MARYLAND

TECHNICAL RESEARCH REPORT

SRC TR 88-28

Asymptotic Nonlinear Filtering and Large Deviations with Application to Observer Design

by

Matthew R. James

SYSTEMS RESEARCH CENTER

UNIVERSITY OF MARYLAND

COLLEGE PARK, MARYLAND 20742

ABSTRACT

Title of Dissertation: Asymptotic Nonlinear Filtering and Large Deviations with Application to Observer Design

Matthew Ronald James, Doctor of Philosophy, 1988

Dissertation directed by: J. S. Baras
Professor, Department of Electrical Engineering,
and Director, Systems Research Center
L. C. Evans
Professor, Department of Mathematics

An important problem in control theory is the design of observers for nonlinear control systems. By observer we mean a deterministic dynamical system which uses observed information to compute an estimate of the state of the control system in such a way that the error decays to zero. Baras and Krishnaprasad have proposed that an observer design might result from a study of an asymptotic nonlinear filtering problem obtained by adding small noise terms to the equations defining the control system. The purpose of this thesis is to study this asymptotic filtering problem and to develop observer designs based on their idea.

Asymptotic nonlinear filtering problems have been studied by several authors, and are closely related to large deviations (Wentzell–Freidlin theory). We prove using vanishing viscosity and control theoretic methods a logarithmic limit result for solutions of the Zakai equation. This limit is characterised by a Hamilton–Jacobi equation which, as noted by Hijab, arises in Mortensen's deterministic minimum energy estimation. We make a careful study of this equation in the light of the relatively recent theory of viscosity solutions due to Crandall and Lions. We study the weak limit of the conditional measures and filters. Inspired by Hijab's large deviation result for pathwise conditional measures, we obtain a large deviation principle "in probability" for the conditional measures, and also a large deviation principle for the distributions of these measures.

This asymptotic analysis suggests that the limiting filter is a candidate observer. We present an exact infinite dimensional observer for uncontrolled observable systems. In the case of uncontrolled nonlinear dynamics and linear observations, Bensoussan obtained a finite dimensional observer which is an approximation to the limiting filter. A detectability condition was used to prove exponential decay of the error, provided the initial condition lies in a bounded region. We extend his approach to the general case of controlled nonlinear dynamics and nonlinear observation. In particular, we obtain an observer for a class of fully nonlinear systems with no constraints on the initial conditions. The Beneš case is considered.

ASYMPTOTIC NONLINEAR FILTERING AND LARGE DEVIATIONS WITH APPLICATION TO OBSERVER DESIGN

by

Matthew Ronald James

Dissertation submitted to the Faculty of the Graduate School
of The University of Maryland in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
1988

Advisory Committee:

Professor J. S. Baras
Professor G. L. Blankenship
Professor L. C. Evans
Professor M. I. Freidlin
Professor P. S. Krishnaprasad

In memory of my father

ACKNOWLEDGEMENT

I wish to thank my advisors, J. S. Baras and L. C. Evans, for introducing me to some interesting problems in applied mathematics and for their expert guidance, assistance and support during my stay in graduate school. I would also like to thank Professor J. J. Benedetto for introducing me to the University of Maryland.

A special thanks to Marianne for being by my side.

I gratefully acknowledge the Systems Research Center for supporting this research (National Science Foundation Grant CDR–85–00108, and Air Force Office of Scientific Research Grant AFOSR–87–0073).

TABLE OF CONTENTS

Chapter 1
Introduction

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

$$\begin{aligned} \dot{x} &= f(x,u), \quad x(0) = x_0, \\ \dot{y} &= h(x), \end{aligned} \tag{1.1}$$

where $x \in I\!R^n$ and $y \in I\!R$. Here, the initial condition x_0 is unknown and y represents observations or partial information regarding the state trajectory.

The *observer problem* consists of computing an estimate $\hat{x}(t)$ of $x(t)$, for which the error decays to zero as $t \to \infty$; that is, to design a system

$$\begin{aligned} \dot{m} &= F(m,u,y), \quad m(0) = m_0, \\ \hat{x} &= G(m), \end{aligned} \tag{1.2}$$

where $m \in \mathcal{M}$ and $\hat{x} \in I\!R^n$, such that (for example)

$$| x(t) - \hat{x}(t) | \leq Ce^{-\gamma t}, \quad \gamma > 0. \tag{1.3}$$

The observer state space $\mathcal{M}$ need not necessarily be finite dimensional. One seeks methods for constructing the system (1.2) from the data f, h given for (1.1).

When the control system (1.1) is linear ($f(x,u) = Ax + Bu$, $h(x) = Cx$), the observer problem was solved by Luenberger [44]. To date various ad hoc methods have been applied in the general nonlinear case to certain special cases (eg. linearisation, choosing convenient local coordinates, etc.), and no general methodology has appeared.

Baras and Krishnaprasad [1] have proposed a method employing nonlinear filtering asymptotics based on the informal idea that "nonlinear observer = limit of nonlinear filters". This is a probabilistic approach to a deterministic problem, where one uses additional probabilistic structure in an effort to gain insight into how to solve the problem.

Associate with (1.1) the family of filtering problems

$$\begin{aligned} dx^\epsilon(t) &= f(x^\epsilon(t), u(t))dt + \sqrt{\epsilon} N dw(t), \quad x^\epsilon(0) = x_0^\epsilon, \\ dy^{\epsilon,\delta}(t) &= h(x^\epsilon(t))dt + \sqrt{\delta} R dv(t), \quad y^{\epsilon,\delta}(0) = 0, \end{aligned} \tag{1.4}$$

where w, v are independent standard Wiener processes, x_0^ϵ is a random variable with distribution μ_0^ϵ and density $C_\epsilon \exp(-S_0(x)/\epsilon)$. Since x_0 is unknown, S_0 is chosen so that $\mu_0^\epsilon \Rightarrow \delta_{m_0}$ as $\epsilon \to 0$ for some $m_0 \in I\!R^n$. The matrices N, R are inserted to give extra flexibility in the design (these are called design parameters); we assume that rank$N = n$ and $R > 0$. Then as ϵ, $\delta \to 0$, the trajectories of (1.4) converge in probability to the trajectories of (1.1) corresponding to the initial condition $x(0) = m_0$.

To obtain a nonlinear observer for (1.1), one then studies the limit of the nonlinear filters for (1.4) and identifies the limiting filter (if indeed one exists). Having accomplished this, one then must try to compute the error between the estimate given by the limiting filter and the actual deterministic trajectory as $t \to \infty$. Thus this approach involves two kinds of asymptotics: small parameter and large time. The latter is very difficult in general, and is related to the observability or detectability of the control system (1.1).

Asymptotic nonlinear filtering problems of the type (1.4) have been studied by a number of authors. Picard [51], Bensoussan [4], and Ji [31] have treated the case $\epsilon = 1$ and $\delta \to 0$. Using a suboptimal filter (similar to the extended

Kalman filter) they obtained by various techniques uniform in time estimates of the error. Rather strong assumptions were made regarding the function h, stronger than observability or detectability. The suboptimal filter has a $1/\delta$ singularity, making it difficult to identify any limiting filter. In any case, the limit would not be deterministic.

Hijab [25] treated the case $\epsilon = \delta \to 0$, and obtained a WKB expansion of pathwise unnormalised conditional measures under certain smoothness assumptions. In [26], Hijab derived a large deviation principle for the pathwise conditional measures. His work in part provided the motivation for the asymptotic filtering method for observer design, since it identifies the limilting filter. Mitter [47] also posed this asymptotic filtering problem. This is the case studied in detail in this dissertation. We shall make some comments on the other cases in Section 7.3. We give a new and detailed analysis of this problem using both PDE and probabilistic techniques. A logarithmic limit result for the unnormalised conditional densities is obtained, and a large deviation principle "in probability" is established for the conditional measures. We also study the weak limit of these conditional measures and of the nonlinear filters.

We now indicate how our analysis is carried out. To simplify the discussion, consider now

$$\begin{aligned} \dot{x}(t) &= f(x(t)), \quad x(0) = x_0, \\ \dot{y}(t) &= h(x(t)), \quad y(0) = 0, \end{aligned} \tag{1.5}$$

with the associated noisy system

$$\begin{aligned} dx^\epsilon(t) &= f(x^\epsilon(t))dt + \sqrt{\epsilon}dw(t), \quad x^\epsilon(0) = x_0^\epsilon, \\ dy^\epsilon(t) &= h(x^\epsilon(t))dt + \sqrt{\epsilon}dv(t), \quad y^\epsilon(0) = 0. \end{aligned} \tag{1.6}$$

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

$$
\begin{aligned}
dq^\epsilon &= A_\epsilon^* q^\epsilon dt + \frac{1}{\epsilon} h q^\epsilon dy^\epsilon, \qquad (1.7)\\
q^\epsilon(x,0) &= C_\epsilon \exp(-S_0(x)/\epsilon),
\end{aligned}
$$

where A_ϵ is the generator of the Markov process x_t^ϵ.

Formally applying the logarithmic transformation

$$
W^\epsilon(x,t) = -\epsilon \log q^\epsilon(x,t),
$$

we obtain a nonlinear parabolic PDE for W^ϵ. Sending $\epsilon \to 0$, then $y^\epsilon \to y$ and $W^\epsilon \to W$, where W satisfies (in the viscosity sense) the Hamilton–Jacobi (HJ) equation

$$
\begin{aligned}
\tfrac{\partial}{\partial t} W &+ \max_{u \in R^n} \left\{ DW \cdot (f(x)+u) - \left(\frac{1}{2} \mid u \mid^2 + \frac{1}{2} h(x)^2 - \dot{y}(t) h(x) \right) \right\} = 0,\\
W(x,0) &= S_0(x). \qquad (1.8)
\end{aligned}
$$

Using the vanishing viscosity method developed by Crandall–Lions [6] and Evans–Ishii [16], we obtain

$$
q^\epsilon(x,t) = \exp\left(-\frac{1}{\epsilon} \left[W(x,t) + o(1) \right] \right)
$$

in probability as $\epsilon \to 0$, uniformly on compact subsets of $I\!R^n \times [0,T]$. This is proved rigorously via the "robust" Zakai equation.

The HJ equation (1.8) is parameterised by the observation path $y(\cdot)$, and characterises the deterministic limiting filter since it can be interpreted as the Bellman equation for an optimal control problem arising in Mortensen's deterministic minimum energy estimation. Consider the auxilliary control system

$$
\begin{aligned}
\dot{x} &= f(x) + u, \qquad (1.9)\\
\dot{y} &= h(x) + v,
\end{aligned}
$$

where u, v are deterministic functions modelling noise. Viewing u as a control and fixing y, we have

$$W(x,t) = \tag{1.10}$$
$$\inf_{u(\cdot)} \left\{ S_0(x_u(0)) + \int_0^t \tfrac{1}{2} \mid u(s) \mid^2 + \tfrac{1}{2} h(x_u(s))^2 - \dot{y}(s) h(x_u(s)) ds : x_u(t) = x \right\}$$

Noting that $(h(x)-\dot{y}(t))^2 = h(x)^2 - 2\dot{y}(t)h(x) + \dot{y}(t)^2$, we see that $\inf_{x \in R^n} W(x,t)$ is a measure of the minimum energy required for (1.9) to produce the given output $y(s) : 0 \leq s \leq t$. If $W(\cdot, t)$ has a unique minimiser $\hat{x}(t)$, we call $\hat{x}(t)$ the deterministic estimate given $y(s) : 0 \leq s \leq t$. This defines a deterministic filter $\bar{\pi}_t[y(\cdot)]$. In fact, we show that

$$\bar{\pi}_t^\epsilon[y(\cdot)] \Rightarrow \bar{\pi}_t[y(\cdot)]$$

as $\epsilon \to 0$, where $\bar{\pi}_t^\epsilon$ is the stochastic filter defined pathwise.

