TR-87-27

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

M.R. James   J.S. Baras

Nonlinear Filtering and Large Deviations: A PDE–Control Theoretic Approach

M. R. James[1] J. S. Baras[2]

February 26, 1987

[1]Mathematics Department and Systems Research Center, University of Maryland, College Park, MD, 20742, USA.
[2]Electrical Engineering Department and Systems Research Center, University of Maryland, College Park, MD, 20742, USA.

Abstract

We consider the asymptotic nonlinear filtering problem $dx = f(x)dt + \sqrt{\epsilon}dw$, $dy = h(x)dt + \sqrt{\epsilon}dv$, and obtain $\lim_{\epsilon \to 0} \epsilon \log q^{\epsilon}(x,t) = -W(x,t)$ for unnormalised conditional densities $q^{\epsilon}(x,t)$ 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 unnormalised conditional measures. The resulting variational problem corresponds to the above control problem.

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 $\epsilon > 0$. 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 $q^\epsilon(x,t)$ as $\epsilon \to 0$, via the Zakai equation (5). We obtain the asymptotic formula

$$q^\epsilon(x,t) = e^{-\frac{1}{\epsilon}(W(x,t)+o(1))}, \tag{2}$$

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

Hijab [10] 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 [13]. In addition, Hijab [11] 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 (c.f. action functional) is exactly the optimal control problem mentioned above.

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

Our method is inspired by the work of Fleming and Mitter [6], and Evans and Ishii [5]. 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.

Acknowledgements: 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.

2 Problem Formulation

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

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

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

$$q_0^\epsilon(x) = C_\epsilon e^{-\frac{1}{\epsilon}S_0(x)} \tag{4}$$

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

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

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

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

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

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

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

the *robust* form of the Zakai equation is

$$\tfrac{\partial}{\partial t}p^\epsilon(x,t) - \frac{\epsilon}{2}\Delta p^\epsilon(x,t) + Dp^\epsilon(x,t)g^\epsilon(x,t) + \frac{1}{\epsilon}V^\epsilon(x,t)p^\epsilon(x,0) = 0, \tag{7}$$
$$p^\epsilon(x,t) \;=\; q_0^\epsilon(x),$$

where

$$g^\epsilon(x,t) \;=\; f(x) - y(t)Dh(x)', \tag{8}$$

$$\begin{aligned} V^\epsilon(x,t) \;&=\; \frac{1}{2}h(x)^2 \;+\; y(t)A_\epsilon h(x) \\ &\quad -\frac{1}{2}y(t)^2 \mid Dh(x) \mid^2 \;+ \epsilon\mathrm{div}\left(f(x) - y(t)Dh(x)'\right). \end{aligned} \tag{9}$$

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 $y \in \Omega_0 = C([0,T], \mathbb{R}^n;\; y(0) = 0)$. This transformation provides a convenient choice of a version of the conditional density, and under our assumptions we can recover the unnormalised density $q^\epsilon(x,t)$ from solutions of (7); see for example Pardoux [14].

Following Fleming and Mitter [6], who considered filtering problems with $\epsilon = 1$, we apply the logarithmic transformation

$$S^\epsilon(x,t) \;=\; -\epsilon \log p^\epsilon(x,t). \tag{10}$$

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

$$\tfrac{\partial}{\partial t}S^\epsilon(x,t) - \frac{\epsilon}{2}\Delta S^\epsilon(x,t) + H^\epsilon(x,t,DS^\epsilon(x,t)) \;= 0, \tag{11}$$
$$S^\epsilon(x,0) \;=\; S_0(x),$$

where

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

Equation (11) is a nonlinear parabolic PDE, which can be interpreted as the Bellman equation for a stochastic control problem [6].

