

Nonlinear Filtering and Large Deviations: A PDE-Control Theoretic Approach*

M. R. JAMES

Mathematics Department and Systems Research Center, University of Maryland, College Park, MD 20742, USA

and

J. S. BARAS

Electrical Engineering Department and Systems Research Center, University of Maryland, College Park, MD 20742, USA

(Received 7 June 1987; in final form 15 September 1987)

We consider the asymptotic nonlinear filtering problem $dx = f(x) dt + \varepsilon^{1/2} dw$, $dy = h(x) dt + \varepsilon^{1/2} dv$, and obtain $\lim_{\varepsilon \rightarrow 0} \varepsilon \log q^\varepsilon(x, t) = -W(x, t)$ for unnormalized conditional densities $q^\varepsilon(x, t)$ using PDE methods. Here, $W(x, t)$ is the value function for a deterministic optimal control problem arising in Mortensen's deterministic estimation, and is the unique viscosity solution of a Hamilton-Jacobi-Bellman equation. Hijab has also studied this filtering problem, and we extend his large deviation result for certain unnormalized conditional measures. The resulting variational problem corresponds to the above control problem.

KEY WORDS: Nonlinear filtering, large deviations, viscosity solutions.

AMS(MOS) SUBJECT CLASSIFICATION: 93E11, 60F10, 49C20.

*Research supported in part through NSF Grant CDR-85-00108, and AFOSR-URI Grant 87-0073.

1. INTRODUCTION

An important problem in nonlinear system theory is the construction of observers for control systems of the form

$$\begin{aligned}\dot{x} &= f(x, u), \\ \dot{y} &= h(x).\end{aligned}\tag{1}$$

Baras and Krishnaprasad [1] have proposed a method for constructing an observer as a limit of nonlinear filters for a family of associated filtering problems (3), parameterised by $\varepsilon > 0$. More recent work in this direction is presented in Baras, Bensoussan and James [2]. It is of interest then to study the asymptotic behaviour of the corresponding unnormalised conditional densities $q^\varepsilon(x, t)$ as $\varepsilon \rightarrow 0$, via the Zakai equation (5). We obtain the asymptotic formula

$$q^\varepsilon(x, t) = \exp\left(-\frac{1}{\varepsilon}(W(x, t) + o(1))\right),\tag{2}$$

as $\varepsilon \rightarrow 0$, where $W(x, t)$ is the value function corresponding to a deterministic optimal control problem, namely that arising in deterministic estimation.

Hijab [11] has studied this asymptotic estimation problem, and obtained a WKB expansion when $W(x, t)$ is smooth. This identifies the limiting filter as Mortensen's deterministic or minimum energy estimator [14]. In addition, Hijab [12] has proved a large deviation principle for the conditional measures for the filtering problem (3). We extend Hijab's large deviation result by allowing random initial conditions in (3), and observe that the resulting variational problem (cf. action functional) is exactly the optimal control problem mentioned above.

The asymptotic formula for the unnormalized conditional densities (Theorem 5.1) and the large deviation principle for the unnormalized conditional measures (Theorem 6.2) characterize the limiting filter in terms of the deterministic estimator.

Our method is inspired by the Work of Fleming and Mitter [7], and Evans and Ishii [6]. A logarithmic transformation is applied to the robust form of the Zakai equation, yielding a Hamilton–Jacobi

equation in the limit. A related Hamilton–Jacobi equation is interpreted as the Bellman equation for the deterministic estimation optimal control problem, of which $W(x, t)$ is the unique viscosity solution. In particular, $W(x, t)$ is not assumed to be smooth.

2. PROBLEM FORMULATION

We consider a family of diffusion processes in \mathbb{R}^n with real valued observations:

$$\begin{aligned} dx^\varepsilon(t) &= f(x^\varepsilon(t)) dt + \varepsilon^{1/2} dw(t), & x^\varepsilon(0) &= x_0^\varepsilon, \\ dy^\varepsilon(t) &= h(x^\varepsilon(t)) dt + \varepsilon^{1/2} dv(t), & y^\varepsilon(0) &= 0. \end{aligned} \tag{3}$$

Here w, v are independent Wiener processes independent of the initial conditions x_0^ε , which have densities

$$q_0^\varepsilon(x) = C_\varepsilon \exp\left(-\frac{1}{\varepsilon} S_0(x)\right) \tag{4}$$

where $\lim_{\varepsilon \rightarrow 0} \varepsilon \log C_\varepsilon = 0$ and $S_0 \geq 0$ is Lipschitz continuous. As $\varepsilon \rightarrow 0$ the trajectories of (3) converge in probability to the trajectory of a corresponding deterministic system.

The *Zakai equation* for an unnormalized conditional density $q^\varepsilon(x, t)$ is

$$\begin{aligned} dq^\varepsilon(x, t) &= A_\varepsilon^* q^\varepsilon(x, t) dt + \frac{1}{\varepsilon} h(x) q^\varepsilon(x, t) dy^\varepsilon(t), \\ q^\varepsilon(x, 0) &= q_0^\varepsilon(x), \end{aligned} \tag{5}$$

where A_ε^* is the formal adjoint of the diffusion operator

$$A_\varepsilon = \frac{\varepsilon}{2} \sum_{i=1}^n \frac{\partial^2}{\partial x_i^2} + \sum_{i=1}^n f_i(x) \frac{\partial}{\partial x_i}.$$

We assume throughout the following: f, h are bounded C^∞ functions with bounded derivatives of orders 1 and 2.