Turning now to the second approach, the Kallianpur–Striebel formula gives a representation for the unnormalised conditional measures σ_t^ϵ:

$$\sigma_t^\epsilon(\phi) = E\phi(x_t^\epsilon) \exp\left(-\frac{1}{\epsilon} \left[\frac{1}{2} \int_0^t h(x_s^\epsilon)^2 ds - \int_0^t h(x_s^\epsilon) dy_s^\epsilon \right] \right) \tag{1.11}$$

for all $\phi \in C_b(I\!R^n)$. Using an extension of a theorem of Varadhan, we obtain the large deviation type estimate

$$\sigma_t^\epsilon(A) \asymp \exp\left(-\frac{1}{\epsilon} \inf_{x \in A} W(x,t) \right)$$

in probability as $\epsilon \to 0$, for Borel sets $A \subset I\!R^n$. The symbol "$\asymp$" denotes logarithmic equivalence. This result allows us to prove a large deviation principle (in a generalised sense) for the conditional measures $\pi_t^\epsilon = \sigma_t^\epsilon / \sigma_t^\epsilon(I\!R^n)$, and to show that

$$\pi_t^\epsilon \Rightarrow \delta_{x(t)}$$

in probability as $\epsilon \to 0$, where $x(\cdot)$ is the deterministic trajectory.

We also obtain a large deviation principle for the distributions of the conditional measures π_t^ϵ. This is a second level of large deviation behaviour for our asymptotic filtering problem.

We can explain *formally* why the deterministic estimator arises naturally in the context of large deviations as follows. In the theory of large deviations, asymptotic probabilities are characterised in terms of variational problems. The Wentzell–Freidlin theory [20] considers small random perturbations of dynamical systems (see Section 2.2.2). If P_X^ϵ is the distribution of $X^\epsilon = \{x^\epsilon(s) : 0 \leq s \leq T\}$ on $\Omega^n = C([0,T], I\!R^n)$, then one has

$$P_X^\epsilon(A) \asymp e^{-\frac{1}{\epsilon}\inf_{\theta \in A} I_X(\theta)}$$

for Borel sets $A \subset \Omega^n$, where the action function I_X is given by

$$I_X(\theta) \;=\; S_0(\theta_0) + \frac{1}{2}\int_0^T \mid \dot{\theta}_s - f(\theta_s) \mid^2 ds$$

for absolutely continuous $\theta \in \Omega^n$, and infinity otherwise. Thus, heuristically,

$$P_X^\epsilon(d\theta) \;\asymp\; e^{-\frac{1}{\epsilon}I_X(\theta)} \text{“} d\theta \text{”}.$$

Now let $P_{Y|X}^\epsilon$ denote the conditional distribution of $Y^\epsilon = \{y^\epsilon(s) : 0 \leq s \leq T\}$ given X^ϵ on $\Omega_0 = \{\eta \in C([0,T], I\!R) : \eta(0) = 0\}$. Then we have

$$P_{Y|X}^\epsilon(d\eta \mid \theta) \;\asymp\; e^{-\frac{1}{2\epsilon}\int_0^T |\dot{\eta}_s - h(\theta_s)|^2 ds} \text{“} d\eta \text{”}.$$

We can compute $P_{X|Y}^\epsilon$, the conditional probability of X^ϵ given Y^ϵ on Ω^n using Bayes' rule:

$$\begin{aligned} & P_{X|Y}^\epsilon(A \mid \eta) \\ \asymp\; & \frac{\int_A \exp\left(-\frac{1}{2\epsilon}\int_0^T \mid \dot{\eta}_s - h(\theta_s) \mid^2 ds\right) P_X^\epsilon(d\theta)}{\int_{\Omega^n} \exp\left(-\frac{1}{2\epsilon}\int_0^T \mid \dot{\eta}_s - h(\theta_s) \mid^2 ds\right) P_X^\epsilon(d\theta)} \end{aligned}$$

$$\asymp \frac{\int_A \exp\left(-\frac{1}{\epsilon}\left[S_0(\theta_0)+\int_0^T \frac{1}{2} \mid \dot\theta_s - f(\theta_s) \mid^2 + \frac{1}{2}h(\theta_s)^2 - \dot\eta_s h(\theta_s) ds\right]\right) \text{“}d\theta\text{”}}{\int_{\Omega^n} \exp\left(-\frac{1}{\epsilon}\left[S_0(\theta_0)+\int_0^T \frac{1}{2} \mid \dot\theta_s - f(\theta_s) \mid^2 + \frac{1}{2}h(\theta_s)^2 - \dot\eta_s h(\theta_s) ds\right]\right) \text{“}d\theta\text{”}}.$$

The expression in the exponent is just the quantity being minimised in (1.10), and the above implies that the measures $P^\epsilon_{X|Y}$ concentrate on the optimal trajectory as $\epsilon \longrightarrow 0$.

Finally, we turn to our observer designs. In general, we cannot obtain the deterministic estimate $\hat{x}(t)$ from a finite set of ODEs, so we must regard (1.8) as giving the (infinite dimensional) dynamics of the limiting filter. One notable exception is the Beneš class of systems, which admit finite dimensional filters. An exact infinite dimensional observer is obtained by setting $S_0 \equiv 0$ in (1.8). If we assume that (1.5) is observable on $[0, T]$, then $\hat{x}(t) = x(t)$ for all $t > 0$ and all $x_0 \in I\!R^n$.

We obtain a finite dimensional observer design in the spirit of extended Kalman filtering. We use an approximate deterministic filter which is defined by an ODE and a Riccati equation as the basis for our observer design:

$$\begin{aligned}
\dot{m}(t) &= f(m(t)) + Q(t)H(m(t))'(\dot{y}(t) - h(m(t))), \\
\dot{Q}(t) &= A(m(t))Q(t) + Q(t)A(m(t))' - Q(t)H(m(t))'H(m(t))Q(t) + I
\end{aligned}$$

where $A(x) = Df(x)$, $H(x) = Dh(x)$. This approach was initiated by A. Bensoussan in [2] for the case $h(x) = Cx$.

If the nonlinearities are small, then we expect that $\hat{x}(t) \approx m(t)$. The large time behaviour of the error $x(t) - m(t)$ depends critically on the initial conditions x_0, m_0 and on the growth of $Q(t)$. Under a uniform detectability condition, which is similar to the well known detectability condition for linear systems, one can bound $\parallel Q(t) \parallel$ uniformly. Then if $\mid x_0 - m_0 \mid < \rho$, where ρ depends on the nonlinearities and design parameters, the error decays exponentially to zero. The Lyapunov function $x'Q(t)^{-1}x$ is used to prove this.

Chapter 2
Probabilistic Framework

Some basic results from the theories of nonlinear filtering and large deviations are recalled below. There is now a substantial literature devoted to these subjects, and in this chapter we present briefly the framework required for treating our asymptotic nonlinear filtering problem. In addition, we prove some auxilliary results for subsequent use.

2.1 The Filtering Problem

Let $(\Omega, \mathcal{F}, P)$ be a complete probability space with a filtration $\{\mathcal{F}_t\}$, on which are defined two independent standard Wiener processes $\{w_t, \mathcal{F}_t\}$, $\{v_t, \mathcal{F}_t\}$ taking values in $\mathbb{R}^n$ and $\mathbb{R}$ respectively. Defined also is an $\mathbb{R}^n$ valued random variable x_0^ϵ independent of the Wiener processes with distribution μ_0^ϵ, which is assumed to have a density

$$p_0^\epsilon(x) \;=\; C_\epsilon e^{-\frac{1}{\epsilon}S_0(x)} \tag{2.1}$$

where C_ϵ is a normalisation constant, S_0 is Lipshitz continuous, convex, and $S_0(x_0) = 0,\ S_0(x) > 0$ if $x \neq x_0 \in \mathbb{R}^n$.

We consider a pair of stochastic differential equations (SDE)

$$dx^\epsilon(t) \;=\; f(x^\epsilon(t))dt + \sqrt{\epsilon}dw(t), \quad x^\epsilon(0) = x_0^\epsilon, \tag{2.2}$$

$$dy^\epsilon(t) \;=\; h(x^\epsilon(t))dt + \sqrt{\epsilon}dv(t), \quad y^\epsilon(0) = 0. \tag{2.3}$$

Here we assume $f \in C_b^1(\mathbb{R}^n, \mathbb{R}^n)$ and $h \in C_b^2(\mathbb{R}^n, \mathbb{R})$. Thus there exists a unique solution of (2.2), (2.3) in the strong sense on $(\Omega, \mathcal{F}, P)$. Then $\{x_t^\epsilon\}$ is a Markov process with generator

$$A_\epsilon \;=\; \frac{\epsilon}{2}\Delta + f(x) \cdot D,$$

where Δ is the Laplacian and D denotes gradient in the x variable.

The process

$$X_t^\epsilon = \{x^\epsilon(s) : 0 \leq s \leq t\}$$

is called the *signal* or *state* process, and cannot be observed directly. Rather, we have available the *observation* process

$$Y_t^\epsilon = \{y^\epsilon(s) : 0 \leq s \leq t\},$$

which consists of a function of the state plus noise. The *filtering problem* is concerned with estimating a functional of the state process using the information contained in the observation process. This information is encoded in the filtration $\{\mathcal{Y}_t^\epsilon\}$ defined by

$$\mathcal{Y}_t^\epsilon = \sigma(Y_t^\epsilon),$$

the σ–algebra generated by the observations up to time t.

Let $\phi \in B_b(\mathbb{R}^n)$. The "best" estimate of $\phi(x_t^\epsilon)$ given $\mathcal{Y}_t^\epsilon$ is the conditional expectation

$$\pi_t^\epsilon(\phi) = E[\phi(x_t^\epsilon) \mid \mathcal{Y}_t^\epsilon].$$

We use the notation $\hat{x}_t^\epsilon$ for the conditional mean $E[x_t^\epsilon \mid \mathcal{Y}_t^\epsilon]$. The task of filtering theory is to obtain representations and recursive equations for computing quantities such as $\pi_t^\epsilon(\phi)$. We refer the reader to the books by Elliott [13], Kallianpur [34], and Liptser and Shiryayev [43] for further details.

One can study filtering problems in more general settings. We choose not to do so here, prefering to focus on the central ideas. The methods used in this thesis can be extended if desired.

2.1.1 Kallianpur–Striebel Formula

Fix $T > 0$. For $t \in [0,T]$ define

$$\Omega_t^n = C([0,t], I\!R^n), \quad \Omega_{0,t} = \{\eta \in C([0,t], I\!R) : \eta(0) = 0\},$$

and equip these spaces with the uniform topology. .We shall always use $\|\cdot\|$ to denote the uniform norm. For $t = T$ we write simply $\Omega^n = \Omega_T^n$, etc. This convention is used in other notation below (c.f. Liptser and Shiryayev [43], Chapter 7). $\mathcal{B}(\mathsf{X})$ denotes the Borel σ–algebra of a topological space X.

If (X, d) is a metric space, then $(\mathcal{P}(\mathsf{X}), \varrho)$ is the space of probability measures on $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ where ϱ is the Prohorov metric (Ethier and Kurtz [15], Chapter 3):

$$\varrho(\nu_1, \nu_2) = \inf\left\{\delta > 0 : \nu_1(C) \le \nu_2(C^\delta) + \delta \text{ for all closed } C \subset \mathsf{X}\right\}$$

for $\nu_1,\ \nu_2 \in \mathcal{P}(\mathsf{X})$, where

$$C^\delta = \left\{\mathsf{x} \in \mathsf{X} : \inf_{\mathsf{y} \in C} d(\mathsf{x}, \mathsf{y}) < \delta\right\}.$$

If (X, d) is complete and separable, then so is $(\mathcal{P}(\mathsf{X}), \varrho)$, and convergence in $(\mathcal{P}(\mathsf{X}), \varrho)$ is equivalent to *weak convergence*:

$$\nu_i \Rightarrow \nu \text{ as } i \to \infty$$

if and only if for all $\varphi \in C_b(\mathsf{X})$

$$\lim_{i \to \infty} \int_{\mathsf{X}} \varphi d\nu_i = \int_{\mathsf{X}} \varphi d\nu.$$

The random variables X_t^ϵ, Y_t^ϵ defined above take values in Ω_t^n, $\Omega_{0,t}$ respectively. Define

$$P_X^\epsilon(A) = P(X^\epsilon \in A) \qquad (A \in \mathcal{B}(\Omega^n)).$$

The following theorem gives a functional integral representation for a conditional expectation, and is known as the Kallianpur–Striebel formula. Refer to Kallianpur and Striebel [33], Kallianpur [34] and Liptser and Shiryayev [43].