Formally letting $\epsilon \to 0$ we obtain a Hamilton–Jacobi equation

$$\tfrac{\partial}{\partial t}S(x,t) + H(x,t,DS(x,t)) \;=\; 0, \tag{13}$$

$$S(x,0) = S_0(x),$$

where

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

$$g_0(x,t) = f(x) - y(t)Dh(x)', \quad (15)$$

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

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

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

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

then, for $y \in \Omega_0 \cap C^1$, $W(x,t)$ 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 $S^\epsilon \to S$ as $\epsilon \to 0$ 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 [13], [10] of deterministic minimum energy estimation.

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

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

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

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

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

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

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 [7], 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*

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

Note that this is a reversal of the standard set–up of dynamic programming [7]. By using standard methods, we see that $W(x, t)$ is continuous and formally satisfies the *Bellman equation*

$$\frac{\partial}{\partial t} W(x, t) + \tilde{H}(x, t, DW(x, t)) = 0, \quad (24)$$

$$W(x, 0) = S_0(x),$$

where

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

$W(x, t)$ 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 minimises $W(x, t)$ over x:

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

Notice that the definition (23) for $W(x,t)$ makes sense for $y \in \Omega_0 \cap C^1$. 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 $y \in \Omega_0 \cap C^1$, 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 f is a complete vector field. Therefore $\mathcal{U}_{x,t} \neq \emptyset$ 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 [4]. Write $C \equiv C(\mathbb{R}^n \times (0,T),\ \mathbb{R})$, and similarly for C^1.

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

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

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

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

If W is both a viscosity subslution 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*

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

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

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

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). □

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

$$\begin{aligned} \mathcal{U}^{\gamma}_{x,t} &= \{(x_0, u) \in \mathcal{U}_{x,t} \;:\; J_t(x_0, u) \leq \gamma\}, \\ B_\epsilon &= \{x' \in \mathbb{R}^n \;:\; \mid x - x' \mid \leq \epsilon\}. \end{aligned}$$

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

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

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

Then $\mid x_u(t - \eta_u) - x \mid = \epsilon$. Let

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

We want to show that $\eta > 0$. Suppose not; $\eta = 0$. Then there is a sequence $(x^n_0, u^n) \in \mathcal{U}^{\gamma}_{x,t}$ with $\eta_{u_n} \to 0$ as $n \to \infty$. Write $x_n = x_{u_n}$, etc.

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

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

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

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

(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

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

Then

$$\begin{aligned} 0 \ < \ \epsilon/2 \ &\leq \ \int_{t-\eta_n}^{t} \mid u_n(s) \mid ds \\ &\leq \ \sqrt{\gamma\eta_n} \text{ for } n \geq N_0, \end{aligned}$$

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

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 $\epsilon > 0$ such that

$$W(x,t) - \phi(x,t) \ \geq \ W(x',t') - \phi(x',t') \tag{30}$$

for all $x' \in B_\epsilon$, $\mid t - t' \mid \leq \epsilon$.

Choose a constant control $u(s) \equiv u \in U$. There is an x_0 such that $(x_0, u) \in \mathcal{U}_{x,t}$. Choose $0 < \delta \leq \epsilon$ such that $x_u(s) \in B_\epsilon$ for $\mid t - s \mid \leq \delta$. Set $t' = t - s$, $x' = x_u(t')$. Select $(x_0', u') \in \mathcal{U}_{x',t'}$ and define

$$\tilde{u}(s) = \begin{cases} u'(s) & 0 \leq s < t' \\ u & t' \leq s \leq t. \end{cases}$$

The Principle of Optimality (29) implies

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

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

$$W(x,t) - \phi(x,t) \;\geq\; W(x_{\tilde{u}}(t-h), t-h) - \phi(x_{\tilde{u}}(t-h), t-h). \qquad (32)$$

Combining (31) and (32) we obtain

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

Letting $h \to 0$ we have

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

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 $\epsilon > 0$ such that

$$W(x,t) - \phi(x,t) \;\leq\; W(x',t') - \phi(x',t',) \qquad (33)$$

for all $x' \in B_\epsilon$, $\mid t' - t \mid \leq \epsilon$.