Defining

$$p^\varepsilon(x, t) = \exp\left(-\frac{1}{\varepsilon} y^\varepsilon(t)h(x)\right) q^\varepsilon(x, t), \quad (6)$$

the *robust* form of the Zakai equation is

$$\begin{aligned} \frac{\partial}{\partial t} p^\varepsilon(x, t) - \frac{\varepsilon}{2} \Delta p^\varepsilon(x, t) + Dp^\varepsilon(x, t)g^\varepsilon(x, t) + \frac{1}{\varepsilon} V^\varepsilon(x, t)p^\varepsilon(x, t) &= 0, \\ p^\varepsilon(x, 0) &= q_0^\varepsilon(x), \end{aligned} \quad (7)$$

where

$$g^\varepsilon(x, t) = f(x) - y(t)Dh(x)', \quad (8)$$

$$V^\varepsilon(x, t) = \frac{1}{2}h(x)^2 + y(t)A_\varepsilon h(x) - \frac{1}{2}y(t)^2 |Dh(x)|^2 + \varepsilon \operatorname{div}(f(x) - y(t)Dh(x)'). \quad (9)$$

Note that (7) is a linear parabolic PDE and the coefficient V^ε depends on the observation path $t \mapsto y(t)$. We shall omit the ε -dependence of y , and view (7) as a functional of the observation path $y \in \Omega_0 = C([0, T], \mathbb{R}; y(0) = 0)$. This transformation provides a convenient choice of a version of the conditional density, and under our assumptions we can recover the unnormalized density $q^\varepsilon(x, t)$ from the solution of (7); see for example Pardoux [15].

Following Fleming and Mitter [7], who considered filtering problems with $\varepsilon = 1$, we apply the logarithmic transformation

$$S^\varepsilon(x, t) = -\varepsilon \log p^\varepsilon(x, t). \quad (10)$$

Then $S^\varepsilon(x, t)$ satisfies

$$\begin{aligned} \frac{\partial}{\partial t} S^\varepsilon(x, t) - \frac{\varepsilon}{2} \Delta S^\varepsilon(x, t) + H^\varepsilon(x, t, DS^\varepsilon(x, t)) &= 0, \\ S^\varepsilon(x, 0) &= S_0(x) - \varepsilon \log C_\varepsilon, \end{aligned} \quad (11)$$

where

$$H^\varepsilon(x, t, \lambda) = \lambda g^\varepsilon(x, t) + \frac{1}{2} |\lambda|^2 - V^\varepsilon(x, t). \tag{12}$$

Equation (11) is a nonlinear parabolic PDE, which can be interpreted as the Bellman equation for a stochastic control problem [7].

Formally letting $\varepsilon \rightarrow 0$ we obtain a Hamilton–Jacobi equation

$$\begin{aligned} \frac{\partial}{\partial t} S(x, t) + H(x, t, DS(x, t)) &= 0, \\ S(x, 0) &= S_0(x), \end{aligned} \tag{13}$$

where

$$H(x, t, \lambda) = \lambda g_0(x, t) + \frac{1}{2} |\lambda|^2 - V(x, t), \tag{14}$$

$$g_0(x, t) = f(x) - y(t) Dh(x)', \tag{15}$$

$$V(x, t) = \frac{1}{2} h(x)^2 + y(t) Dh(x) f(x) - \frac{1}{2} y(t)^2 |Dh(x)|^2. \tag{16}$$

Note that $g^\varepsilon \rightarrow g_0$, $V^\varepsilon \rightarrow V$ and $H^\varepsilon \rightarrow H$ uniformly on compact subsets.

We shall interpret solutions of (13) in the viscosity sense. If we define

$$W(x, t) = S(x, t) - y(t)h(x), \quad y \in \Omega_0, \tag{17}$$

then, for $y \in \Omega_0 \cap C^1([0, T], \mathbb{R})$, $W(x, t)$ satisfies a Hamilton–Jacobi equation, which in Section 3 is presented as the Bellman equation for the deterministic estimation control problem.

Our main task is to prove that $S^\varepsilon \rightarrow S$ as $\varepsilon \rightarrow 0$ uniformly on compact subsets. From this the asymptotic formula (2) will follow (Theorem 5.1).

3. DETERMINISTIC ESTIMATION

We begin by reviewing Mortensen’s method [14], [10] of deterministic minimum energy estimation.

Given an observation record $\mathcal{Y}_t = \{y(s), 0 \leq s \leq t\}$, $0 \leq t \leq T$, of the deterministic system

$$\begin{aligned} \dot{x} &= f(x) + u, & x(0) &= x_0, \\ \dot{y} &= h(x) + v, & y(0) &= 0, \end{aligned} \quad (18)$$

we wish to estimate the state at time y , the initial condition x_0 being unknown. Define

$$\tilde{J}_t(x_0, u, v) = S_0(x_0) + \frac{1}{2} \int_0^t (|u(s)|^2 + v(s)^2) ds. \quad (19)$$

A minimum energy input triple (x_0^*, u^*, v^*) given \mathcal{Y}_t is a triple that minimises \tilde{J}_t subject to the constraint that the trajectory of (18) produces the output \mathcal{Y}_t . By replacing $v(s)$ by $\dot{y}(s) - h(x(s))$ in (19) and omitting the $\dot{y}(s)^2$ term, we can formulate an equivalent unconstrained optimal control problem. Define

$$J_t(x_0, u) = S_0(x_0) + \int_0^t L(x(s), u(s), s) ds, \quad (20)$$

where

$$L(x, u, s) = \frac{1}{2}|u|^2 + \frac{1}{2}h(x)^2 - \dot{y}(s)h(x). \quad (21)$$

We now minimise J_t over pairs (x_0, u) . The *deterministic* or minimum energy estimate $\hat{x}(t)$ given \mathcal{Y}_t is defined to be the endpoint of the optimal trajectory $s \mapsto x^*(s)$, $0 \leq s \leq t$, corresponding to a minimum energy pair (x_0^*, u^*) : $\hat{x}(t) = x^*(t)$.