Theorem (*Kallianpur–Striebel Formula*) *Let* $\Phi \in B_b(\Omega^n, \mathbb{R})$. *Then*

$$E[\Phi(X^\epsilon) \mid \mathcal{Y}^\epsilon](\omega) \;=\; \frac{\int_{\Omega^n} \Phi(\theta)\Lambda^\epsilon(\theta, Y^\epsilon(\omega))P_X^\epsilon(d\theta)}{\int_{\Omega^n} \Lambda^\epsilon(\theta, Y^\epsilon(\omega))P_X^\epsilon(d\theta)} \quad P\text{–a.e. } \omega \in \Omega, \tag{2.4}$$

where

$$\Lambda_t^\epsilon(\theta, \eta) \;=\; \exp\left(-\frac{1}{\epsilon}\left[\frac{1}{2}\int_0^t h(\theta_s)^2 ds - \int_0^t h(\theta_s)d\eta_s\right]\right) \; (\theta \in \Omega^n, \; \eta \in \Omega_0) \tag{2.5}$$

We use this formula to define several random measures as follows. Define

$$\Sigma^\epsilon(A)(\omega) \;=\; \int_A \Lambda^\epsilon(\theta, Y^\epsilon(\omega))P_X^\epsilon(d\theta) \quad (A \in \mathcal{B}(\Omega^n), \; \omega \in \Omega), \tag{2.6}$$

$$\Pi^\epsilon(A)(\omega) \;=\; \frac{\Sigma^\epsilon(A)(\omega)}{\Sigma^\epsilon(\Omega^n)(\omega)}. \tag{2.7}$$

Thus Π^ϵ is a regular version of the conditional probability:

$$P(X^\epsilon \in A \mid \mathcal{Y}^\epsilon)(\omega) \;=\; \Pi^\epsilon(A)(\omega) \quad P\text{–a.s.}, \quad (A \in \mathcal{B}(\Omega^n)).$$

We are also interested in corresponding conditional measures on $\mathbb{R}^n$. Let $t \in [0, T]$ and define

$$\sigma_t^\epsilon(A)(\omega) \;=\; \Sigma^\epsilon\left(\{\theta : \theta_t \in A\}\right)(\omega) \quad (A \in \mathcal{B}(\mathbb{R}^n), \quad \omega \in \Omega), \tag{2.8}$$

$$\pi_t^\epsilon(A)(\omega) \;=\; \frac{\sigma_t^\epsilon(A)(\omega)}{\sigma_t^\epsilon(\mathbb{R}^n)(\omega)}. \tag{2.9}$$

So π^ϵ is a regular version of the conditional probability:

$$P\left(x_t^\epsilon \in A \mid \mathcal{Y}_t^\epsilon\right)(\omega) = \pi_t^\epsilon(A)(\omega) \quad P\text{–a.s.} \;\; (A \in \mathcal{B}(\mathbb{R}^n)),$$

and if $\phi \in B_b(I\!R^n)$ we have

$$\begin{aligned}\pi_t^\epsilon(\phi)(\omega) &= E_P[\phi(x_t^\epsilon) \mid \mathcal{Y}_t^\epsilon](\omega) \\ &= \frac{\int_{\Omega^n} \phi(\theta_t)\Lambda_t^\epsilon(\theta, Y^\epsilon(\omega))P_X^\epsilon(d\theta)}{\int_{\Omega^n} \Lambda_t^\epsilon(\theta, Y^\epsilon(\omega))P_X^\epsilon(d\theta)} \qquad P\text{–a.s.}\end{aligned} \tag{2.10}$$

The measures (Σ^ϵ, σ_t^ϵ) Π^ϵ, π_t^ϵ are (unnormalised) conditional measures.

π^ϵ is a *random measure:*

$$\begin{aligned}\pi_t^\epsilon : \Omega &\rightarrow \mathcal{P}(I\!R^n) \\ \omega &\mapsto \pi_t^\epsilon(\omega).\end{aligned} \tag{2.11}$$

Pathwise or *robust* analogues of these measures, which we next define, are useful from a practical point of view, as well as for asymptotic analysis (Hijab [25], [26], and Chapter 5 herein). Refer to Davis [11], Sussmann [53], and others.

Using the integration by parts formula and Girsanov's theorem, we proceed as follows:

$$\begin{aligned}&\int_{\Omega^n} \Phi(\theta)\Lambda^\epsilon(\theta, Y^\epsilon(\omega))P_X^\epsilon(d\theta) = \\ &\int_\Omega \Phi(X^\epsilon(\tilde{\omega})) \exp\left(-\frac{1}{\epsilon}\left[\frac{1}{2}\int_0^T h(x_s^\epsilon(\tilde{\omega}))^2 ds - \int_0^T h(x_s^\epsilon(\tilde{\omega}))dy_s^\epsilon(\omega)\right]\right) P(d\tilde{\omega}) \\ = &\int_\Omega \Phi(X^\epsilon(\tilde{\omega})) \exp\left(-\frac{1}{\epsilon}\left[-y_T^\epsilon(\omega)h(x_T^\epsilon(\tilde{\omega})) + \int_0^T V^\epsilon(x_s^\epsilon(\tilde{\omega}), y_s^\epsilon(\omega))ds\right]\right) \\ &\qquad \Gamma_T^\epsilon(X^\epsilon(\tilde{\omega}), Y^\epsilon(\omega), \tilde{\omega})P(d\tilde{\omega}) \qquad P\text{–a.s.} \\ = &\int_{\Omega^n} \Phi(\theta)\bar{\Lambda}^\epsilon(\theta, Y^\epsilon(\omega))\check{P}_{[Y^\epsilon(\omega)]}^\epsilon(d\theta),\end{aligned} \tag{2.12}$$

where

$$V^\epsilon(x,y) = \frac{1}{2}h(x)^2 + yA_\epsilon h(x) - \frac{1}{2}y^2 \mid Dh(x) \mid^2 \quad (x \in I\!R^n,\ y \in I\!R), \tag{2.13}$$

$$\begin{aligned}\Gamma_t^\epsilon(\theta, \eta, \tilde{\omega}) &= \exp\left[-\frac{1}{\sqrt{\epsilon}}\int_0^t \eta_s Dh(\theta_s)dw_s(\tilde{\omega}) - \frac{1}{2\epsilon}\int_0^t \eta_s^2 \mid Dh(\theta_s) \mid^2 ds\right] \\ &\qquad (\theta \in \Omega^n,\ \eta \in \Omega_0,\ \tilde{\omega} \in \Omega),\end{aligned} \tag{2.14}$$

$$\bar{\Lambda}^\epsilon_t(\theta,\eta) = \exp\left(-\frac{1}{\epsilon}\left[-\eta_t h(\theta_t) + \int_0^t V^\epsilon(\theta_s,\eta_s)ds\right]\right) \quad (\theta \in \Omega^n,\ \eta \in \Omega_0). \quad (2.15)$$

Here $\check{P}^\epsilon_{[\eta]}$ is the distribution on Ω^n of the diffusion

$$d\check{x}^\epsilon(t) = g(\check{x}^\epsilon(t),\eta(t))dt + \sqrt{\epsilon}dw(t), \quad \check{x}^\epsilon(0) = x_0^\epsilon,$$

where

$$g(x,y) = f(x) - yDh(x)' \quad (x \in I\!R^n, y \in I\!R). \tag{2.16}$$

Notice that $\bar{\Lambda}^\epsilon$ does not involve stochastic integration, and thus is well defined for all $\eta \in \Omega_0$, and not just on a set of ϵ–Wiener measure one. Further, $\bar{\Lambda}^\epsilon$ depends continuously on $\eta \in \Omega_0$. These properties are inherited by the measures we next define:

$$\bar{\Sigma}^\epsilon(A)[\eta] = \int_A \bar{\Lambda}^\epsilon(\theta,\eta)\check{P}^\epsilon_{[\eta]}(d\theta) \quad (A \in \mathcal{B}(\Omega^n),\ \eta \in \Omega_0), \tag{2.17}$$

$$\bar{\Pi}^\epsilon(A)[\eta] = \frac{\bar{\Sigma}^\epsilon(A)[\eta]}{\bar{\Sigma}^\epsilon(\Omega^n)[\eta]}, \tag{2.18}$$

and also

$$\bar{\sigma}^\epsilon_t(A)[\eta] = \bar{\Sigma}^\epsilon(\{\theta : \theta_t \in A\})[\eta] \quad (A \in \mathcal{B}(I\!R^n),\ \eta \in \Omega_0), \tag{2.19}$$

$$\bar{\pi}^\epsilon_t(A)[\eta] = \frac{\bar{\sigma}^\epsilon_t(A)[\eta]}{\bar{\sigma}^\epsilon_t(I\!R^n)[\eta]}. \tag{2.20}$$

In contrast to (2.11), this can be interpreted as a *pathwise filter* $\bar{\pi}^\epsilon$:

$$\begin{aligned} \bar{\pi}^\epsilon_t : \Omega_{0,t} &\to \mathcal{P}(I\!R^n) \\ \eta &\mapsto \bar{\pi}^\epsilon_t[\eta] \end{aligned} \tag{2.21}$$

These measures also define versions of the above conditional probabilities via

$$\Sigma(A)^\epsilon(\omega) = \bar{\Sigma}^\epsilon(A)[Y^\epsilon(\omega)] \quad P\text{–a.s. } (A \in \mathcal{B}(\Omega^n),\ \omega \in \Omega), \tag{2.22}$$

and in particular we have

$$\pi^\epsilon_t(\phi)(\omega) = \frac{\int_{\Omega^n} \phi(\theta_t)\bar{\Lambda}^\epsilon(\theta, Y^\epsilon(\omega))\check{P}^\epsilon_{[Y^\epsilon(\omega)]}(d\theta)}{\int_{\Omega^n} \bar{\Lambda}^\epsilon(\theta, Y^\epsilon(\omega))\check{P}^\epsilon_{[Y^\epsilon(\omega)]}(d\theta)} \quad P\text{–a.s.} \tag{2.23}$$

We now prove that these robust measures are continuous functions of $\eta \in \Omega_0$. This has been established by several authors; for example Kushner [41], Sussmann [53].

Proposition 2.1 *The map $\bar{\pi}_t^\epsilon$ defined by* (2.20), (2.21) *is locally Lipshitz continuous.*

Proof: 1. Let $K \subset \Omega_0$ be compact, and $\eta, \zeta \in K$. We will show that there exists a constant $C_{K,\epsilon} > 0$ such that

$$\varrho(\bar{\pi}_t^\epsilon[\eta], \bar{\pi}_t^\epsilon[\zeta]) \;\leq\; C_{K,\epsilon} \parallel \eta - \zeta \parallel . \tag{2.24}$$

2. For any $M > 0$, there exists $C > 0$ such that

$$J(M) = \sup_{\|\eta\| \leq M} E \exp\left(-\frac{1}{\sqrt{\epsilon}} \int_0^t \eta_s Dh(x_s^\epsilon) dw_s \right) \leq e^{\frac{C}{\epsilon}}. \tag{2.25}$$

We have

$$\int_0^t \eta_s^2 \mid Dh(x_s^\epsilon) \mid^2 ds \;\leq\; \int_0^T \parallel \eta \parallel^2 \parallel Dh \parallel^2 ds \leq C.$$

Then by Lemma 7.1.2 of Kallianpur [34],

$$\begin{aligned} 1 \;&=\; E \exp\left(-\frac{1}{\epsilon} \left[\sqrt{\epsilon} \int_0^t \eta_s Dh(x_s^\epsilon) dw_s - \frac{1}{2} \int_0^t \eta_s^2 \mid Dh(x_s^\epsilon) \mid^2 ds \right] \right) \\ &\geq\; e^{-\frac{C}{\epsilon}} E \exp\left(-\frac{1}{\sqrt{\epsilon}} \int_0^t \eta_s Dh(x_s^\epsilon) dw_s \right), \end{aligned}$$

which implies (2.25).

