4}$$

where $J(u; y)$ is as given in Theorem A.

As a corollary, if we choose $\Omega^n = G = C$, we have

$$\lim_{\varepsilon\downarrow 0} \varepsilon \log Q^\varepsilon_{x|y}(\Omega^n) = -\inf\{J(u, y) \mid u \text{ in } L^2\}. \tag{2.5}$$

This last fact, coupled with (2.1) and Theorem D yields Theorem A. In fact one can show that $\{P^\varepsilon_{x|y}\}$ admits large deviations.

3. An application. For φ in $B(\mathbb{R}^n)$ let $\mu^\varepsilon_t(\varphi)$, $\hat{\varphi}_\varepsilon(t)$: $\Omega^p \to \mathbb{R}$ be given by

$$\mu^\varepsilon_t(\varphi) = E^{P^\varepsilon_x}(\exp(-\Lambda(t)/\varepsilon)\varphi(x(t))) \tag{3.1}$$

and

$$\hat{\varphi}_\varepsilon(t) = \mu^\varepsilon_t(\varphi)/\mu^\varepsilon_t(1).$$

The random variable $\hat{\varphi}_\varepsilon(t)(\mu^\varepsilon_t(\varphi))$ is the (unnormalized) conditional expectation of $t \to \varphi(x^\varepsilon(t))$ given $y^\varepsilon(s)$, $0 \leq s \leq t$, in the sense that

$$\hat{\varphi}_\varepsilon(t)\circ y^\varepsilon = E(\varphi(x^\varepsilon(t)) \mid \mathscr{Y}^\varepsilon_t)$$

almost surely on $\Omega^m \times \Omega^p$, where $\mathscr{Y}^\varepsilon_t = \sigma[y^\varepsilon(s), 0 \leq s \leq t]$, $t \leq T$, and y^ε is given by (0.4). A measure of accuracy of the corresponding "filter" is then

$$\frac{1}{2}\int_0^T \hat{h}_\varepsilon(t)^2\, dt - \int_0^T \hat{h}_\varepsilon(t)\, dy(t), \tag{3.2}$$

representing the L^2-norm of the difference between $y(t)$, $0 \leq t \leq T$, and $\hat{h}_\varepsilon(t)$, $0 \leq t \leq T$. We note that the second integral in (3.2) is an Itô stochastic integral. In this section we prove the following theorem, under our assumptions (i), (ii), (iii), (iv).

THEOREM E. *For all $\varepsilon > 0$, there is one and only one continuous version of the* *residual* (3.2). *As* $\varepsilon \downarrow 0$, *these versions converge to*

$$\inf\left\{\frac{1}{2}\int_0^T [u^2 + h(x_u)^2]\,dt - \int_0^T h(x_u)\,dy \mid u \text{ in } L^2([0, T]; \mathbb{R}^m)\right\},$$

for all y *in* Ω^p.

The proof of Theorem E is based on the well-known observation that for each $\varepsilon > 0$, (3.2) is $W^\varepsilon -$ a.s. equal to $-\varepsilon \log Q^\varepsilon_{x|y}(\Omega^n)$. Given this observation, Theorem E then follows from (2.5) and Theorem D.

To derive this observation, we note the following well-known result concerning $\mu^\varepsilon_t(\varphi)$.

LEMMA. *Let* A^ε *be given by* (0.1) *and let* P^ε_x *be a solution to the martingale problem corresponding to* A^ε. *Let* $h\colon \mathbb{R}^n \to \mathbb{R}^p$ *be a bounded measurable map and let* $\mu^\varepsilon_t(\varphi)$ *be given by* (3.1), *where* $\Lambda(t)$ *is as in Section* 2. *Then for all* φ *in* $C_0^\infty(\mathbb{R}^n)$, $0 \le s \le t \le T$, *we have*

$$\mu^\varepsilon_t(\varphi) - \mu^\varepsilon_s(\varphi) - \int_s^t \mu^\varepsilon_r(A^\varepsilon\varphi)\,dr = \frac{1}{\varepsilon}\int_s^t \mu^\varepsilon_r(h\varphi)\,dy(r)$$

$W^\varepsilon -$ *a.s. on* Ω^p.