Next, we use dynamic programming to study this problem. The controls $t \mapsto u(t)$ take values $u \in U = \mathbb{R}^n$, and are square integrable. Given such a control, let x_u denote the corresponding trajectory (given a specified initial condition). Following the general scheme presented in Fleming and Rishel [8], define a class of admissible pairs (x_0, u) by

$$\mathcal{U}_{x,t} = \{(x_0, u): x_u(0) = x_0, x_u(t) = x\}; \quad (22)$$

that is, pairs for which the corresponding trajectory passes through a

specified point x at time t . Define a *value function*

$$W(x, t) = \inf_{(x_0, u) \in \mathcal{U}_{x,t}} J_t(x_0, u). \tag{23}$$

Note that this is a reversal of the standard set-up of dynamic programming [8]. By using standard methods, we see that $W(x, t)$ is continuous and formally satisfies the *Bellman equation*

$$\begin{aligned} \frac{\partial}{\partial t} W(x, t) + \tilde{H}(x, t, DW(x, t)) &= 0, \\ W(x, 0) &= S_0(x), \end{aligned} \tag{24}$$

where

$$\tilde{H}(x, t, \lambda) = \max_{u \in U} \{ \lambda(f(x) + u) - L(x, u, t) \}. \tag{25}$$

$W(x, t)$ is the minimum value (if it is attained) of J_t subject to the end point condition $x_u(t) = x$. To obtain $\hat{x}(t)$, one minimizes $W(x, t)$ over x :

$$W(\hat{x}(t), t) \leq W(x, t) \quad \text{for all } x \in \mathbb{R}^n. \tag{26}$$

Notice that the definition (23) for $W(x, t)$ makes sense for $y \in \Omega_0 \cap C^1([0, T], \mathbb{R})$. We can directly interpret (13) as the Bellman equation of another optimal control problem (see (40)–(42) below), with $S(x, t)$ as its value function. This makes sense for all $y \in \Omega_0$, since \dot{y} does not appear. Thus defining $W(x, t)$ by (17) is valid for any $y \in \Omega_0$. If $y \in \Omega_0 \cap C^1([0, T], \mathbb{R})$, these definitions coincide.

Now we prove that $W(x, t)$ is the unique viscosity solution of the Hamilton–Jacobi–Bellman equation (24). Our assumptions imply that $\mathcal{U}_{x,t} \neq \emptyset$ for all $x \in \mathbb{R}^n$, $0 \leq t \leq T$, and consequently $W(x, t) < \infty$. We do not assume existence of optimal controls.

The following definition is taken from Crandall, Evans and Lions [5]. Write $C \equiv C(\mathbb{R}^n \times (0, T), \mathbb{R})$, and similarly $C^1 = C^1(\mathbb{R}^n \times (0, T), \mathbb{R})$.

DEFINITION Let $W \in C$. We say that W is a viscosity subsolution of (24) provided that for all $\phi \in C^1$ the following property holds:

if $W - \phi$ attains a local maximum at a point (x, t) , then

$$\frac{\partial}{\partial t} \phi(x, t) + \tilde{H}(x, t, D\phi(x, t)) \leq 0. \quad (27)$$

We say that W is a viscosity supersolution of (24) provided that for all $\phi \in C^1$ the following property holds:

if $W - \phi$ attains a local minimum at a point (x, t) , then

$$\frac{\partial}{\partial t} \phi(x, t) + \tilde{H}(x, t, D\phi(x, t)) \geq 0. \quad (28)$$

If W is both a viscosity subsolution and supersolution, we say that W is a viscosity solution of (24).

LEMMA 3.1 (Principle of Optimality) *Let $0 \leq t_1 \leq t_2 \leq t$, and choose $(x_0, u) \in \mathcal{U}_{x, t}$. Then*

$$W(x_u(t_2), t_2) \leq W(x_u(t_1), t_1) + \int_{t_1}^{t_2} L(x_u(s), u(s), s) ds. \quad (29)$$

Proof Let $(\tilde{x}_0, \tilde{u}) \in \mathcal{U}_{x_u(t_1), t_1}$. Define

$$\bar{u}(s) = \begin{cases} \tilde{u}(s) & 0 \leq s \leq t_1 \\ u(s) & t_1 \leq s \leq t_2 \end{cases}.$$

Then $\bar{u} \in \mathcal{U}_{x_u(t_2), t_2}$, and hence

$$W(x_u(t_2), t_2) \leq S_0(\tilde{x}_0) + \int_0^{t_1} L(x_{\tilde{u}}(s), \tilde{u}(s), s) ds + \int_{t_1}^{t_2} L(x_u(s), u(s), s) ds.$$

Taking the infimum of the right-hand side over $(\tilde{x}_0, \tilde{u}) \in \mathcal{U}_{x_u(t_1), t_1}$ we obtain (29). \square

Fix (x, t) and choose $\gamma > W(x, t)$. Define

$$\mathcal{U}_{x, t}^\gamma = \{(x_0, u) \in \mathcal{U}_{x, t} : J_t(x_0, u) \leq \gamma\},$$

$$B_\varepsilon = \{x' \in \mathbb{R}^n : |x - x'| \leq \varepsilon\}.$$

LEMMA 3.2 Fix $\varepsilon > 0$. Then there exists $\eta > 0$ such that if $(x_0, u) \in \mathcal{U}_{x,t}^\gamma$ then $x_u(t-h) \in B_\varepsilon$ for all $0 \leq h \leq \eta$.

Proof Note that $x_u(t) = x \in B_\varepsilon$. Define

$$\eta_u = \sup \{h > 0 : x_u(s) \in B_\varepsilon \text{ for all } s \in [t-h, t]\}.$$

Then $|x_u(t - \eta_u) - x| = \varepsilon$. Let

$$\eta = \inf_{(x_0, u) \in \mathcal{U}_{x,t}^\gamma} \eta_u.$$

We want to show that $\eta > 0$. Suppose not; $\eta = 0$. Then there is a sequence $(x_0^n, u^n) \in \mathcal{U}_{x,t}^\gamma$ with $\eta_{u^n} \rightarrow 0$ as $n \rightarrow \infty$. Write $x_n = x_{u^n}$, etc.