3. Let $F \subset \mathbb{R}^n$ be closed, and set $M = \sup_{\eta \in K} \parallel \eta \parallel$. Following the method of Sussmann [53], we shall prove that there exists $C > 0$ depending on K, ϵ but not on F such that

$$\mid \bar{\sigma}_t^\epsilon(F)[\eta] - \bar{\sigma}_t^\epsilon(F)[\zeta] \mid \leq \; C \parallel \eta - \zeta \parallel .$$

Note that we can write

$$\bar{\sigma}_t^\epsilon(F)[\eta] = EI_{x_t^\epsilon \in F} \exp\left(\frac{1}{\epsilon}G(X^\epsilon, \eta)\right)$$

where

$$G(X^\epsilon, \eta) = \eta_t h(x_t^\epsilon) - \int_0^t \frac{1}{2}h(x_s^\epsilon)^2 + \eta_s A_\epsilon h(x_s^\epsilon) ds - \sqrt{\epsilon}\int_0^t \eta_s Dh(x_s^\epsilon) dw_s.$$

Then

$$\begin{aligned} &\mid \bar{\sigma}_t^\epsilon(F)[\eta] - \bar{\sigma}_t^\epsilon(F)[\zeta] \mid \leq E\left(e^{\frac{1}{\epsilon}G(X^\epsilon,\eta)} - e^{\frac{1}{\epsilon}G(X^\epsilon,\zeta)}\right) \\ &= E\textstyle\int_0^1 \exp\left(\frac{1}{\epsilon}G(X^\epsilon, \lambda\eta) + \frac{1}{\epsilon}G(X^\epsilon, (1-\lambda)\zeta)\right) d\lambda \frac{1}{\epsilon}\left(G(X^\epsilon, \eta) - G(X^\epsilon, \zeta)\right). \end{aligned}$$

Define

$$F(X^\epsilon, \eta) \; = \; \eta_t h(x_t^\epsilon) - \int_0^t \eta_s A_\epsilon h(x_s^\epsilon) ds - \sqrt{\epsilon}\int_0^t \eta_s Dh(x_s^\epsilon) dw_s.$$

Note that for $\gamma > 0$,

$$\mid x \mid \leq \; \gamma \cosh(\frac{x}{\gamma}).$$

Using this inequality with $\gamma = \parallel \eta - \zeta \parallel$, we get

$$\begin{aligned} \mid \bar{\sigma}_t^\epsilon(F)[\eta] - \bar{\sigma}_t^\epsilon(F)[\zeta] \mid \; &\leq \; \frac{\parallel \eta - \zeta \parallel}{2} E\int_0^1 e^{-\frac{1}{\epsilon}\int_0^t h(x_s^\epsilon)^2 ds} \\ &\left(\exp\left[\frac{1}{\epsilon}F(X^\epsilon, \lambda\eta + (1-\lambda)\zeta) + \frac{\eta - \zeta}{\parallel \eta - \zeta \parallel}\right]\right. \\ &\left. + \exp\left[\frac{1}{\epsilon}F(X^\epsilon, \lambda\eta + (1-\lambda)\zeta) - \frac{\eta - \zeta}{\parallel \eta - \zeta \parallel}\right]\right) d\lambda \\ &\leq \; \parallel \eta - \zeta \parallel e^{\frac{C}{\epsilon}} J(M+1), \end{aligned}$$

as required.

4. Next we have

$$\begin{aligned} \mid \bar{\pi}_t^\epsilon(F)[\eta] - \bar{\pi}_t^\epsilon(F)[\zeta] \mid \; &\leq \; \frac{1}{\bar{\sigma}_t^\epsilon(\mathbb{R}^n)[\eta]\bar{\sigma}_t^\epsilon(\mathbb{R}^n)[\zeta]} \\ &\left(\bar{\sigma}_t^\epsilon(F)[\eta]\bar{\sigma}_t^\epsilon(\mathbb{R}^n)[\zeta] - \bar{\sigma}_t^\epsilon(F)[\zeta]\bar{\sigma}_t^\epsilon(\mathbb{R}^n)[\eta]\right) \\ &\leq \; C_{K,\epsilon} \parallel \eta - \zeta \parallel. \end{aligned}$$

5. Let $\delta = C_{K,\epsilon} \parallel \eta - \zeta \parallel$. Then

$$\begin{aligned} \bar{\pi}_t^\epsilon(F)[\eta] &\leq \bar{\pi}_t^\epsilon(F)[\zeta] + \mid \bar{\pi}_t^\epsilon(F)[\eta] - \bar{\pi}_t^\epsilon(F)[\zeta] \mid \\ &\leq \bar{\pi}_t^\epsilon(F^\delta)[\zeta] + \delta, \end{aligned}$$

which implies (2.24). ∎

2.1.2 Zakai Equation

Consider the unnormalised conditional measures σ_t^ϵ defined by (2.8). Define a new measure P_0 on $(\Omega, \mathcal{F})$ by

$$\frac{dP_0}{dP} = \Lambda_T^\epsilon(X^\epsilon, Y^\epsilon)^{-1}.$$

Then, under P_0,

$$\tilde{y}^\epsilon(t) = \frac{1}{\sqrt{\epsilon}} y^\epsilon(t)$$

is a standard Wiener process with respect to the filtration $\{\mathcal{Y}_t^\epsilon\}$.

Under our assumptions, σ_t^ϵ has a density:

$$\sigma_t^\epsilon(A) = \int_A q^\epsilon(x,t)dx \quad (A \in \mathcal{B}(\mathbb{R}^n)).$$

The next theorem gives (linear) stochastic partial differential equations satisfied by σ_t^ϵ and q^ϵ. These are the *Zakai* equations of nonlinear filtering, in weak and strong form. See, for example, Elliott [13], Liptser and Shiryayev [43], Kallianpur [34], etc.

Theorem *(Zakai Equation) Let $\phi \in C_c^2(\mathbb{R}^n)$. We have*

$$\sigma_t^\epsilon(\phi) = \mu_0^\epsilon(\phi) + \int_0^t \sigma_s^\epsilon(A_\epsilon\phi)ds + \frac{1}{\epsilon}\int_0^t \sigma_s^\epsilon(\phi h)dy_s^\epsilon \qquad P\text{–a.s.}, \tag{2.26}$$

and

$$\begin{aligned} dq^\epsilon(x,t) &= A_\epsilon^* q^\epsilon(x,t)dt + \frac{1}{\epsilon}h(x)q^\epsilon(x,t)dy^\epsilon(t) \qquad P\text{–a.s.}, \qquad (2.27) \\ q^\epsilon(x,0) &= p_0^\epsilon(x). \end{aligned}$$

The Zakai equation can be regarded as part of an *infinite dimensional realisation* of the stochastic filter π^ϵ (2.11), as discussed in Hijab [25].

From Section 2.1.1, we have a Feynman–Kac type representation for the solution of the SPDE (2.26):

$$\sigma_t^\epsilon(\phi) = \int_{\Omega^n} \phi(\theta_t)\Lambda_t^\epsilon(\theta, Y^\epsilon)P_X^\epsilon(d\theta) \qquad P\text{–a.s.} \qquad (2.28)$$

A similar representation can be obtained for the solution of (2.27) using the robust Zakai equation derived below. As noted in the previous section, such pathwise versions will be useful in the sequel.

Define

$$p^\epsilon(x,t)(\omega) = \exp\left(-\frac{1}{\epsilon}y_t^\epsilon(\omega)h(x)\right) q^\epsilon(x,t)(\omega). \qquad (2.29)$$

Using Itô's rule, p^ϵ satisfies

$$\begin{aligned} &\tfrac{\partial}{\partial t}p^\epsilon(x,t)(\omega) - \frac{\epsilon}{2}\Delta p^\epsilon(x,t)(\omega) + g(x, y_t^\epsilon(\omega)) \cdot Dp^\epsilon(x,t)(\omega) \qquad (2.30) \\ &+ \frac{1}{\epsilon}\left(V^\epsilon(x, y_t^\epsilon(\omega)) + \epsilon \operatorname{div} g(x, y_t^\epsilon(\omega))\right) p^\epsilon(x,t) = 0, \\ &p^\epsilon(x,0)(\omega) = p_0^\epsilon(x), \qquad P\text{–a.s.} \end{aligned}$$

and note that this equation does not involve stochastic integration. It is a PDE with random coefficients since it depends on $\omega \in \Omega$ via Y^ϵ. Thus fix $\eta \in \Omega_0$ and denote by $p^\epsilon(x,t) = p^\epsilon(x,t)[\eta] \in C_b^{1,2}(I\!R^n \times [0,T])$ the unique positive solution of

$$\frac{\partial}{\partial t}p^\epsilon(x,t)-\frac{\epsilon}{2}\Delta p^\epsilon(x,t)+g(x,\eta_t)\cdot Dp^\epsilon(x,t)+\frac{1}{\epsilon}\left(V^\epsilon(x,\eta_t)+\epsilon\mathrm{div}g(x,\eta_t)\right)p^\epsilon(x,t)=0, \tag{2.31}$$

$$p^\epsilon(x,0) \;=\; p_0^\epsilon(x),$$

We refer to (2.31) as the *robust* or *pathwise* Zakai equation, and it forms part of an infinite dimensional realisation of the pathwise filter $\bar{\pi}^\epsilon$ (2.21). This defines a version of the solutions to (2.27), (2.30) via

$$p^\epsilon(x,t)(\omega) \;=\; p^\epsilon(x,t)[Y^\epsilon(\omega)] \quad P\text{–a.s.} \tag{2.32}$$

Let $\breve{P}_{x,t,[\eta]}$ denote the distribution on Ω_t^n of the diffusion:

$$\begin{aligned} d\breve{x}^\epsilon(s) &= -g(\breve{x}^\epsilon(s),\eta(t-s))ds+\sqrt{\epsilon}dw(s),\;\; 0\le s\le t, \\ \breve{x}^\epsilon(0) &= x \end{aligned}$$

Then using the Feynman–Kac formula (see Freidlin [21], Theorem 2.2, page 132), we have

$$p^\epsilon(x,t)[\eta] = \tag{2.33}$$

$$\int_{\Omega_t^n} p_0^\epsilon(\theta_t)\exp\left(-\frac{1}{\epsilon}\left[\int_0^t\left(V^\epsilon(\theta_s,\eta(t-s))+\epsilon\mathrm{div}g(\theta_s,\eta(t-s))\right)ds\right]\right)\breve{P}_{x,t,[\eta]}(d\theta).$$

This together with (2.32), (2.29) yields a functional integral representation for q^ϵ P–a.s. Note also:

$$\pi_t^\epsilon(\phi)(\omega) \;=\; \frac{\int_{R^n}\phi(x)e^{\frac{1}{\epsilon}y_t^\epsilon(\omega)h(x)}p^\epsilon(x,t)[Y^\epsilon(\omega)]dx}{\int_{R^n}e^{\frac{1}{\epsilon}y_t^\epsilon(\omega)h(x)}p^\epsilon(x,t)[Y^\epsilon(\omega)]dx} \quad P\text{–a.s.} \tag{2.34}$$

We now give a bound and continuity result for $p^\epsilon[\eta]$.

Proposition 2.2 *Let $K\subset\Omega^n$ be compact. Then there exists a positive constant C such that*

$$p^\epsilon(x,t)[\eta] \;\le\; e^{-\frac{1}{\epsilon}(C|x|-C)}$$

for all $\eta \in K$ *and all* $(x,t) \in I\!R^n \times [0,T]$, *and*

$$\sup_{(x,t)\in I\!R^n\times[0,T]} \mid p^\epsilon(x,t)[\eta] - p^\epsilon(x,t)[\zeta] \mid \leq C \parallel \eta - \zeta \parallel$$

for all $\eta,\ \zeta \in K$.

Proof: To obtain the bound, use the representation formula (2.33) and reduce to a gaussian integral with Girsanov's Theorem. Then bound this integral.

The proof of the continuity result is similar to that of Proposition 2.1 and will be omitted. ∎

2.2 Large Deviations

Let (X, d) be a complete separable metric space, equipped with the Borel $\sigma-$algebra $\mathcal{B}(\mathsf{X})$. For $\epsilon > 0$ let $\{\mathsf{P}^\epsilon\}$ be a family of probability measures on X with $\mathsf{P}^\epsilon \Rightarrow \delta_{\mathsf{x}_0}$ as $\epsilon \to 0$, for some $\mathsf{x}_0 \in \mathsf{X}$ (weak convergence of probability measures). Large deviations is concerned with estimating the rate at which $\mathsf{P}^\epsilon(A) \to 0$ as $\epsilon \to 0$ for events A with $\mathsf{x}_0 \notin A$; that is, to study asymptotic probabilities of "rare" events. These asymptotics are characterised by a deterministic variational problem: minimise an *action* or *rate* function over the set A.

A general formulation for studying such problems was given by Varadhan [54], [55], and the measures $\{\mathsf{P}^\epsilon\}$ are then said to satify a *large deviation principle* (LDP). Freidlin and Wentzell [56], [20] have studied extensively problems concerned with small random perturbations of dynamical systems. There are close links to PDE problems with small parameters; see for example Varadhan [55], Freidlin and Wentzell [20], Freidlin [21], and the references contained therein.

A large deviation principle for a family $\{\mathsf{P}^\epsilon\}$ can be obtained using purely probabilistic methods (for example, Ellis [14], who treats statistical mechanics problems using large deviations). Sometimes the problem can be expressed as an equivalent PDE problem, which may be solved using probabilistic methods via a representation formula (for example, Freidlin [21]), or stochastic control techniques (Fleming [17]), or by using direct analytic (PDE) arguements (for example, Evans and Ishii [16]).