PROOF. Follows from (1.2), the fact that $z^\varepsilon(t) \equiv \exp(-\Lambda(t)/\varepsilon)$ satisfies

$$z^\varepsilon(t) - z^\varepsilon(s) = \frac{1}{\varepsilon}\int_s^t h(x(r))z^\varepsilon(r)\,dy(r)$$

for $0 \le s \le t \le T$, and a Fubini-type argument similar to that on page 87 of [8].

Continuing now with the proof of Theorem E, note that by the result of this lemma, for h bounded $\mu^\varepsilon_t(1)$ satisfies

$$\mu^\varepsilon_t(1) - \mu^\varepsilon_s(1) = \frac{1}{\varepsilon}\int_s^t \hat{h}_\varepsilon(r)\mu^\varepsilon_r(1)\,dy(r)$$

for $0 \le s \le t \le T$, and $\mu^\varepsilon_0(1) = 1$, for all $\varepsilon > 0$. Solving for the unique solution of this stochastic integral equation, we see that

$$\mu^\varepsilon_t(1) = \exp\frac{1}{\varepsilon}\left(\int_0^t \hat{h}_\varepsilon(s)\,dy(s) - \frac{1}{2}\int_0^t \hat{h}_\varepsilon(s)^2\,ds\right) \quad \text{a.s.} - W^\varepsilon$$

for $0 \le t \le T$. Since $\mu^\varepsilon_T(1) = Q^\varepsilon_{x|y}(\Omega^n)$, we conclude that $-\varepsilon \log Q^\varepsilon_{x|y}(\Omega^n)$ and (3.2) are a.s. equal. Applying Theorem D now yields the result.

4. Appendix. The purpose of this appendix is to give a self-contained proof of the fact that $\{P^\varepsilon_x\}$ admits large deviations with I-functional

$$I(\omega) = \inf\left\{\frac{1}{2}\int_0^T u^2\,dt \mid x_u = \omega\right\}. \tag{A.1}$$

Before we begin we give some elementary examples of large deviations which will be of use later on.

If ξ is an $N(0, 1)$ random variable in $\mathbb{R}^m$, and P^ε is the distribution of $\sqrt{\varepsilon}\,\xi$, then $\{P^\varepsilon\}$ admits large deviation with I-functional $I_1(u) = u^2/2$, u in $\mathbb{R}^m$.

If $\xi_1, \xi_2, \cdots, \xi_N$ is an i.i.d. sequence of $N(0, 1)$ random variables in $\mathbb{R}^m$ and P^ε is the distribution of $\sqrt{\varepsilon}(\xi_1, \cdots, \xi_N)$ on $\mathbb{R}^{mN}$, then here $I_2(u) = (u_1^2 + \cdots + u_N^2)/2$, u_k in $\mathbb{R}^m$.

Suppose now that a given family $\{P^\varepsilon\}$ admits large deviations on Ω with I-functional I_2 and $\alpha\colon \Omega \to \Omega'$ is continuous. Then it is elementary to verify that $\{P^\varepsilon \circ \alpha^{-1}\}$ admits large deviations on Ω' with I-functional

$$I_3(\omega') = \inf\{I_2(\omega) \mid \alpha(\omega) = \omega'\}.$$

For example, consider the process

$$\text{(A.2)} \qquad u^N(t) = \sqrt{N/T}\,\xi_k \qquad (k-1)T/N \le t < kT/N, \quad k = 1, \cdots, N.$$

Then $t \to u^N(t)$ is in $L^2([0, T]; \mathbb{R}^m)$ (henceforth denoted by L^2); if P^ε is the distribution of $\sqrt{\varepsilon}\,u^N$ on L^2 then we have

$$\begin{aligned} I_3(u) &= \frac{1}{2}\int_0^T u^2\,dt, \quad \text{if} \quad u(t) = u([Nt/T]/N), \quad 0 \le t \le T, \\ &= +\infty \qquad\qquad \text{otherwise.} \end{aligned}$$