Now f is continuous, so there is a constant $K > 0$ such that $|f(x')| \leq K$ for all $x' \in B_\varepsilon$. Then

$$\begin{aligned} 0 < \varepsilon &= |x - x_n(t - \eta_n)| \\ &\leq \int_{t-\eta_n}^t (|f(x_n(s))| + |u_n(s)|) ds \\ &\leq K\eta_n + \int_{t-\eta_n}^t |u_n(s)| ds. \end{aligned}$$

Choose $N_0 > 0$ such that $n \geq N_0$ implies $K\eta_n < \varepsilon/2$. Then

$$0 < \varepsilon/2 \leq \int_{t-\eta_n}^t |u_n(s)| ds \text{ for } n \geq N_0.$$

(Note that if U is bounded, then the lemma follows from this inequality.)

Next, since $(x_0^n, u^n) \in \mathcal{U}_{x,t}^\gamma$ it follows that

$$\int_{t-\eta_n}^t |u(s)|^2 ds \leq \gamma.$$

Then

$$0 < \varepsilon/2 \leq \int_{t-\eta_n}^t |u_n(s)| ds$$

$$\leq \sqrt{\gamma\eta_n} \quad \text{for } n \geq N_0,$$

using the Cauchy–Schwarz inequality, which is impossible since $\sqrt{\eta_n} \rightarrow 0$. Consequently $\eta > 0$ proving the lemma. \square

THEOREM 3.1 *The value function $W(x, t)$ defined by (23) is the unique viscosity solution of the Hamilton–Jacobi–Bellman equation (24).*

Proof First we show that $W(x, t)$ is a viscosity subsolution. Let $\phi \in C^1$ and suppose that $W - \phi$ attains a local maximum at (x, t) . Then there exists $\varepsilon > 0$ such that

$$W(x, t) - \phi(x, t) \geq W(x', t') - \phi(x', t') \quad (30)$$

for all $x' \in B_\varepsilon$, $|t - t'| \leq \varepsilon$.

Choose a constant control $\tilde{u}(s) \equiv u \in U$. There is an x_0 such that $(x_0, u) \in \mathcal{U}_{x,t}$. Choose $0 < \delta \leq \varepsilon$ such that $x_u(s) \in B_\varepsilon$ for $|t - s| \leq \delta$.

The Principle of Optimality (29) implies

$$W(x, t) \leq W(x_u(t-h), t-h) + \int_{t-h}^t L(x_u(s), \tilde{u}(s), s) ds. \quad (31)$$

If $0 \leq h \leq \delta$, then (30) gives

$$W(x, t) - \phi(x, t) \geq W(x_u(t-h), t-h) - \phi(x_u(t-h), t-h). \quad (32)$$

Combining (31) and (32) we obtain

$$\frac{\phi(x_u(t-h), t-h) - \phi(x, t)}{-h} - \frac{1}{h} \int_{t-h}^t L(x_u(s), \tilde{u}(s), s) ds \leq 0.$$

Letting $h \rightarrow 0$ we have

$$\frac{\partial}{\partial t} \phi(x, t) + D\phi(x, t)(f(x) + u) - L(x, u, t) \leq 0.$$

But this holds for all $u \in U$, hence (27) and so $W(x, t)$ is a subsolution of (24).

To see that $W(x, t)$ is a viscosity supersolution, let $\phi \in C^1$ and suppose that $W - \phi$ attains a local minimum at (x, t) . Then there exists an $\varepsilon > 0$ such that

$$W(x, t) - \phi(x, t) \leq W(x', t') - \phi(x', t') \tag{33}$$

for all $x' \in B_\varepsilon$, $|t' - t| \leq \varepsilon$.

Suppose, contrary to (28), that there exists a $\theta > 0$ such that

$$\frac{\partial}{\partial t} \phi(x, t) + \tilde{H}(x, t, D\phi(x, t)) < -\theta < 0.$$

By continuity, reducing $\varepsilon > 0$ if necessary,

$$\frac{\partial}{\partial t} \phi(x', t') + \max_{u \in U} \{D\phi(x', t')(f(x') + u) - L(x', u, t')\} < -\theta < 0 \tag{34}$$

for all $x' \in B_\varepsilon$, $|t - t'| \leq \varepsilon$. Let $\gamma > W(x, t)$ and let η be given as in Lemma 3.2. By the Principle of Optimality (29) we have

$$W(x, t) = \inf_{(x_0, u) \in \mathcal{U}_{x,t}^\gamma} \left\{ W(x_u(t-h), t-h) + \int_{t-h}^t L(x_u(s), u(s), s) ds \right\}. \tag{35}$$

Let $0 < h < \eta \wedge \varepsilon$, and choose $(x_0, u) \in \mathcal{U}_{x,t}^\gamma$ such that

$$W(x_u(t-h), t-h) + \int_{t-h}^t L(x_u(s), u(s), s) ds \leq W(x, t) + \frac{\theta h}{2}. \tag{36}$$

Since $x_u(t-h) \in B_\varepsilon$, we have from (33)

$$W(x_u(t-h), t-h) - \phi(x_u(t-h), t-h) \geq W(x, t) - \phi(x, t). \tag{37}$$

Combining (36) and (37) we have

$$-\frac{\theta}{2} \leq \frac{\phi(x_u(t-h), t-h) - \phi(x, t)}{-h} - \frac{1}{h} \int_{t-h}^t L(x_u(s), u(s), s) ds. \quad (38)$$

However, for $t-h \leq s \leq t$, $x_u(s) \in B_\varepsilon$ and $|t-s| < \varepsilon$, so from (34) we have

$$\frac{\partial}{\partial t} \phi(x_u(s), s) + D\phi(x_u(s), s)(f(x_u(s)) + u(s)) - L(x_u(s), u(s), s) < -\theta.$$

Integrating, we obtain

$$\frac{\phi(x, t) - \phi(x_u(t-h), t-h)}{h} - \frac{1}{h} \int_{t-h}^t L(x_u(s), u(s), s) ds < -\theta. \quad (39)$$

But (38) and (39) contradict each other, so we must have $\theta \leq 0$; proving (28). Thus $W(x, t)$ is a supersolution of (24).