Our interest is in the small parameter asymptotics of the filtering problem, that is, in large deviation results for conditional measures and related asymptotics of the filtering equations. Below we define a LDP for families of *random* measures.

2.2.1 Varadhan's Formulation

Let X be defined as above, and let $\{\mathsf{P}^\epsilon\}$ be a family of probability measures on X.

Definition (*Varadhan [54]*) *We say that* $\{\mathsf{P}^\epsilon\}$ *obeys the* <u>*large deviation principle*</u> *(LDP) with* <u>*action function*</u> $I : \mathsf{X} \to [0, \infty]$ *provided:*

(i) *I is lower semicontinuous;*

(ii) *For each $M > 0$, the set $\{\mathsf{x} \in \mathsf{X} : I(\mathsf{x}) \leq M\}$ is compact;*

(iii) *For any closed set $C \subset \mathsf{X}$,*

$$\limsup_{\epsilon \to 0} \epsilon \log \mathsf{P}^\epsilon(C) \leq -\inf_{\mathsf{x} \in C} I(\mathsf{x});$$

(iv) *For any open set $G \subset \mathsf{X}$,*

$$\liminf_{\epsilon \to 0} \epsilon \log \mathsf{P}^\epsilon(G) \geq -\inf_{\mathsf{x} \in G} I(\mathsf{x}).$$

Typically, $I(\mathsf{x}_0) = 0$ and $I(\mathsf{x}) > 0$ if $\mathsf{x} \neq \mathsf{x}_0$.

EXAMPLE Let $\mathsf{X} = I\!R^n$, and consider the random variables x_0^ϵ with distribution μ_0^ϵ and density p_0^ϵ defined by (2.1). Then for regular $A \in \mathcal{B}(I\!R^n)$ one has

$$\lim_{\epsilon \to 0} \mu_0^\epsilon(A) = -\inf_{x \in A} S_0(x).$$

So if $x_0 \notin A$ the decay rate is given by

$$\mu_0^\epsilon(A) \asymp e^{-\frac{1}{\epsilon}\inf_{x \in A} S_0(x)},$$

as $\epsilon \to 0$. The symbol "$\asymp$" denotes logarithmic equivalence in the sense of (iii) and (iv) of the above definition. In terms of densities,

$$\lim_{\epsilon \to 0} \epsilon \log p_0^\epsilon(x) = -S_0(x).$$

Here the action function is $I(x) = S_0(x)$. ///

There are two important theorems for obtaining new large deviation results from old ones. Each has a number of variants.

Theorem (*Contraction Principle, Varadhan [54]*) *Let* $\{\mathsf{P}^\epsilon\}$ *obey the LDP with action function* I. *Let* $F_\epsilon : \mathsf{X} \to \mathsf{Y}$ *be continuous, where* Y *is another complete separable metric space. Assume* $\lim_{\epsilon \to 0} F_\epsilon = F$ *exists uniformly on compact subsets of* X. *Define* $\{\mathsf{Q}^\epsilon\}$ *on* Y *by* $\mathsf{Q}^\epsilon = \mathsf{P}^\epsilon F_\epsilon^{-1}$. *Then* $\{\mathsf{Q}^\epsilon\}$ *obey the LDP with action function*

$$J(\mathsf{y}) = \inf_{\mathsf{x}: F(\mathsf{x}) = \mathsf{y}} I(\mathsf{x}).$$

If y *is not in the range of* F, *then set* $J(\mathsf{y}) = +\infty$.

Theorem (*Varadhan's Theorem [54], [55]*) *Let* $\{\mathsf{P}^\epsilon\}$ *obey the LDP with action function* I. *Let* F^ϵ, $F : \mathsf{X} \to I\!R$ *be bounded continuous functions such that* $F^\epsilon \to F$ *uniformly as* $\epsilon \to 0$. *Then for any closed set* $C \subset \mathsf{X}$ *and any open set*

$G \subset \mathsf{X}$, *we have:*

$$\limsup_{\epsilon \to 0} \epsilon \log \int_C \exp\left[-\frac{1}{\epsilon} F(\mathsf{x})\right] \mathsf{P}^\epsilon(d\mathsf{x}) \;\le\; -\inf_{\mathsf{x}\in C} [I(\mathsf{x}) + F(\mathsf{x})],$$

$$\liminf_{\epsilon \to 0} \epsilon \log \int_G \exp\left[-\frac{1}{\epsilon} F(\mathsf{x})\right] \mathsf{P}^\epsilon(d\mathsf{x}) \;\ge\; -\inf_{\mathsf{x}\in G} [I(\mathsf{x}) + F(\mathsf{x})].$$

The second result is an extension to function space of Laplace's asymptotic method. The next result is a version of Laplace's method (c.f. Freidlin and Wentzell [20], Chapter 3).

Theorem 2.1 *(Laplace's Asymptotic Method) Let $F : \mathbb{R}^n \to \bar{\mathbb{R}}$ be Borel measurable, bounded below. Assume that for each $\epsilon > 0$,*

$$a(\epsilon) \;=\; \int_{R^n} e^{-\frac{1}{\epsilon}F(x)} dx < \infty.$$

(i) If

$$\limsup_{\epsilon \to 0} \epsilon \log a(\epsilon) \;\le\; M$$

for some constant $M < \infty$, then for any closed subset C of $\mathbb{R}^n$, we have

$$\limsup_{\epsilon \to 0} \epsilon \log \int_C e^{-\frac{1}{\epsilon}F(x)} dx \;\le\; -\inf_{x\in C} F(x).$$

(ii) *If F is upper semicontinuous, then for any open subset G of $\mathbb{R}^n$, we have*

$$\liminf_{\epsilon \to 0} \epsilon \log \int_G e^{-\frac{1}{\epsilon}F(x)} dx \;\ge\; -\inf_{x\in G} F(x).$$

Proof: *Upper bound.* Write $m = \inf_{x\in C} F(x)$ and assume $m < \infty$. Let $0 < \lambda < 1$. Then

$$\int_C e^{-\frac{1}{\epsilon}F(x)} dx \;=\; \int_C e^{-\frac{\lambda}{\epsilon}F(x)} e^{-\frac{(1-\lambda)}{\epsilon}F(x)} dx$$

$$\begin{aligned} &\le e^{-\frac{(1-\lambda)}{\epsilon}m}\int_C e^{-\frac{\lambda}{\epsilon}F(x)}dx \\ &\le e^{-\frac{(1-\lambda)}{\epsilon}m}a(\frac{\epsilon}{\lambda}), \end{aligned}$$

and so

$$\epsilon\log\int_C e^{-\frac{1}{\epsilon}F(x)}dx \;\le\; -(1-\lambda)m+\lambda(\frac{\epsilon}{\lambda})\log a(\frac{\epsilon}{\lambda}).$$

Therefore

$$\limsup_{\epsilon\to 0}\epsilon\log\int_C e^{-\frac{1}{\epsilon}F(x)}dx \;\le\; -(1-\lambda)m+\lambda M.$$

Let $\lambda\to 0$ to obtain the required inequality.

Lower bound. Now write $m=\inf_{x\in G}F(x)$, and assume $m<\infty$. For any $\delta>0$ define

$$G_\delta \;=\; \{x\in G : F(x)<m+\delta,\ |\,x\,|<R\}$$

where $R>0$ is chosen large enough to ensure $G_\delta\neq\emptyset$. Then G_δ is a bounded open subset of G, and

$$\begin{aligned} \int_G e^{-\frac{1}{\epsilon}F(x)}dx &\ge \int_{G_\delta} e^{-\frac{1}{\epsilon}(m+\delta)}dx \\ &\ge L^n(G_\delta)e^{-\frac{1}{\epsilon}(m+\delta)}, \end{aligned}$$

and hence

$$\liminf_{\epsilon\to 0}\epsilon\log\int_G e^{-\frac{1}{\epsilon}F(x)}dx \;\ge\; -(m+\delta).$$

This holds for all $\delta>0$, hence the lower bound follows. ∎

2.2.2 Wentzell–Freidlin Theory

The Wentzell–Freidlin theory is concerned with small random perturbations of the dynamical system

$$\dot{x}(s) \;=\; f(x(s)),\quad x(0)=x_0, \tag{2.35}$$

for $s \in [0,T]$, given by, for example, the SDE (2.2) with small parameter $\epsilon > 0$. The parameter ϵ is a measure of the intensity of the noise, and as $\epsilon \to 0$, the trajectories of (2.2) are "close" to the trajectory of (2.35) with high probability. However, there is a positive (exponentially small) probability that the trajectories of (2.2) will be far from the trajectory of (2.35). The behaviour of such rare events is captured by a LDP.

We assume that x_0^ϵ has density given by (2.1). Noting that $P(x_t^\epsilon \in A) = \int_{R^n} P(x_t^\epsilon \in A \mid x_0^\epsilon = x)\mu_0^\epsilon(dx)$, the following slight extension of Theorem 1.2, page 45, of Freidlin and Wentzell [20] is easily obtained.

Lemma 2.1 *Under the above assumptions we have:*

(i) *there is a constant $C > 0$ depending on $T > 0$ such that*

$$\sup_{0 \le t \le T} E \mid x_t^\epsilon - x_t \mid^2 \le \epsilon C;$$

(ii) *for all $t > 0$ and $\delta > 0$,*

$$\lim_{\epsilon \to 0} P(\{\max_{0 \le s \le t} \mid x^\epsilon(s) - x(s) \mid > \delta\}) = 0.$$

Assume temporarily that $x_0^\epsilon = x$ *a.s.* Let P_x^ϵ denote the distribution of X^ϵ on Ω^n. Fix $T > 0$ and consider the control system

$$\begin{aligned} \dot{x}_u(t) &= f(x_u(t)) + u(t), \quad 0 \le t \le T, \\ x(0) &= x, \end{aligned} \tag{2.36}$$

where $u : [0,T] \to I\!R^n$ is measurable. For $\theta \in \Omega^n$ define

$$I_x(\theta) = \inf_u \{\frac{1}{2}\int_0^T \mid u(t) \mid^2 dt \ : \ x_u = \theta, \ x_u(0) = x\} \tag{2.37}$$

$$= \begin{cases} \frac{1}{2}\int_0^T \mid \dot{\theta}_t - f(\theta_t) \mid^2 dt & \text{if } \theta \text{ is absolutely continuous} \\ & \text{and } \theta(0) = x, \\ \infty & \text{otherwise} \end{cases}$$

Theorem *(Freidlin and Wentzell [20], Theorem 1.1, page 104) $\{P_x^\epsilon\}$ obey the LDP with action function $I_x(\theta)$.*

Now suppose again that x_0^ϵ has distribution μ_0^ϵ, and recall that P_X^ϵ is the distribution of X^ϵ on Ω^n:

$$P_X^\epsilon(A) \;=\; \int_{R^n} P_x^\epsilon(A) C_\epsilon e^{-\frac{1}{\epsilon}S_0(x)} dx.$$

Define for $\theta \in \Omega^n$:

$$I_X(\theta) = \inf_u \left\{ S_0(x_u(0)) + \frac{1}{2}\int_0^T \mid u(t) \mid^2 dt \;:\; x_u = \theta \right\}. \tag{2.38}$$

Theorem 2.2 *$\{P_X^\epsilon\}$ obey the LDP with action function $I_X(\theta)$.*

Proof: *Upper bound.* For any $\delta > 0$, from the previous theorem we have, for $C \subset \Omega^n$ closed,

$$P_x^\epsilon(C) \;\leq\; \exp\left(-\frac{1}{\epsilon}\left[\inf_{\theta \in C} I_x(\theta) - \delta\right]\right),$$

for $\epsilon > 0$ sufficiently small. Then

$$P^\epsilon(C) \;\leq\; C_\epsilon \int_{R^n} \exp\left(-\frac{1}{\epsilon}\left[S_0(x) + \inf_{\theta \in C} I_x(\theta) - \delta\right]\right) dx$$

and hence

$$\begin{aligned} \limsup_{\epsilon \to 0} \epsilon \log P^\epsilon(C) \;&\leq\; -\inf_{x \in R^n}\left[S_0(x) + \inf_{\theta \in C} I_x(\theta)\right] + \delta \\ &=\; -I_X(\theta) + \delta, \end{aligned}$$

using Theorem 2.1. Since $\delta > 0$ was arbitrary, the upper bound follows.