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

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

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

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

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

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

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

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

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

$$W(x_u(t-h), t-h) - \phi(x_u(t-h), t-h) \;\geq\; W(x,t) - \phi(x,t). \qquad (37)$$

Combining (36) and (37) we have

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

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

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

Integrating, we obtain

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

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 [12], Theorem 1. In fact, since $S_0(x)$ is uniformly continuous, it follows that $W(x,t)$ is also uniformly continuous. □

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

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

We wish to minimise

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

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

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

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

Theorem 4.1 *For every compact subset $Q \subset I\!R^n \times [0,T]$, there exists $\epsilon_0 > 0$ and $K > 0$ such that for $0 < \epsilon < \epsilon_0$ we have*

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

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

To prove (43), we use a comparison theorem which depends on the maximum principle for linear parabolic PDE. Let $B_R \subset I\!R^n$ 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 [10]*) Define*

$$\mathcal{L}w = \tfrac{\partial}{\partial t}w - \frac{\epsilon}{2}\Delta w + Dwb^\epsilon,$$

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

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

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

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

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

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

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

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

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–Ishii [5]). 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

$$\begin{aligned}
\tilde{\mathcal{L}}v \;&=\; \mu - \frac{\epsilon}{2}\left(\frac{2n}{(R^2 - | x |^2)^2} + \frac{8 | x |^2}{(R^2 - | x |^2)^3}\right) \\
&\quad + \sum_{i=1}^{n} \frac{2g^{\epsilon i}x_i}{(R^2 - | x |^2)^2} + \frac{2 | x |^2}{(R^2 - | x |^2)^4} - V^\epsilon \\
&\geq\; \mu - \epsilon C\left(\frac{1}{(R^2 - | x |^2)^2} + \frac{1 | x |^2}{(R^2 - | x |^2)^3}\right) \\
&\quad + \frac{2 | x |^2}{(R^2 - | x |^2)^4} - C \\
&\geq\; 0 \text{ in } Q_R^0,
\end{aligned}$$

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

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

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

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

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

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

for all sufficiently small $\epsilon > 0$.

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 [5], as suggested by Evans. To simplify the notation we write $v = S^\epsilon$, which from (11) satisfies

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

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 ς such that $\varsigma \equiv 1$ on Q and $\varsigma \equiv 0$ near $\partial Q'$, and define

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

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

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

$$0 \ \le \ z_t - \frac{\epsilon}{2} z_i z_i \tag{49}$$

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

$$\begin{aligned} 0 \ \le \ & 2\varsigma\varsigma_t v_k v_k + 2\varsigma^2 v_k v_{kt} - \lambda v_t \\ & -\epsilon\varsigma_i\varsigma_i v_k v_k - \epsilon\varsigma\varsigma_i v_k v_k - 4\epsilon\varsigma\varsigma_i v_k v_{ki} \\ & -\epsilon\varsigma^2 v_{ki} v_{ki} - \epsilon\varsigma^2 v_k v_{kii} + \frac{\epsilon}{2}\lambda v_{ii} \\ \le \ & -\epsilon C \varsigma^2 \mid D^2 v \mid + 2\varsigma^2 v_k \left(v_t - \frac{\epsilon}{2} v_{ii} \right)_k \\ & + \lambda \left(-v_t + \frac{\epsilon}{2} v_{ii} \right) + C \mid Dv \mid^2 \end{aligned}$$

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

$$\begin{aligned} 0 \ \le \ & -v_k \left(\varsigma^2 v_i v_i \right)_k - g^{\epsilon i} \left(\varsigma^2 v_k v_k \right)_i + \frac{\lambda}{2} v_i v_i + C\varsigma \mid Dv \mid^3 \\ & + C \mid Dv \mid^2 + \lambda C \mid Dv \mid + \lambda C. \end{aligned}$$