The uniqueness assertion follows from Ishii [13, Theorem 1]. In fact, since $S_0(x)$ is uniformly continuous, it follows that $W(x, t)$ is also uniformly continuous. \square

Finally, we state an optimal control problem for which $S(x, t)$ is the value function. Consider the dynamics

$$\dot{x} = g_0(x, s) + u, \quad x(0) = x_0. \quad (40)$$

We wish to minimize

$$I_t(x_0, u) = S_0(x_0) + \int_0^t \left(\frac{1}{2}|u(s)|^2 + V(x_u(s), s)\right) ds. \quad (41)$$

Denote by $\mathcal{F}_{x,t}$ the corresponding class of admissible pairs (x_0, u) .

Define

$$S(x, t) = \inf_{(x_0, u) \in \mathcal{F}_{x,t}} I_t(x_0, u). \quad (42)$$

The above arguments can be used to prove the following.

THEOREM 3.2 *The value function $S(x, t)$ defined by (42) is the unique viscosity solution of the Hamilton–Jacobi equation (13).*

4. SOME ESTIMATES

Let $S^\varepsilon(x, t)$ be the solution of (11). In this section we obtain estimates for $|S^\varepsilon|$ and $|DS^\varepsilon|$ on compact subsets independent of the parameter ε . These estimates will be used in Section 5 to prove that $S^\varepsilon \rightarrow S$.

THEOREM 4.1 *For every compact subset $Q \subset \mathbb{R}^n \times [0, T]$, there exists $\varepsilon_0 > 0$ and $K > 0$ such that for $0 < \varepsilon < \varepsilon_0$ we have*

$$|S^\varepsilon(x, t)| \leq K, \quad \text{for all } (x, t) \in Q, \tag{43}$$

$$|DS^\varepsilon(x, t)| \leq K, \quad \text{for all } (x, t) \in Q. \tag{44}$$

To prove (43), we use a comparison theorem which depends on the maximum principle for linear parabolic PDE. Let $B_R \subset \mathbb{R}^n$ denote the closed ball centred at 0 with radius $R > 0$, write $\Gamma_R = B_R \times \{0\} \cup \partial B_R \times [0, T]$ and define $Q_R = B_R \times [0, T]$, denoting by Q_R^0 its interior.

LEMMA 4.1 (Maximum Principle, Friedman [11]) *Define*

$$\mathcal{L}w = \frac{\partial}{\partial t} w - \frac{\varepsilon}{2} \Delta w + Dw b^\varepsilon,$$

where b^ε is smooth. If $\mathcal{L}w \leq 0$ ($\mathcal{L}w \geq 0$) in Q_R^0 , then

$$w(x, t) \leq \sup_{(z, s) \in \Gamma_R} w(z, s) \left(\inf_{(z, s) \in \Gamma_R} w(z, s) \leq w(x, t) \right)$$

for all $(x, t) \in Q_R$.

LEMMA 4.2 (Comparison Theorem) *Let S^ϵ be a solution of (11), and define*

$$\tilde{\mathcal{L}}v = \frac{\partial}{\partial t} v - \frac{\epsilon}{2} \Delta v + Dv g^\epsilon + \frac{1}{2} |Dv|^2 - V^\epsilon.$$

Let $w = v - S^\epsilon$. If $\tilde{\mathcal{L}}v \geq 0$ ($\tilde{\mathcal{L}}v \leq 0$) in Q_R^0 , and if $S^\epsilon \leq v$ ($v \leq S^\epsilon$) on Γ_R , then $S^\epsilon \leq v$ ($v \leq S^\epsilon$) in Q_R^0 .

Proof If $\tilde{\mathcal{L}}v \geq 0$, then

$$\frac{\partial}{\partial t} w - \frac{\epsilon}{2} \Delta w + Dw g^\epsilon + \frac{1}{2} (|Dv|^2 - |DS^\epsilon|^2) \geq 0.$$

Now $|Dv|^2 - |DS^\epsilon|^2 = Dw(Dv + DS^\epsilon)$. Set

$$b^\epsilon = g^\epsilon + \frac{1}{2} (Dv + DS^\epsilon).$$

Then $\mathcal{L}w \geq 0$ and on Γ_R , $w(z, s) \geq 0$. Hence $w(x, t) \geq 0$ for all $(x, t) \in Q_R$ by Lemma 5.1. \square

Proof of Theorem 4.1 We now construct a function v such that $\tilde{\mathcal{L}}v \geq 0$ in Q_R^0 and $S^\epsilon \leq v$ on Γ_R , independent of (sufficiently small) $\epsilon > 0$ (Evans and Ishii [6]). Define

$$v(x, t) = \frac{1}{R^2 - |x|^2} + \mu t + M, \tag{45}$$

where the constants $\mu > 0$, $M > 0$ are to be chosen.