Lower bound. Note that if $G \subset \Omega^n$ is open, then

$$x \mapsto \inf_{\theta \in G} I_x(\theta)$$

is upper semicontinuous. The lower bound follows using

$$P_x^\epsilon(G) \ \geq \ \exp\left(-\frac{1}{\epsilon}\left[\inf_{\theta \in G} I_x(\theta) + \delta\right]\right)$$

and Theorem 2.1. ∎

Let $P^\epsilon_{X,Y}$, P^ϵ_Y denote the distributions on $\Omega^n \times \Omega_0$, Ω_0 of (X^ϵ, Y^ϵ), Y^ϵ respectively. Define for $(\theta, \eta) \in \Omega^n \times \Omega_0$,

$$I_{X,Y}(\theta,\eta) = \begin{cases} \frac{1}{2}S_0(\theta_0) + \frac{1}{2}\int_0^T \mid \dot{\theta}_s - f(\theta_s) \mid^2 + \mid \dot{\eta}_s - h(\theta_s) \mid^2 ds \\ \qquad \text{if } (\theta,\eta) \text{ is absolutely continuous,} \\ \infty \qquad \text{otherwise;} \end{cases} \tag{2.39}$$

and for $\eta \in \Omega_0$,

$$I_Y(\eta) \ = \ \inf_{\theta \in \Omega^n} I_{X,Y}(\theta, \eta). \tag{2.40}$$

Corollary 2.1 *$\{P^\epsilon_{X,Y}\}$ and $\{P^\epsilon_Y\}$ obey the LDP with action functions $I_{X,Y}(\theta,\eta)$ and $I_Y(\eta)$ respectively.*

2.2.3 Large Deviations for Random Measures

As before let $(\Omega, \mathcal{F}, P)$ be a probability space, and let X be a complete separable metric space. We are interested in the following situation: $\{\mathsf{Q}^\epsilon\}$ is a family of random probability measures on X with

$$\mathsf{Q}^\epsilon \stackrel{P}{\Rightarrow} \delta_{\mathsf{x}_0}$$

for some $\mathsf{x}_0 \in \mathsf{X}$. That is, the probability measures Q^ϵ converge weakly in probability to the point mass at x_0: for all $\alpha > 0$,

$$\lim_{\epsilon \to 0} P\left(\varrho(\mathsf{Q}^\epsilon, \delta_{\mathsf{x}_0}) > \alpha\right) \;=\; 0.$$

Let now $\{\mathsf{Q}^\epsilon\}$ be a family of random probability measures on X.

Definition 2.1 *We say that* $\{\mathsf{Q}^\epsilon\}$ *obeys the* <u>*large deviation principle in probability*</u> *(LDPP) with action function* $J : \mathsf{X} \to [0, \infty]$ *provided:*

(i) *J is lower semicontinuous;*

(ii) *For each $M > 0$, the set $\{\mathsf{x} \in \mathsf{X} : J(\mathsf{x}) \leq M\}$ is compact;*

(iii) *For any closed set $C \subset \mathsf{X}$,*

$$\limsup_{\epsilon \to 0} \epsilon \log \mathsf{Q}^\epsilon(C) \leq -\inf_{\mathsf{x} \in C} J(\mathsf{x}) \quad \text{in probability;}$$

(iv) *For any open set $G \subset \mathsf{X}$,*

$$\liminf_{\epsilon \to 0} \epsilon \log \mathsf{Q}^\epsilon(G) \geq -\inf_{\mathsf{x} \in G} J(\mathsf{x}) \quad \text{in probability.}$$

Remark This definition is smiliar to a definition by Ji [31], who considered filtering problems where a small parameter appears in the observation equation only, and not in the state equation. ///

Chapter 3
The Observer Problem

In this chapter we define what we mean by the term *observer*. The definition given is rather general, in as much as the observer state space may be infinite dimensional, as is typically the case in nonlinear filtering. We discuss the well known definitions concerning *observability* in the light of our definition, and introduce a notion of *uniform detectability* which extends the definition of detectability for linear systems.

The asymptotic filtering method is applied to detectable linear systems. Finally, we review several approaches proposed by other authors.

3.1 Problem Statement

We wish to design an observer for the nonlinear control system

$$\begin{aligned} \dot{x} &= f(x,u), \quad x(0) = x_0, \\ y &= h(x), \end{aligned} \tag{3.1}$$

with *state* $x(t) \in I\!R^n$, *control* $u(t) \in [-1,1]^m$, and *observation* $y(t) \in I\!R^p$. The initial condition $x_0 \in I\!R^n$ is unknown.

The problem is to use the available information $t \mapsto u(t)$ and $t \mapsto y(t)$ to compute an estimate $\hat{x}(t)$ of the state $x(t)$ such that the error $e(t) = x(t) - \hat{x}(t)$ converges to zero as $t \to \infty$.

In general, some a priori knowledge concerning the initial condition x_0 is required. We quantify this by saying that x_0 belongs to some class $\mathcal{I}$ which depends on: the problem data f, h; the initial estimate $\hat{x}_0$; and on various design parameters. For example, $\mathcal{I} = \{x_0 \in \mathbb{R}^n \ : \mid x_0 - \hat{x}_0 \mid < \rho\}$. It is desirable that $\mathcal{I}$ be as large as possible, to maximise the permissible initial uncertainty.

Definition 3.1 *An* <u>*observer*</u> *for the control system* (3.1) *is a dynamical system*

$$\begin{aligned} \dot{m} &= F(m, u, y), \quad m(0) = m_0, \\ \hat{x} &= G(m), \end{aligned} \tag{3.2}$$

such that

$$\lim_{t \to \infty} \mid x(t) - \hat{x}(t) \mid = 0 \tag{3.3}$$

for all $x_0 \in \mathcal{I}$, where $\mathcal{I}$ is a subset of $\mathbb{R}^n$.

Here, $m(t) \in \mathcal{M}$ and $\hat{x}(t) \in \mathbb{R}^n$. The observer state space $\mathcal{M}$ may be infinite dimensional (c.f. nonlinear filtering). Of course, observers for which $\mathcal{M} = \mathbb{R}^l$, for some $l < \infty$, are most useful.

3.2 Observability and Detectability

For an observer design to be successful, the measurements $t \to u(t)$ and $t \to y(t)$ must contain "enough" information about the state trajectory. This information, together with $\mathcal{I}$, is used by the observer to estimate the state. In an effort to make precise what "enough" means, in this section, we study various notions of observability.

Observability addresses the question: what does the information contained in the measurements reveal about the states? More precisely, consider the control system (3.1) and let γ_u denote the flow corresponding to a control $u : t \mapsto u(t)$. Let x_0^1, x_0^2 denote two initial conditions. Can the states x_0^1 and x_0^2 be distinguished by comparing their observation records for some control u?

Definition *The control system* (3.1) *is* <u>*observable*</u> *if given any* $x_0^1 \neq x_0^2 \in \mathbb{R}^n$ *there exists a control* u *such that*

$$h\left(\gamma_u(t)x_0^1\right) \neq h\left(\gamma_u(t)x_0^2\right) \tag{3.4}$$

for some $t \geq 0$. *If* (3.4) *holds for some* u, *we say that the states* x_0^1 *and* x_0^2 *are* <u>*distinguishable*</u> *with respect to the control* u.

Note that in general, $p \leq n$ and $h : \mathbb{R}^n \to \mathbb{R}^p$ is not one to one: h is often a submersion. Also, the control enters the definition in an essential way. There may be many controls for which x_0^1 and x_0^2 are *indistinguishable.*

In general, it is difficult to give criteria for testing whether or not a given system is observable. A number of authors have addressed this issue, most notably in the pioneering work of Hermann and Krener [24]. See also Isidori [28].

There is an algebraic test for a weak form of local observability. Let U be a neighbourhood of $x_0 \in \mathbb{R}^n$. Define

$$\begin{aligned} I(x_0, U) \;=\; & \{x_0^1 \in U \;:\; x_0^1 \text{ is not distinguishable from } x_0 \text{ with} \\ & \quad \text{respect to } u \text{ and } \gamma_u(t)x_0 \in U,\ \gamma_u(t)x_0^1 \in U \text{ for all } t\}. \end{aligned}$$

Definition *The control system* (3.1) *is said to have the* <u>*local distinguishability property*</u> *if every* $x_0 \in \mathbb{R}^n$ *has an open neighbourhood* V *such that for all open neighbourhoods* U *of* x_0, $U \subset V$, *one has* $I(x_0, U) \cap V = \{x_0\}$.

Let

$$\langle L_f \mid dh \rangle = \operatorname{Span}\left\{L^i_{f(\cdot,u)}(dh^j) : u \in [-1,1]^m,\ j = 1, \ldots, p;\ i = 1, 2, \ldots\right\} \tag{3.5}$$

denote the smallest codistribution containing $\{dh^1, \cdots, dh^p\}$ which is invariant under the vector fields $\{f(\cdot, u) \ : \ u \in [-1,1]^m\}$. Here, $L_{f(\cdot,u)}$ denotes the Lie derivative with respect to the vector field $f(\cdot, u)$, and dh^j is the gradient 1–form for the real valued function h^j, $j = 1, \ldots, p$, and the span is taken with respect to the ring $C^\infty(I\!R^n)$. This is called the *observability codistribution* [24].

Definition *The control system* (3.1) *is said to satisfy the* *observability rank condition* *at* $x_0 \in I\!R^n$ *if there exists a neighbourhood* U *of* x_0 *such that*

$$\operatorname{rank}\langle L_f \mid dh\rangle(x) \;=\; n \tag{3.6}$$

for all $x \in U$.

Theorem *(Hermann–Krener [24]) If the control system* (3.1) *satisfies the observabilty rank condition at* x_0*, then the system has the local distinguishability property at* x_0*. Conversely, if the control system* (3.1) *has the local distinguishability property, then the observability rank condition is satisfied generically.*

The idea is that if the observability rank condition fails, then the leaves of the foliation induced by the annihilator of the observability codistribution form equivalence classes of indistinguishable states. Such local results amount essentially to obtaining information by "differentiating the output".

The definition of observability is connected with the problem of determining exactly the unknown state. As we shall see in Chapter 6, it turns out that, rather surprisingly, a notion weaker than observability is sufficient for successful observer construction. We are not necessarily trying to recover the state exactly. Rather, it suffices to estimate the state trajectory asymptotically as $t \to \infty$. One can construct observers for *detectable* linear systems (defined in the example below) which are *not* observable. This was first noted by Wonham; see [59], Chapter 3. The situation is similar for nonlinear systems; see Chapter 6.

Define

$$\begin{aligned} A(x,u) &= Df(x,u), \\ H(x) &= Dh(x), \end{aligned}$$

where D denotes the gradient in the x variables. *Assume* that

$$\begin{aligned} \| A \| &= \sup_{x \in \mathbb{R}^n,\ |u| \leq 1} \| A(x,u) \| < \infty, \\ \| H \| &= \sup_{x \in \mathbb{R}^n} \| H(x) \| < \infty. \end{aligned}$$

Definition 3.2 *The control system* (3.1) *is* <u>*uniformly detectable*</u> *if there exists a bounded continuous matrix valued map* $x, u \mapsto \Lambda(x,u)$ *and a constant* $\alpha_0 > 0$ *such that*

$$\eta' \left(A(x,u) + \Lambda(x,u) H(x) \right) \eta \ \leq\ -\alpha_0 \mid \eta \mid^2 \tag{3.7}$$

for all $\eta \in \mathbb{R}^n,\ x \in \mathbb{R}^n,\ u \in [-1,1]^m$.

A disadvantage of this condition is that it is in general difficult to check, and $\Lambda(x,u)$ may be hard to compute. No simple rank–type condition exists to date. A simple but less general condition is the following.

Definition 3.3 *The control system* (3.1) *is* <u>*uniformly of full rank*</u> *if there exists a constant* $s_0 > 0$ *such that*

$$H(x)' H(x) \ \geq s_0 I \tag{3.8}$$

for all $x \in \mathbb{R}^n$.