This together with (48) implies

$$\frac{\lambda}{2} \mid Dv \mid^2 \leq \; C\varsigma \mid Dv \mid^3 + C \mid Dv \mid^2 + \lambda C \mid Dv \mid + \lambda C. \tag{50}$$

Let

$$\lambda \; = \; \mu[\max \varsigma \mid Dv \mid +1] \tag{51}$$

Then

$$\varsigma \mid Dv \mid^3 \leq \mid Dv \mid^2 [\max \varsigma \mid Dv \mid +1],$$

and from (50),

$$\frac{\mu}{2} \mid Dv \mid^2 \leq \; C \mid Dv \mid^2 + C\lambda\mu. \tag{52}$$

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

$$\mid Dv \mid^2 \leq \; C\lambda \text{ at } (x_0, t_0). \tag{53}$$

This implies

$$z \; \leq \; C\lambda \text{ in } Q'. \tag{54}$$

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

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

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

$$\max \varsigma^2 \mid Dv \mid^2 \leq \; \max z + C\lambda \; \leq \; C\lambda,$$

and using the definition (51) we have

$$\max \varsigma^2 \mid Dv \mid^2 \leq \; C\mu[\max \varsigma \mid Dv \mid +1]$$

which implies

$$\varsigma \mid Dv \mid \leq C \text{ in } Q',$$

and hence

$$\mid Dv \mid \leq \; C \text{ in } \bar{Q}.$$

This completes the proof of Theorem 4.1. □

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*

$$\lim_{\epsilon \to 0} \epsilon \log q^\epsilon(x,t) \;=\; -W(x,t) \tag{55}$$

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

Proof: From Theorem 4.1 and the Arzela–Ascoli theorem, there is a subsequence $\epsilon_k \to 0$ 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, $S^\epsilon \to S$ as $\epsilon \to 0$.

From this we have

$$\lim_{\epsilon \to 0} \epsilon \log q^\epsilon(x,t) \;=\; -\left(S(x,t) - y(t)h(x)\right)$$

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

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 [11]. Fix x_0 and consider the stochastic differential equation (3), with initial condition $x_0^\epsilon = x_0$ for all $\epsilon > 0$. Let $Q^\epsilon_{x|(y,x_0)}$ 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 $t \to u(t)$, let x_u denote the corresponding trajectory of (18). Hijab [11] proved the following.

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

$$\liminf_{\epsilon\to 0} \epsilon \log Q^{\epsilon}_{x|(y,x_0)}(\mathcal{O}) \geq -I(x_0, y, \mathcal{O})$$

$$\limsup_{\epsilon\to 0} \epsilon \log Q^{\epsilon}_{x|(y,x_0)}(\mathcal{C}) \leq -I(x_0, y, \mathcal{C})$$

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

$$I(x_0, y, \mathcal{A}) = \inf_u \left\{ \frac{1}{2}\int_0^T \left(| u(s) |^2 + h(x_u(s))^2 \right) ds - \int_0^T h(x_u(s)) dy(s) \mid x_u(0) = x_0,\ x_u \in \mathcal{A} \right\}, \tag{56}$$

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

Now let the initial conditions of (3) be random with unnormalised density defined by (4). Let $Q^{\epsilon}_{(x,x_0)|y}$ be an unnormalised joint conditional measure of $(x^{\epsilon}, x_0^{\epsilon})$ 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 bounded subset $\mathcal{C}_0$ of $\mathbb{R}^n$, we have*

$$\liminf_{\epsilon\to 0} \epsilon \log Q^{\epsilon}_{(x,x_0)|y}(\mathcal{O} \times \mathcal{O}_0) \geq -J(\mathcal{O} \times \mathcal{O}_0, y) \tag{57}$$

$$\limsup_{\epsilon\to 0} \epsilon \log Q^{\epsilon}_{(x,x_0)|y}(\mathcal{C} \times \mathcal{C}_0) \leq -J(\mathcal{C} \times \mathcal{C}_0, y) \tag{58}$$

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 [8].