We write v_i for v_{x_i} , etc. Then

$$\begin{aligned} \tilde{\mathcal{L}}v &= \mu - \frac{\epsilon}{2} \left(\frac{2n}{(R^2 - |x|^2)^2} + \frac{8|x|^2}{(R^2 - |x|^2)^3} \right) \\ &\quad + \sum_{i=1}^n \frac{2g^{\epsilon_i} x_i}{(R^2 - |x|^2)^2} + \frac{2|x|^2}{(R^2 - |x|^2)^4} - V^\epsilon \\ &\geq \mu - \epsilon C \left(\frac{1}{(R^2 - |x|^2)^2} + \frac{|x|^2}{(R^2 - |x|^2)^3} \right) + \frac{2|x|^2}{(R^2 - |x|^2)^4} - C \end{aligned}$$

$$\geq 0 \text{ in } Q_R^0,$$

for all small $\varepsilon > 0$, provided μ is chosen sufficiently large. Choose M so large that

$$S_0(x) \leq M \text{ for all } x \in B_R.$$

Now $v(x, t) \rightarrow \infty$ as $|x| \rightarrow R$ uniformly in $t \in [0, T]$, hence

$$S^\varepsilon \leq v \text{ in } Q_R^0,$$

and since v is continuous in Q_R^0 , there is a constant $K > 0$ depending on R such that

$$S^\varepsilon(x, t) \leq K \text{ for all } (x, t) \in Q_{R/2},$$

for all sufficiently small $\varepsilon > 0$.

Similarly we can find a lower bound for S^ε on $Q_{R/2}$.

Next we estimate the gradient, using a variant of the techniques used in Evans and Ishii [6], as suggested to us by L. C. Evans. To simplify the notation we write $v = S^\varepsilon$, which from (11) satisfies

$$v_t - \frac{\varepsilon}{2} v_{ii} + \frac{1}{2} v_i v_i + v_i g^{xi} - V^\varepsilon = 0, \tag{46}$$

where we have used the summation convention. Let $Q \subset \subset Q' \subset \subset \mathbb{R}^n \times (0, T)$, where Q, Q' are open and “ $\subset \subset$ ” means “compactly contained in”. Choose ζ such that $\zeta \equiv 1$ on Q and $\zeta \equiv 0$ near $\partial Q'$, and define

$$z = \zeta^2 v_k v_k - \lambda v \tag{47}$$

where the constant $\lambda > 0$ is to be chosen.

Suppose that z attains its maximum over \bar{Q}' at $(x_0, t_0) \in Q'$. Then we have

$$z_i = 0 \tag{48}$$

and

$$0 \leq z_t - \frac{\varepsilon}{2} z_{ii} \tag{49}$$

at the point (x_0, t_0) . Then at this point, using (49),

$$\begin{aligned} 0 &\leq 2\zeta\zeta_t v_k v_k + 2\zeta^2 v_k v_{kt} - \lambda v_t - \varepsilon\zeta_i \zeta_t v_k v_k - \varepsilon\zeta_{ii} v_k v_k - 4\varepsilon\zeta\zeta_t v_k v_{kt} \\ &\quad - \varepsilon\zeta^2 v_{ki} v_{ki} - \varepsilon\zeta^2 v_k v_{kii} + \frac{\varepsilon}{2} \lambda v_{ii} \\ &\leq -\varepsilon C \zeta^2 |D^2 v| + 2\zeta^2 v_k \left(v_t - \frac{\varepsilon}{2} v_{ii} \right)_k + \lambda \left(-v_t + \frac{\varepsilon}{2} v_{ii} \right) + C |Dv|^2 \end{aligned}$$

for ε sufficiently small. Using (46) we find that

$$0 \leq -v_k (\zeta^2 v_i v_i)_k - g^{ei} (\zeta^2 v_k v_k)_i + \frac{\lambda}{2} v_i v_i + C \zeta |Dv|^3 + C |Dv|^2 + \lambda C |Dv| + \lambda C.$$

This together with (48) implies

$$\frac{\lambda}{2} |Dv|^2 \leq C \zeta |Dv|^3 + C |Dv|^2 + \lambda C |Dv| + \lambda C. \quad (50)$$

Let

$$\lambda = \mu [\max \zeta |Dv| + 1], \quad (51)$$

where $\mu > 1$ is to be chosen. Then

$$\zeta |Dv|^3 \leq |Dv|^2 [\max \zeta |Dv| + 1],$$

and from (50),

$$\frac{\mu}{2} |Dv|^2 \leq C |Dv|^2 + C \lambda \mu. \quad (52)$$

Choosing μ so large that $\mu/4 \leq \mu/2 - C$, we have from (52)

$$|Dv|^2 \leq C \lambda \quad \text{at } (x_0, t_0). \quad (53)$$

This implies

$$z \leq C \lambda \quad \text{in } Q'. \quad (54)$$

If it happened that $(x_0, t_0) \in \partial Q'$, then

$$z = -\lambda v \leq C\lambda \quad \text{at } (x_0, t_0),$$

and this also implies (54). But from (54),

$$\max \zeta^2 |Dv|^2 \leq \max z + C\lambda \leq C\lambda,$$

and using the definition (51) we have

$$\max \zeta^2 |Dv|^2 \leq C\mu[\max \zeta |Dv| + 1]$$

which implies

$$\zeta |Dv| \leq C \quad \text{in } Q',$$

and hence

$$|Dv| \leq C \quad \text{in } \bar{Q}.$$

This completes the proof of Theorem 4.1. \square

5. MAIN RESULT

We are now in a position to state and prove our main result.

THEOREM 5.1 *Under the above assumptions, we have*

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \log q^\varepsilon(x, t) = -W(x, t) \tag{55}$$

uniformly on compact subsets of $\mathbb{R}^n \times [0, T]$, where $W(x, t)$ is defined by (17).

Proof Using Crandall and Lions [4, Lemma 5.2] we obtain uniform estimates of the Hölder norm of S^ε as a function of $t \in [0, T]$. From Theorem 4.1 and the Arzela–Ascoli theorem, there is a subsequence $\varepsilon_k \rightarrow 0$ such that S^{ε_k} converges uniformly on compact subsets to a continuous function \tilde{S} . By the “vanishing viscosity” theorem (Crandall and Lions [3]), \tilde{S} is a viscosity solution of (13). By uniqueness, Theorem 3.2, $\tilde{S} = S$. In fact, $S^\varepsilon \rightarrow S$ as $\varepsilon \rightarrow 0$.