EXAMPLE *(Linear Systems)* Consider

$$\begin{aligned} \dot{x} &= Ax + Bu, \\ y &= Cx. \end{aligned} \tag{3.9}$$

The system (3.9) (or the pair (C,A)) is *observable* if

$$\bigcap_{i=1}^{n} \ker(CA^{i-1}) = \emptyset.$$

Here,

$$\langle L_f \mid dh \rangle = \mathrm{span}\{C, CA, \cdots, CA^{n-1}\}$$

and so observability is equivalent to the rank condition

$$\mathrm{rank}[C, CA, \cdots, CA^{n-1}] = n.$$

Let $\mathcal{S}(A)$ denote the stable subspace of A, that is, the span of the (generalised) eigenvectors corresponding to eigenvalues with strictly negative real part. The pair (C,A) is *detectable* if

$$\bigcap_{i=1}^{n} \ker(CA^{i-1}) \subset \mathcal{S}(A).$$

This is equivalent to the existence of a matrix Λ such that the eigenvalues of the matrix $A + \Lambda C$ have strictly negative real parts (Wonham [59]). Thus uniform detectability for (C, A) implies detectability, but *not* conversely. States x_0^1, x_0^2 are indistinguishable if

$$x_0^1 - x_0^2 \in \bigcap_{i=1}^{n} \ker(CA^{i-1}).$$

So observability implies detectability, but not conversely. ///

Thus uniform detectability requires roughly speaking that states which are unobservable decay exponentially to zero.

Remark Filtering theory does not generally address the observability issue. Mitter [45] mentions observability in connection with his discussion of the extended Kalman filter; we discuss this further in Chapter 7. The conditional expectation gives the conditional minimum variance estimate, but the error does not in general converge to zero as $t \to \infty$; see Kunita [36]. ///

3.3 Application to Linear Systems

In this section we provide a complete description of the asymptotic filtering method as it applies to detectable linear systems. For further details, refer to Baras, Bensoussan and James [2].

The method constructs explicitly an observer for the linear system

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t), \quad x(0) = x_0, \\ y(t) &= Cx(t), \end{aligned} \tag{3.10}$$

as the asymptotic limit of (Kalman) filters for a family of associated filtering problems

$$\begin{aligned} dx^\epsilon(t) &= Ax^\epsilon(t)dt + Bu(t)dt + \sqrt{\epsilon}Ndw(t), \quad x^\epsilon(0) = x_0^\epsilon, \\ d\xi^\epsilon(t) &= Cx^\epsilon(t)dt + \sqrt{\epsilon}Rdv(t), \quad \xi^\epsilon(0) = 0. \end{aligned} \tag{3.11}$$

Such a construction is suggested by the fact that for certain choices of $Q_0^\epsilon = \mathrm{cov}(x_0^\epsilon)$, the filters are independent of ϵ, as discussed in Baras and Krishnaprasad [1].

The work of Hijab [25], [26] is indispensible here in deriving a large deviation principle for certain conditional measures (see Chapter 5), and identifying the limit of the filters for (3.11) as an associated deterministic estimator (Chapter 4).

We assume that $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$, $y(t) \in \mathbb{R}^p$, and $t \mapsto u(t)$ is piecewise continuous.

The *observer* problem here consists of constructing a dynamical system

$$\begin{aligned} \dot{m}(t) &= Em(t) + Fu(t) + Gy(t), \quad m(0) = m_0, \\ \hat{x}(t) &= Hm(t), \end{aligned} \tag{3.12}$$

so that the error

$$e(t) = x(t) - \hat{x}(t) \tag{3.13}$$

decays exponentially fast to zero, at a rate controlled by the designer, independent from the choice of m_0 and x_0. Here the matrices E, F, G and H are possibly time-varying and the dimension of $m(t)$ is not necessarily n.

Solutions to this problem are well known, first given by Luenberger [44]. In particular, if the pair (C,A) is *detectable*, then there exists a matrix Γ such that the matrix $A + \Gamma C$ has eigenvalues in the open left half plane. Then set

$$E = A + \Gamma C, \quad F = B, \quad G = -\Gamma, \quad H = I.$$

In this case the error (3.13) satisfies

$$\dot{e}(t) = (A + \Gamma C)e(t), \quad e(0) = x_0 - m_0,$$

and the eigenvalues of $A + \Gamma C$ can be arbitrarily assigned by the designer if and only if (C,A) is *observable*. Typically Γ is selected by transforming (C, A) into the observer standard form and choosing the coefficients of Γ to achieve a desired characteristic polynomial for $A+\Gamma C$. Alternatively, one may employ the grammian approach for stabilisation of Kleinman [37].

Consider the system (3.10). Define $\xi(t) = \int_0^t y(s)ds$, so that (3.10) becomes

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t), \quad x(0) = x_0, \\ \dot{\xi}(t) &= Cx(t), \quad \xi(0) = 0. \end{aligned} \tag{3.14}$$

Then associate with (3.14) the family of filtering problems (3.11), where w, v are independent standard k-dimensional, respectively p-dimensional Brownian motions. The initial condition x_0^ϵ is Gaussian, independent from w, v with $E(x_0^\epsilon) = m_0{}^\epsilon$, $\mathrm{cov}(x_0^\epsilon)=Q_0^\epsilon$, where Q_0^ϵ is positive definite. The matrix R is assumed positive definite.

As is well known, the minimum variance estimate $\hat{x}^\epsilon(t) = E(x(t) \mid \xi^\epsilon(s), 0 \leq s \leq t)$ for the linear Gaussian filtering problem (3.11) is given by the Kalman filter [10]

$$\begin{aligned} d\hat{x}^\epsilon(t) &= A\hat{x}^\epsilon(t)dt + Bu(t)dt + Q^\epsilon(t)C'(RR')^{-1}(d\xi^\epsilon(t) - C\hat{x}^\epsilon(t)dt), \\ \hat{x}^\epsilon(0) &= m_0{}^\epsilon, \end{aligned} \tag{3.15}$$

where Q^ϵ satisfies the Riccati equation

$$\begin{aligned} \dot{Q}^\epsilon(t) &= AQ^\epsilon(t) + Q^\epsilon(t)A' - Q^\epsilon(t)C'(RR')^{-1}CQ^\epsilon(t) + NN', \quad (3.16) \\ Q^\epsilon(0) &= Q_0^\epsilon/\epsilon. \end{aligned}$$

Note that these filters depend on ϵ only via the matrix Q_0^ϵ/ϵ. In fact, if we choose $Q_0^\epsilon = \epsilon Q_0$, then all the filters are independent of ϵ and identical with the filter for $\epsilon = 1$.

Define

$$\begin{aligned} \dot{m}(t) &= Am(t) + Bu(t) + Q(t)C'(RR')^{-1}(y(t) - Cm(t)), \quad (3.17) \\ m(0) &= m_0, \end{aligned}$$

where $Q(t)$ is the solution of the Riccati equation

$$\begin{aligned} \dot{Q}(t) &= AQ(t) + Q(t)A - Q(t)C'(RR')^{-1}CQ(t) + NN', \quad (3.18) \\ Q(0) &= Q_0 \end{aligned}$$

As is discussed in Chapter 4, $m(t) = \hat{x}(t)$ is the deterministic minimum energy estimate of $x(t)$ given the observation record $Y_t = \{y(s); 0 \leq s \leq t\}$. This is closely related to a large deviation principle for the conditional measures (Chapter 5), and (3.17), (3.18) is the limiting filter. In fact, if $\lim_{\epsilon \to 0} m_0^\epsilon = m_0$ and $\lim_{\epsilon \to 0} Q_0^\epsilon = Q_0$, one can prove:

$$\lim_{\epsilon \to 0} E \mid \hat{x}^\epsilon(t) - m(t) \mid^2 = 0$$

uniformly on $[0, T]$. In fact, $m(\cdot)$ coincides with the solution of (3.10) for the initial condition $x(0) = m_0$.

We are interested in the asymptotic behaviour of $e(t) = x(t) - m(t)$ as $t \to \infty$, and we make the natural assumption that the pair (C, A) is *detectable.* We also assume rank $N = n$ and $R > 0$. It is important to obtain bounds on $\| Q(t) \|$ and $\| P(t) \|$, where $P(t) = Q(t)^{-1}$. As will be explained in Section 6.2, there exist constants $q > 0,\ p > 0$ such that

$$\| Q(t) \| \leq q, \quad \| P(t) \| \leq p$$

for all $t \geq 0$. As in [2] and Section 6.3, one can use the Lyapunov function $x'P(t)x$ to prove

$$\mid x(t) - m(t) \mid \leq K \mid x_0 - m_0 \mid e^{-\gamma t},$$

for all $t \geq 0$, for some $K > 0,\ \gamma > 0$.

Thus the deterministic filter (3.17), (3.18) is an observer for the linear control system (3.10). This design is a useful alternative to the standard Luenberger design, and can be easily be extended to time varying systems provided that we assume uniform complete observability (Kalman's terminology [35]).

3.4 Other Approaches

In this section we give a brief survey of some observer designs proposed by other authors that have come to our attention. This discussion will serve to place our approach in perspective.

Since the linear theory is well understood, it is perhaps natural to use linear theory where possible in designing observers for nonlinear systems. One obvious approach is to linearise the system (3.1) about an equilibrium x_0, u_0: $f(x_0, u_0) = 0$. Expanding $x(t) = x_0 + \delta x(t) + \ldots$, $u(t) = u_0 + \delta u(t) + \ldots$, we obtain the linear system

$$\begin{aligned} \delta\dot{x}(t) &= A_0 \delta x(t) + B_0 \delta u(t), \\ \delta y(t) &= C_0 \delta x(t). \end{aligned} \tag{3.19}$$

Then if (C_0, A_0) is detectable, one might design an observer for (3.19) and use it to estimate $x(t)$, the state of (3.1). This of course would be valid only for small excursions from the equilibrium.

Another approach is to try to find coordinates in which the system exhibits suitable linear features. For example, Krener and Respondek [39] consider the system

$$\begin{aligned} \dot{\xi} &= f(\xi, u), \quad \xi(0) \approx \xi_0, \\ \phi &= h(\xi), \end{aligned} \tag{3.20}$$

in some neighbourhood of a given point ξ_0. Here, $\xi \in I\!R^n$, $\phi \in I\!R^p$. Under certain conditions, they find coordinate transformations $\Xi : I\!R^n \to I\!R^n$, $\Phi : I\!R^p \to I\!R^p$,

$$\xi = \Xi(x), \quad \phi = \Phi(y),$$

so that (3.20) becomes

$$\begin{aligned} \dot{x} &= Ax + \gamma(y, u), \\ y &= Cx, \end{aligned} \tag{3.21}$$

where (C, A) is an observable pair in observer standard form. Then Γ can be chosen so that $A + \Gamma C$ has eigenvalues with negative real parts and an observer for (3.21) is

$$\dot{z} = (A + \Gamma C)z - \Gamma y + \gamma(y, u),$$

with error equation ($e = x - z$)

$$\dot{e} = (A + \Gamma C)e.$$

To obtain an observer for (3.20), one writes the ODE for $\zeta = \Xi(z)$:

$$\dot{\zeta} = F(\zeta, \phi, u).$$

Then at least on a compact region containing ξ_0, the error $\xi - \zeta$ decays exponentially to zero.

Most of their paper [24] is devoted to conditions under which the above procedure is possible. These conditions are related to local observability, and the computation of the coordinate transformations requires the solution of certain PDEs. This approach has been extended by Levine and Marino [42], where the nonlinear system is immersed in a linear system of higher dimension; the transformations are local immersions rather than local diffeomorphisms.

Kuo, Elliott and Tarn [40] consider the uncontrolled system

$$\begin{aligned} \dot{x} &= f(x), \\ y &= h(x). \end{aligned} \tag{3.22}$$

They propose the following design:

$$\dot{z} = f(z) + B(y - h(z)).$$

If there exists a constant matrix B and a positive definite constant matrix Q such that

$$Q\,(Df(x) - BDh(x))$$

is uniformly negative definite, then the error decays exponentially to zero for all initial values. Their proof uses the Lyapunov function $e'Qe$. Notice in particular that B and Q must be constant, and no constructive procedure for selecting these matrices is given.

Observer design for bilinear systems

$$\begin{aligned} \dot{x} &= \left(A + \sum_{i=1}^{m} u_i B_i\right) x \;=\; A(u)x, \\ y &= Cx, \end{aligned} \tag{3.23}$$

is nontrivial. A survey of some design methods is presented in Derese, Stevens and Noldus [12]. A candidate observer is

$$\dot{z} = (A(u) + \Gamma(u)C)\,z - \Gamma(u)y, \tag{3.24}$$

with error equation

$$\dot{e} = (A(u) + \Gamma(u)C)\,e. \tag{3.25}$$

Then in order that (3.24) be an observer for (3.23), one needs to find a matrix valued function $\Gamma(\cdot)$ such that (3.25) is asymptotically stable for all controls $t \mapsto u(t)$. The results discussed in [12] address this issue, and sufficient conditions are obtained which are applicable in various cases. In general, no constructive method for computing $\Gamma(\cdot)$ is available, a difficulty shared with uniform detectability. Williamson [58] views (3.23) as a time varying linear system and seeks to transform this system into a canonical form.