Now we make a specific choice of $\xi_1, \cdots, \xi_N$ by setting

$$\text{(A.3)} \qquad \xi_k = (b(kT/N) - b((k-1)T/N))\sqrt{N/T},$$

$k = 1, \cdots, N$, where $t \to b(t)$ is as defined in Section 1. We now set $x^{\varepsilon,N} \equiv x_v$ where $v = \sqrt{\varepsilon}\,u^N$. Since $u \mapsto x_u$ is certainly continuous when u is piecewise constant we have that the distributions of $t \to x^{\varepsilon,N}(t)$ on Ω^n admit large deviations with

$$I_4(\omega) = \inf\left\{\frac{1}{2}\int_0^T u^2\,dt \,\middle|\, x_u = \omega \quad \text{and} \quad u(t) = u([Nt/T]/N), \quad 0 \le t \le T\right\}.$$

Thus for C closed in Ω^n,

$$\limsup_{\varepsilon\downarrow 0} \varepsilon \log W(x^{\varepsilon,N} \in C) \le -\inf\{I_4(\omega) \mid \omega \text{ in } C\}.$$

Since $I_4(\omega) \ge I(\omega)$ as given by (A.1), we arrive at

$$\text{(A.4)} \qquad \limsup_{\varepsilon\downarrow 0} \varepsilon \log W(x^{\varepsilon,N} \in C) \le -\inf\{I(\omega) \mid \omega \text{ in } C\},$$

for all $N \ge 1$.

LEMMA 1. *$u_i \to u$ weakly in L^2 implies that $x_{u_i} \to x_u$ in Ω^n.*

PROOF. Let B be a ball in $\mathbb{R}^n$ containing $x_u(t)$, $0 \le t \le T$, and let τ_i be the first exit time of x_{u_i} from B. Let K be a bound for $f, g_1, \cdots, g_m$ and their first derivatives on B and the L^2-norms of $u_1, u_2, \cdots$. Since both x_{u_i} and x_u are solutions of (0.2) starting at x_0, for $t < \tau_i \wedge T$ we have

$$e_i(t) \le K \int_0^t e_i(s)(1 + (u_i(s)))\, ds + r_i(t)$$

where

$$e_i(t) = \max_{0\le s\le t} | x_{u_i}(s) - x_u(s) |$$

and

$$r_i(t) = \max_{0\le s\le t} \left| \int_0^s g(x_u(r))(u_i(r) - u(r))\, dr \right|.$$

Now an application of Gromwall's inequality yields

$$e_i(\tau_i \wedge T) \le C(K, T) r_i(T).$$

But Ascoli's theorem guarantees that $r_i(T) \to 0$ is $i \uparrow \infty$ and thus for i sufficiently large $\tau_i > T$ and hence $x_{u_i} \to x_u$ in Ω^n.

LEMMA 2. *For all $\delta > 0$ and $K > 0$ there exists a weak neighborhood U of 0 in L^2 such that $\| u \|_2 \le K$ and $\| x_u - x_v \|_\infty \ge \delta$ imply $u - v$ is in U^c.*

PROOF. This follows from Lemma 1.

Lemma 1 also implies that I is lower semicontinuous: if $\omega_i \to \omega$ in Ω^n then

$$I(\omega) \le \liminf_{i\to\infty} I(\omega_i).$$

To show this, it is enough to assume the right-hand side is finite; let ω_i' be a subsequence such that $\lim I(\omega_i') = \liminf I(\omega_i)$. Choose u_i in L^2 such that $½ \int_0^T u_i^2\, dt \le I(\omega_i') + i^{-1}$ and $x_{u_i} = \omega_i'$. Since then $\{u_i\}$ is bounded in L^2, by passing to a subsequence we can assume that u_i converges weakly to some u in L^2, and hence $x_{u_i} \to x_u$ in Ω^n. Thus

$$I(\omega) = I(x_u) \le \frac{1}{2}\int_0^T u^2\, dt \le \liminf \frac{1}{2}\int_0^T u_i^2\, dt \le \lim I(\omega_i'),$$

which implies the result.