From this we have

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \log q^\varepsilon(x, t) = -(S(x, t) - y(t)h(x))$$

uniformly on compact subsets, for $y \in \Omega_0$. Using the definition (17) of $W(x, t)$ completes the proof. \square

This pathwise asymptotic result can be used to obtain a corresponding probabilistic result. Let $z \in \Omega_0$ denote the limit in probability of $y^\varepsilon \in \Omega_0$. Recall that S^ε, W , etc. are functionals of $y \in \Omega_0$. Define $W = W[z], S^\varepsilon = S^\varepsilon[y^\varepsilon]$, etc. Then $\lim_{\varepsilon \rightarrow 0} \varepsilon \log q^\varepsilon(x, t) = -\lim_{\varepsilon \rightarrow 0} (S^\varepsilon(x, t) - y^\varepsilon(t)h(x)) = -W(x, t)$ in probability.

6. LARGE DEVIATIONS

We have seen that the optimal control problem associated with deterministic estimation plays a key role in studying the asymptotics of the Zakai equation (5). In this section we shall see that this control problem is exactly the variational problem arising in a large deviation principle for certain conditional measures.

We begin by reviewing the results in Hijab [12]. Fix x_0 and consider the stochastic differential equation (3), with initial condition $x_0^\varepsilon = x_0$ for all $\varepsilon > 0$. Let $Q_{x|y, x_0}^\varepsilon$ be an unnormalised conditional measure on $\Omega^n = C([0, T], \mathbb{R}^n)$ of x^ε given $y \in \Omega_0$ and the initial condition x_0 . As in Section 3, given a control $t \rightarrow u(t)$, let x_u denote the corresponding trajectory of (18). Hijab [12] proved the following.

THEOREM 6.1 *For any open subset \mathcal{O} and any closed subset \mathcal{C} of Ω^n ,*

$$\liminf_{\varepsilon \rightarrow 0} \varepsilon \log Q_{x|y, x_0}^\varepsilon(\mathcal{O}) \geq -I(x_0, y, \mathcal{O})$$

$$\limsup_{\varepsilon \rightarrow 0} \varepsilon \log Q_{x|y, x_0}^\varepsilon(\mathcal{C}) \leq -I(x_0, y, \mathcal{C})$$

where for $\mathcal{A} \subset \Omega^n$,

$$I(x_0, y, \mathcal{A}) = \inf_u \left\{ \frac{1}{2} \int_0^T (|u(s)|^2 + h(x_u(s))^2) ds \right\}$$

$$-\int_0^T h(x_u(s)) dy(s) \Big|_{x_u(0)=x_0, x_u \in \mathcal{A}} \Big\}, \tag{56}$$

with the understanding that the infimum over an empty set is infinite.

Now let the initial conditions of (3) be random with density defined by (4). Let $Q_{(x, x_0)|y}^\varepsilon$ be an unnormalized joint conditional measure of $(x^\varepsilon, x_0^\varepsilon)$ on $\Omega^n \times \mathbb{R}^n$ given $y \in \Omega_0$.

THEOREM 6.2 *For any open subset \mathcal{O} and any closed subset \mathcal{C} of Ω^n , and for any open subset \mathcal{O}_0 and any closed subset \mathcal{C}_0 of \mathbb{R}^n , we have*

$$\liminf_{\varepsilon \rightarrow 0} \varepsilon \log Q_{(x, x_0)|y}^\varepsilon(\mathcal{O} \times \mathcal{O}_0) \geq -J(\mathcal{O} \times \mathcal{O}_0, y) \tag{57}$$

$$\limsup_{\varepsilon \rightarrow 0} \varepsilon \log Q_{(x, x_0)|y}^\varepsilon(\mathcal{C} \times \mathcal{C}_0) \leq -J(\mathcal{C} \times \mathcal{C}_0, y) \tag{58}$$

where for $\mathcal{A} \times \mathcal{A}_0 \subset \Omega^n \times \mathbb{R}^n$,

$$J(\mathcal{A} \times \mathcal{A}_0, y) = \inf_{x_0 \in \mathcal{A}_0} \{S_0(x_0) + I(x_0, y, \mathcal{A})\}. \tag{59}$$

To prove this theorem we employ the following version of Laplace's asymptotic method, adapted from Freidlin and Wentzell [9].

LEMMA 6.1 *Let $f, g: \mathbb{R}^n \rightarrow \mathbb{R}$ be Borel measurable, bounded below, and let C_ε be a family of positive real numbers such that $\lim_{\varepsilon \rightarrow 0} \varepsilon \log C_\varepsilon = 0$. Assume that f is upper semicontinuous. Then for any open subset A and any closed subset B of \mathbb{R}^n we have*

$$\liminf_{\varepsilon \rightarrow 0} \varepsilon \log \int_A C_\varepsilon \exp\left(-\frac{1}{\varepsilon} f(x)\right) dx \geq -\inf_{x \in A} f(x), \tag{60}$$

$$\limsup_{\varepsilon \rightarrow 0} \varepsilon \log \int_B C_\varepsilon \exp\left(-\frac{1}{\varepsilon} g(x)\right) dx \leq -\inf_{x \in B} g(x). \tag{61}$$

Proof Let $m = \inf_{x \in A} f(x)$. If $m = \infty$, the result is clear; so assume $m < \infty$. For any $\delta > 0$ define

$$A_\delta = \{x \in A: f(x) < m + \delta, |x| < R\},$$

where R is chosen large enough to ensure $A_\delta \neq \emptyset$. Then A_δ is a bounded open subset of A , and

$$\begin{aligned} \int_A C_\varepsilon \exp\left(-\frac{1}{\varepsilon} f(x)\right) dx &\geq \int_{A_\delta} C_\varepsilon \exp\left(-\frac{1}{\varepsilon} (m + \delta)\right) dx \\ &\geq K_\delta C_\varepsilon \exp\left(-\frac{1}{\varepsilon} (m + \delta)\right), \end{aligned}$$

and hence

$$\liminf_{\varepsilon \rightarrow 0} \varepsilon \log \int_A C_\varepsilon \exp\left(-\frac{1}{\varepsilon} f(x)\right) dx \geq -(m + \delta).$$

This holds for all $\delta > 0$, hence (60) follows.