Lemma 6.1 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be Borel measurable, bounded below, and let C_ϵ be a family of positive real numbers such that $\lim_{\epsilon\to 0} \epsilon \log C_\epsilon = 0$. Then for any Borel subset A and any bounded Borel subset B of $\mathbb{R}^n$ we have*

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

$$\limsup_{\epsilon\to 0} \epsilon \log \int_B C_\epsilon \exp\left(-\frac{1}{\epsilon} f(x)\right) dx \;\leq\; -\inf_{x\,\in\,B} f(x). \tag{61}$$

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

$$A_\delta \;=\; \{x \in A \;:\; f(x) \leq m + \delta,\ \mid x \mid \leq R\},$$

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

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

and hence

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

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

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

$$\int_B C_\epsilon \exp\left(-\frac{1}{\epsilon} f(x)\right) dx \;\leq\; \int_B C_\epsilon \exp\left(-\frac{1}{\epsilon} m\right) dx,$$

from which (61) follows. □

Proof of Theorem 6.2: From Theorem 6.1, for any $\delta > 0$ there exists $\epsilon_0 > 0$ such that for $0 < \epsilon < \epsilon_0$,

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

Then

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

Applying (60) we have

$$
\liminf_{\epsilon\to 0} \epsilon \log Q^{\epsilon}_{(x,x_0)|y}(\mathcal{O}\times\mathcal{O}_0) \ \geq\ -J(\mathcal{O}\times\mathcal{O}_0, y) - \delta.
$$

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

The estimate (58) follows from

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

for ϵ sufficiently small, using (61). □

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 $x^*(s)$, $0 \leq s \leq T$.

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, pp1126–1127.

[2] J. S. Baras, A. Bensoussan and M. R. James, *Dynamic Observers as Asymptotic Limits of Recursive Filters: Special Cases*, Technical Report SRC–TR–86–79, Systems Research Center, University of Maryland, December, 1986. (Submitted)

[3] M. G. Crandall and P. L. Lions, *Viscosity Solutions of Hamilton–Jacobi Equations*, Trans. AMS, 277, No.1, May 1983, pp1–42.

[4] M. G. Crandall, L. C. Evans and P. L. Lions, *Some Properties of Viscosity Solutions of Hamilton–Jacobi Equations*, Trans. AMS, 282, No.2, April 1984, pp487–502.

[5] 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 linéire, 2, No.1, 1985, pp1–20.

[6] W. H. Fleming and S. K. Mitter, *Optimal Control and Nonlinear Filtering for Nondegenerate Diffusion Processes*, Stochastics, 8, 1982, pp63–77.

[7] W. H. Fleming and R. W. Rishel, *Deterministic and Stochastic Optimal Control*, Springer–Verlag, New York, 1975.

[8] M. I. Freidlin and A. D. Wentzell, *Random Perturbations of Dynamical Systems*, Springer–Verlag, New York, 1984.

[9] A. Friedman, *Partial Differential Equations of Parabolic Type*, Prentice–Hall, Englewood Cliffs NJ, 1964.

[10] O. Hijab, *Minimum Energy Estimation*, PhD Dissertation, University of California, Berkely, December, 1980.

[11] O. Hijab, *Asymptotic Bayesian Estimation of a First Order Equation with Small Diffusion*, Annals of Probability, 12, 1984, pp890–902.

[12] H. Ishii, *Existence and Uniqueness of Solutions of Hamilton–Jacobi Equations*, Preprint.

[13] R. E. Mortensen, *Maximum–Likelihood Recursive Nonlinear Filtering*, J. Opt. Theory and Appl., 2, No. 6, 1968, pp386–394.

[14] E. Pardoux, *Équations du Filtrage non Linéaire de la Prédiction et du Lissage*, Stochastics, 6, 1982, pp193–231.