The above results together imply that $\{\omega \mid I(\omega) \le M\}$ is compact and so (LD2) and (LD3) are established.

We now prove (LD4). For u in L^2 set

$$A^{\varepsilon,u} = A^\varepsilon + g_1 u_1 + \cdots + g_m u_m;$$

let $b^{\varepsilon,u}\colon \Omega^m \to \Omega^m$ be given by

$$b^{\varepsilon,u}(t) = b(t) + \frac{1}{\sqrt{\varepsilon}} \int_0^t u(s)\, ds.$$

Let $W^{\varepsilon,u}$ denote the image of Wiener measure W on Ω^n under the map $b^{\varepsilon,u}$. The solution of (1.1) provides a map $x^\varepsilon\colon \Omega^m \to \Omega^n$ determined almost surely W. Since by Cameron-Martin

$$\frac{dW}{dW^{\varepsilon,u}} = \exp\left(-\frac{1}{\sqrt{\varepsilon}}\int_0^T u\,db + \frac{1}{2\varepsilon}\int_0^T u^2\,dt\right),$$

we see that W and $W^{\varepsilon,u}$ are equivalent and so the image of $W^{\varepsilon,u}$ under the map x^ε is a well-defined distribution $P_x^{\varepsilon,u}$ on Ω^n. It is easily seen then that $P_x^{\varepsilon,u}$ is governed by $A^{\varepsilon,u}$ in the sense that for all φ in $C_0^\infty(\mathbb{R}^n)$ and $0 \le s \le t \le T$,

$$E^{P_x^{\varepsilon,u}}\left(\varphi(x(t)) - \varphi(x(s)) - \int_s^t A^{\varepsilon,u}(\varphi)(x(r))\,dr \mid \mathscr{F}_s\right) = 0.$$

Now for $\varepsilon = 0$, $P_x^{\varepsilon,u}$ is simply the Dirac mass δ_{x_u} supported at x_u. Therefore as $\varepsilon \downarrow 0$ $P_x^{\varepsilon,u}$ converges to δ_{x_u} and so

$$\lim_{\varepsilon\downarrow 0} P_x^{\varepsilon,u}(G^c) = 0$$

for any open set G in Ω^n containing x_u.

Now let $0 < \theta < 1$ and let G be an open set in Ω^n with x_u in G. Then

$$\begin{aligned}
P_x^\varepsilon(G) &= W(x^\varepsilon \in G)\\
&= E^{W^{\varepsilon,u}}\left(\exp\left(-\frac{1}{\sqrt{\varepsilon}}\int_0^T u\,db + \frac{1}{2\varepsilon}\int_0^T u^2\,dt\right), x^\varepsilon \in G\right)\\
&= E^{W}\left(\exp\left(-\frac{1}{\sqrt{\varepsilon}}\int_0^T u\,db^{\varepsilon,u} + \frac{1}{2\varepsilon}\int_0^T u^2\,dt\right), x^\varepsilon\circ b^{\varepsilon,u} \in G\right)\\
&= E^{W}\left(\exp\left(-\frac{1}{\sqrt{\varepsilon}}\int_0^T u\,db - \frac{1}{2\varepsilon}\int_0^T u^2\,dt\right), x^\varepsilon\circ b^{\varepsilon,u} \in G\right)\\
&\ge W\left(x^\varepsilon\circ b^{\varepsilon,u} \in G,\ \theta\int_0^T u\,db \le \left(\left(\int_0^T u^2\,dt\right)^{1/2}\right)\right)\\
&\quad \times \exp\left(-\frac{1}{\theta\sqrt{\varepsilon}}\left(\int_0^T u^2\,dt\right)^{1/2} - \frac{1}{2\varepsilon}\int_0^T u^2\,dt\right)\\
&\ge (1 - \theta^2 - P_x^{\varepsilon,u}(G^c)) \times \exp\left(-\frac{1}{\theta\sqrt{\varepsilon}}\left(\int_0^T u^2\,dt\right)^{1/2} - \frac{1}{2\varepsilon}\int_0^T u^2\,dt\right).
\end{aligned}$$

Now taking the logarithm of both sides and the limit as $\varepsilon \downarrow 0$, we have

$$\liminf_{\varepsilon\downarrow 0} \varepsilon \log P_x^\varepsilon(G) \ge -\frac{1}{2}\int_0^T u^2\,dt.$$

Taking the supremum over all x_u in G, we arrive at (LD4).