Next, write $m = \inf_{x \in B} g(x)$ and assume $m < \infty$. Then

$$\int_B C_\varepsilon \exp\left(-\frac{1}{\varepsilon} g(x)\right) dx \leq \int_B C_\varepsilon \exp\left(-\frac{1}{\varepsilon} m\right) dx,$$

from which (61) follows when B is bounded. The general case follows by approximation. \square

Proof of Theorem 6.2 From Theorem 6.1, for any $\delta > 0$ there exists $\varepsilon_0 > 0$ such that for $0 < \varepsilon < \varepsilon_0$,

$$Q_{x|(y, x_0)}^\varepsilon(\mathcal{O}) \geq \exp\left(-\frac{1}{\varepsilon} (I(x_0, y, \mathcal{O}) + \delta)\right).$$

Then

$$\begin{aligned} Q_{(x, x_0)|y}^\varepsilon(\mathcal{O} \times \mathcal{O}_0) &= \int_{\mathcal{O}_0} Q_{x|(y, x_0)}^\varepsilon(\mathcal{O}) q_0^\varepsilon(x_0) dx_0 \\ &\geq \int_{\mathcal{O}_0} C_\varepsilon \exp\left(-\frac{1}{\varepsilon} (S_0(x_0) + I(x_0, y, \mathcal{O}) + \delta)\right) dx_0. \end{aligned}$$

Note that $I(x_0, y, \mathcal{O})$ is upper semicontinuous as a function of $x_0 \in \mathbb{R}^n$ since \mathcal{O} is open. Applying (60) we have

$$\liminf_{\varepsilon \rightarrow 0} \varepsilon \log Q_{(x, x_0)|y}^\varepsilon(\mathcal{O} \times \mathcal{O}_0) \geq -J(\mathcal{O} \times \mathcal{O}_0, y) - \delta.$$

However, $\delta > 0$ was arbitrary; hence (57).

The estimate (58) follows from

$$Q_{x|y, x_0}^\varepsilon(\mathcal{C}) \leq \exp\left(-\frac{1}{\varepsilon}(I(x_0, y, \mathcal{C}) - \delta)\right)$$

for ε sufficiently small, using (61). \square

Note that the variational problem (59) corresponds to the optimal control problem (18)–(23) discussed in Section 3. Theorem 6.2 implies that the limiting measure is concentrated on the optimal initial condition x_0^* and optimal trajectory $x^*(s)$, $0 \leq s \leq T$.

Acknowledgments

We wish to thank Professor L. C. Evans for his time and invaluable assistance. We also thank Dr. Alan Weiss for his useful comments regarding large deviations, and the referees for their careful reading of the paper.

References

- [1] J. S. Baras and P. S. Krishnaprasad, Dynamic observers as asymptotic limits of recursive filters, *IEEE Proc.* 21st CDC, Orlando, Florida, Dec., 1982, pp. 1126–1127.
- [2] J. S. Baras, A. Bensoussan and M. R. James, Dynamic observers as asymptotic limits to recursive filters: Special cases, Technical Report SRC-TR-86-79, Systems Research Center, University of Maryland, Dec., 1986. To appear, *SIAM J. Applied Math.*
- [3] M. G. Crandall and P. L. Lions, Viscosity solutions of Hamilton–Jacobi equations, *Trans. AMS* **277**(1) (1983), 1–42.
- [4] M. G. Crandall and P. L. Lions, Two approximations of solutions of Hamilton–Jacobi equations, *Math. Comput.* **43**(167) (1984), 1–19.
- [5] M. G. Crandall, L. C. Evans and P. L. Lions, Some properties of viscosity solutions of Hamilton–Jacobi equations, *Trans. AMS.* **282**(2) (1984), 487–502.

- [6] L. C. Evans and H. Ishii, A PDE approach to some asymptotic problems concerning random differential equations with small noise intensities, *Ann. Inst. Henri Poincaré-Analyse non linéaire* **2**(1) (1985), 1–20.
- [7] W. H. Fleming and S. K. Mitter, Optimal control and nonlinear filtering for nondegenerate diffusion processes, *Stochastics* **8** (1982), 63–77.
- [8] W. H. Fleming and R. W. Rishel, *Deterministic and Stochastic Optimal Control*, Springer-Verlag, New York, 1975.
- [9] M. I. Freidlin and A. D. Wentzell, *Random Perturbations of Dynamical Systems*, Springer-Verlag, New York, 1984.
- [10] A. Friedman, *Partial Differential Equations of Parabolic Type*, Prentice-Hall, Englewood Cliffs, NJ, 1964.
- [11] O. Hijab, Minimum energy estimation, Ph.D. Dissertation, University of California, Berkeley, Dec., 1980.
- [12] O. Hijab, Asymptotic Bayesian estimation of a first order equation with small diffusion, *Annals of Probability* **12** (1984), 890–902.
- [13] H. Ishii, Existence and uniqueness of solutions of Hamilton–Jacobi equations, *Funkcialaj Ekvacioj* **29** (1986), 167–188.
- [14] R. E. Mortensen, Maximum-likelihood recursive nonlinear filtering, *J. Opt. Theory and Appl.* **2**(6) (1968), 386–394.
- [15] E. Pardoux, Équations du filtrage non linéaire de la prédiction et du lissage, *Stochastics* **6** (1982), 193–231.