To establish (LD5), we need a simple device first. Let $N \ge 1$ and partition

$[0, T]$ into N equal pieces. For any v in L^2, let v^N be given by

$$v^N(t) = \frac{N}{T} \int_{(k-1)T/N}^{kT/N} v(s)\, ds \quad \text{for} \quad (k-1)T/N \le t < kT/N.$$

Then v^N is also in L^2 and $v^N \to v$ in L^2 as $N \uparrow \infty$.

Let b^N denote the process equal to b at the time points kT/N and linearly interpolated in between. Let $u^N = db^N/dt$; then u^N is in L^2 and is given by (A.2) where ξ_k is given by (A.3). Note that for any v in L^2

$$\int_0^T v u^N\, dt = \int_0^T v^N\, db.$$

Recall that $x^{\varepsilon,N} = x_{\sqrt{\varepsilon} u^N}$ and that $u \mapsto x_u$ is continuous in the weak L^2 topology. Then according to Lemma 2, for $\delta > 0$ and $K > 0$, there are $v_1, \cdots, v_p$ in L^2 and $a > 0$ such that the set $G_{K,\delta} = \{(u, v) \mid \|u\|_2 < K \text{ and } \|x_u - x_v\|_\infty > \delta\}$ is contained in the set of (u, v) satisfying $|\langle v_i, u - v\rangle| \ge a$ from some $1 \le i \le p$ ($\langle , \rangle$ is the L^2-inner product). Hence for all $N, M \ge 1$

$$\begin{aligned} &W(\|x^{\varepsilon,N} - x^{\varepsilon,M}\|_\infty > \delta \text{ and } \sqrt{\varepsilon}\, \|u^N\|_2 < K) \\ &\quad = W(\sqrt{\varepsilon}(u^N, u^M) \text{ in } G_{K,\delta}) \\ &\quad \le \max\nolimits_{1\le i\le p} W\left(\sqrt{\varepsilon}\left|\int_0^T v_i(u^N - u^M)\, dt\right| \ge a\right) \\ &\quad = \max\nolimits_{1\le i\le p} W\left(\sqrt{\varepsilon}\left|\int_0^T (v_i^N - v_i^M)\, db\right| \ge a\right) \\ &\quad \le \max\nolimits_{1\le i\le p} \frac{2\sqrt{\varepsilon}\, \|v_i^N - v_i^M\|_2}{a\sqrt{2\pi}} \cdot \exp\left(-\frac{a^2}{2\varepsilon\, \|v_i^N - v_i^M\|_2^2}\right). \end{aligned} \tag{A.5}$$

Now as $M \uparrow \infty$, $t \to (x^{\varepsilon,N}(t), x^{\varepsilon,M}(t))$ converges in distribution to $t \to (x^{\varepsilon,N}(t), x^\varepsilon(t))$ and so

$$\begin{aligned} &W(\|x^{\varepsilon,N} - x^\varepsilon\|_\infty > \delta) \\ &\qquad \le \liminf\nolimits_{M\uparrow\infty} W(\sqrt{\varepsilon}(u^N, u^M) \in G_{K,\delta}) + W(\sqrt{\varepsilon}\, \|u^N\|_2 \ge K). \end{aligned}$$

But according to (A.5),

$$\lim\nolimits_{N\uparrow\infty} \limsup\nolimits_{\varepsilon\downarrow 0} \liminf\nolimits_{M\uparrow\infty} \varepsilon \log W(\sqrt{\varepsilon}(u^N, u^M) \in G_{K,\delta}) = -\infty.$$

Therefore

$$\begin{aligned} &\lim\nolimits_{N\uparrow\infty} \limsup\nolimits_{\varepsilon\downarrow 0} \varepsilon \log W(\|x^{\varepsilon,N} - x^\varepsilon\|_\infty > \delta) \\ &\qquad \le \lim\nolimits_{N\uparrow\infty} \limsup\nolimits_{\varepsilon\downarrow 0} \varepsilon \log W(\sqrt{\varepsilon}\, \|u^N\|_2 \ge K) \\ &\qquad \le \lim\nolimits_{N\uparrow\infty} -\tfrac{1}{2} K^2 = -\tfrac{1}{2} K^2 \end{aligned} \tag{A.6}$$

for all K and all $\delta > 0$.

Now if C is a closed set in Ω^n and C_δ is its open δ-neighborhood then

$$\begin{aligned} P_x^\varepsilon(C) &= W(x^\varepsilon \in C, \| x^\varepsilon - x^{\varepsilon,N}\|_\infty > \delta) + W(x^\varepsilon \in C, \| x^\varepsilon - x^{\varepsilon,N}\|_\infty \le \delta) \\ &\le W(x^\varepsilon \in C, \| x^\varepsilon - x^{\varepsilon,N}\|_\infty \le \delta) + W(\| x^\varepsilon - x^{\varepsilon,N}\|_\infty > \delta) \\ &\le W(x^{\varepsilon,N} \in \bar{C}_\delta) + W(\| x^\varepsilon - x^{\varepsilon,N}\|_\infty > \delta). \end{aligned}$$

Combining this last inequality with (A.4), (A.6) we see that

$$\lim \sup_{\varepsilon\downarrow 0} \varepsilon \log P_x^\varepsilon(C) \le -\inf\{I(\omega) \mid \omega \in \bar{C}_\delta\} \le -\inf\{I(\omega) \mid \omega \in C_{2\delta}\}.$$

Now (LD2) and (LD3) imply that

$$\sup_{\delta>0}\inf\{I(\omega) \mid \omega \in C_{2\delta}\} = \inf\{I(\omega) \mid \omega \in C\}$$

and so letting $\delta \downarrow 0$, (LD5) obtains.

REFERENCES

[1] VARADHAN, S. R. S. (1966). Asymptotic probabilities and differential equations. *Comm. Pure Appl. Math.* **XIX** 261–286.

[2] VARADHAN, S. R. S. (1967). Diffusion processes in a small time interval. *Comm. Pure Appl. Math.* **XX** 659–685.

[3] GLASS, M. (1969). Perturbation of a first order equation by a small diffusion. Ph.D. Dissertation, New York University.

[4] WENTZELL, A. D. and FREIDLIN, M. I. (1970). Small random perturbations of dynamical systems. *Russian Math. Surveys* **25** 1–56 [*Uspehi Mat. Nauk.* **25** (1970) 3–55].

[5] AZENCOTT, R. (1978). *Lecture Notes in Math.* **774**. Springer, Berlin.

[6] KALLIANPUR, G. and STRIEBEL, C. (1968). Estimation of stochastic processes. *Ann. Math. Statist.* **39** 785–801.

[7] IKEDA, N. and WATANABE, S. (1981). *Stochastic Differential Equations and Diffusion Processes.* North-Holland, Amsterdam.

[8] STROOCK, D. W. and VARADHAN, S. R. S. (1979). *Multidimensional Diffusion Processes.* Springer, Berlin.

[9] HIJAB, O. (1982). Asymptotic nonlinear filtering. *Proceedings of the 21st IEEE Conference on Decision and Control.* Orlando, Florida.

[10] FRIEDMAN, A. (1976). *Stochastic Differential Equations and Applications, Vol.* 2. Academic, New York.

MATHEMATICS DEPARTMENT
THE OHIO STATE UNIVERSITY
COLUMBUS, OHIO 43210