We finally remark that observer theory for linear systems remains an active area of research: O' Reilly [49]. For linear time varying systems, it seems impossible to avoid using the observability grammian; no simple rank–type condition is known. See also Section 6.4.1.

Chapter 4
Deterministic Estimation

In 1968, Mortensen [48] proposed a deterministic method for state estimation (filtering), based on the idea that the most likely state trajectory has minimum energy (in a certain sense). The resulting filter is infinite dimensional, being characterised by a Hamilton–Jacobi equation, and generically the estimate is not computable by a system of finite dimensional equations. This difficulty is shared with the stochastic method of filtering recalled in Chapter 2. In his thesis [25], Hijab studied Mortensen's method and compared it with the stochastic approach. He was interested in the finite dimensional computability issue, as well as small parameter asymptotics.

We employ the relatively recent notion of *viscosity solution* for Hamilton–Jacobi equations and give a new treatment of Mortensen's method. This is fundamental to the use of PDE techniques for proving the small parameter asymptotics in Chapter 5. In addition, we obtain a continuity result when the deterministic estimator is well defined, and introduce an approximate deterministic estimator which will be the basis of our observer design in Chapter 6.

4.1 Formulation

We begin by reviewing Mortensen's method [48], [25] of deterministic minimum energy estimation.

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

$$\dot{x} = f(x) + u, \quad x(0) = x_0, \tag{4.1}$$

$$\dot{y} = h(x) + v, \quad y(0) = 0,$$

we wish to estimate the state at time t, the initial condition x_0 being unknown. Here, $u \in L^2([0,T], \mathbb{R}^n)$, $v \in L^2([0,T], \mathbb{R})$ are deterministic models for noise. We assume that $f \in C_b^1(\mathbb{R}^n, \mathbb{R}^n)$ and $h \in C_b^2(\mathbb{R}^n, \mathbb{R})$. Then

$$Y_t \in \Omega_{0,t}^{H^1} = \{\eta \in \Omega_{0,t} : \eta \text{ is absolutely continuous and } \| \eta \|_{H^1} < \infty\}$$

where

$$\| \eta \|_{H^1} = \left(\int_0^t \dot{\eta}_s^2 ds \right)^{\frac{1}{2}}.$$

The most likely path is one for which u, v have least energy and produce the given observation record Y_t.

More precisely, define

$$\tilde{J}_t(x_0, u, v) = S_0(x_0) + \frac{1}{2} \int_0^t \left(| u(s) |^2 + v(s)^2 \right) ds, \tag{4.2}$$

where S_0 is Lipshitz continuous, convex, and $S_0(m_0) = 0$, $S_0(x) > 0$ if $x \neq m_0 \in \mathbb{R}^n$. A minimum energy input triple (x_0^*, u^*, v^*) given Y_t is a triple that minimises $\tilde{J}_t$ subject to the constraint that the trajectory of (4.1) produces the output Y_t. By replacing $v(s)$ by $\dot{y}(s) - h(x(s))$ in (4.2) and omitting the $\dot{y}(s)^2$ term, we can formulate an equivalent unconstrained optimal control problem. Define

$$J_t(x_0, u) = S_0(x_0) + \int_0^t L(x(s), u(s), s) ds, \tag{4.3}$$

where

$$L(x, u, s) = \frac{1}{2} | u |^2 + \frac{1}{2} h(x)^2 - \dot{y}(s) h(x). \tag{4.4}$$

We now minimise J_t over pairs (x_0, u). The *deterministic* or minimum energy *estimate* $\hat{x}(t)$ given Y_t is defined to be the endpoint of the optimal trajectory

$s \mapsto x^*(s),\ 0 \le s \le t$, corresponding to a *minimum energy pair* (x_0^*, u^*) : $\hat{x}(t) = x^*(t)$.

Next, we use dynamic programming to study this problem. Given a control $t \mapsto u(t)$, let x_u denote the corresponding trajectory (given a specified initial condition). Following the general scheme presented in Fleming and Rishel [19], define a class of admissible pairs (x_0, u) by

$$\mathcal{U}_{x,t} \;=\; \{(x_0, u) \in I\!R^n \times L^2([0,T], I\!R^n) \;:\; x_u(0) = x_0,\ x_u(t) = x\}; \tag{4.5}$$

that is, pairs for which the corresponding trajectory passes through a specified point x at time t. Define a *value function*

$$W(x,t) \;=\; \inf_{(x_0,u)\,\in\,\mathcal{U}_{x,t}} J_t(x_0, u). \tag{4.6}$$

Note that this is a reversal of the standard set–up of dynamic programming [19]. By using standard methods, we see that $W(x,t)$ is continuous and formally satisfies the *Hamilton–Jacobi–Bellman* (HJB) equation

$$\frac{\partial}{\partial t} W(x,t) \;+\; \tilde{H}(x,t,DW(x,t)) \;=\; 0, \tag{4.7}$$

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

where

$$\tilde{H}(x,t,\lambda) \;=\; \max_{u\,\in\,U} \{\lambda \cdot (f(x) + u) - L(x,u,t)\}. \tag{4.8}$$

$W(x,t)$ is the minimum value of J_t subject to the end point condition $x_u(t) = x$.

The deterministic estimate is given by:

$$\hat{x}(t) \;=\; \underset{x\,\in\,I\!R^n}{\operatorname{argmin}}\; W(x,t), \tag{4.9}$$

and we say that the deterministic estimator is *well defined* if $\hat{x}(t)$ is the unique minimiser of $W(\cdot,t)$ for each t and each observation record. When we wish to make explicit the dependence on the observation path, we write $W(x,t)[Y]$, etc. As described by Hijab [25], this gives a *deterministic estimator (filter)* $\bar{\pi}$:

$$
\begin{aligned}
\bar{\pi}_t \; : \; \Omega^{H^1}_{0,t} &\rightarrow \mathcal{P}(I\!R^n) \\
Y_t &\mapsto \delta_{\hat{x}(t)[Y_t]}.
\end{aligned} \tag{4.10}
$$

In control theoretic terms, we can regard the HJB equation (4.7) as part of an *infinite dimensional realisation* of the map $\bar{\pi}$ (c.f. Hijab [25]).

Proposition 4.1 *For any $Y_t \in \Omega^{H^1}_{0,t}$, there exists a minimum energy pair (x_0^*, u^*). If the map $\breve{J}_t : \Omega^n \rightarrow \bar{I\!R}$ defined by*

$$
\breve{J}_t(\theta) = S_0(\theta_0) + \int_0^t \frac{1}{2} \mid \dot{\theta}_s - f(\theta_s) \mid^2 + \frac{1}{2} h(\theta_s)^2 - \dot{y}_s h(\theta_s) ds \tag{4.11}
$$

is strictly convex, then the minimum energy pair (x_0^*, u^*) is unique. If this holds for each $0 \leq t \leq T$, then the deterministic estimator (4.10) *is well defined.*

Proof: Notice that

$$
\breve{J}_t(\theta) \;=\; J_t(\theta_0, \dot{\theta} - f(\theta)).
$$

One can check that $\breve{J}_t$ is lower semicontinuous (uniform convergence) and for each $M > 0$, $\{\theta \in \Omega^n : \breve{J}_t(\theta) \leq M\}$ is compact (uniform topology). Hence there exists $\theta^* \in \Omega^n$ such that

$$
\breve{J}_t(\theta^*) \;=\; \min_{\theta \in \Omega^n} \breve{J}_t(\theta).
$$

Then $(\theta_0^*, \dot{\theta}^* - f(\theta^*))$ is a minimum energy pair. The remaining assertions are easily verified. ∎

EXAMPLE (*Linear Systems; Krener [38]*) Consider the system

$$
\begin{aligned}
\dot{x} &= Ax + Nu, \; x(0) = x_0, \\
\dot{y} &= Cx + Rv,
\end{aligned} \tag{4.12}
$$

with energy or cost function

$$\begin{aligned} J_t(x_0,u) &= \frac{1}{2}(x_0-m_0)'P_0(x_0-m_0) \\ &+\int_0^t \frac{1}{2} \mid u(s) \mid^2 + \frac{1}{2} \mid R^{-1}Cx(s) \mid^2 - \dot{y}(s)(RR')^{-1}Cx(s)ds, \end{aligned}$$

where $P_0 > 0$, $R > 0$, and $\mathrm{rank}N = n$. The value function has the explicit form

$$W(x,t) = \frac{1}{2}(x-\hat{x}(t))'P(t)(x-\hat{x}(t)) + \int_0^t \frac{1}{2} \mid R^{-1}C\hat{x}(s) \mid^2 - \dot{y}(s)(RR')^{-1}C\hat{x}(s)ds,$$

where $\hat{x}(t)$ is the deterministic estimate, given by

$$\dot{\hat{x}}(t) = A\hat{x}(t) + P(t)^{-1}C'(RR')^{-1}(\dot{y}(t) - C\hat{x}(t)), \quad \hat{x}(0) = m_0, \tag{4.13}$$

and $P(t)$ is the positive definite solution of the Riccati equation

$$\begin{aligned} \dot{P}(t) &= -P(t)A - A'P(t) - P(t)NN'P(t) + C'(RR')^{-1}C, \qquad (4.14) \\ P(0) &= P_0. \end{aligned}$$

Thus the deterministic estimator for the linear system (4.12) is well defined and has a finite dimensional realisation (4.13), (4.14). ///

4.2 Viscosity Solution

We now turn to the HJB equation (4.7). If $Y \in \Omega_0^{H^1}$, then the Hamiltonian $\tilde{H}$ (4.8) is in general only measurable in t. In this section we restrict ourselves to

$$Y_t \in \Omega_{0,t}^{C^1} = \left\{\eta \in C^1([0,t], \mathbb{R}) : \eta(0) = 0\right\}.$$

This space is equipped with the norm

$$\| \eta \|_{C^1} = \sup_{0\leq s\leq t} \mid \dot{\eta}(s) \mid .$$

Then $\tilde{H} \in C(\mathbb{R}^n \times [0,T] \times \mathbb{R}^n; \mathbb{R})$.

Now we prove that $W(x,t)$ is the unique viscosity solution of the Hamilton–Jacobi–Bellman equation (4.7). Our assumptions imply that $\mathcal{U}_{x,t} \neq \emptyset$ for all $x \in \mathbb{R}^n$, $0 \leq t \leq T$, and consequently $W(x,t) < \infty$. The analysis does not assume existence of optimal controls.

The following definition is taken from Crandall, Evans and Lions [9]. Write $C \equiv C(\mathbb{R}^n \times (0,T),\ \mathbb{R})$, and similarly for $C^1 \equiv C^1(\mathbb{R}^n \times (0,T), \mathbb{R})$.

Definition *Let $W \in C$. We say that W is a viscosity subsolution of* (4.7) *provided that for all $\phi \in C^1$ the following property holds:*

if $W - \phi$ attains a local maximum at a point $(x,t) \in \mathbb{R}^n \times (0,T)$, then

$$\tfrac{\partial}{\partial t}\phi(x,t) \ + \ \tilde{H}(x,t,D\phi(x,t)) \ \leq \ 0. \tag{4.15}$$

We say that W is a viscosity supersolution of (4.7) *provided that for all $\phi \in C^1$ the following property holds:*

if $W - \phi$ attains a local minimum at a point $(x,t) \in \mathbb{R}^n \times (0,T)$, then

$$\tfrac{\partial}{\partial t}\phi(x,t) \ + \ \tilde{H}(x,t,D\phi(x,t)) \ \geq \ 0. \tag{4.16}$$

If W is both a viscosity subsolution and supersolution, we say that W is a viscosity solution of (4.7).

Lemma 4.1 *(Principle of Optimality) Let $0 \leq t-h \leq t$. Then*

$$W(x,t) \ = \inf_{(x_0,u)\in\mathcal{U}_{x,t}} \left\{ W(x_u(t-h),t-h) \ + \ \int_{t-h}^{t} L(x_u(s),u(s),s)ds \right\}. \tag{4.17}$$

Proof: Let $(x_0,u) \in \mathcal{U}_{x,t}$ and choose $(\tilde{x}_0,\tilde{u}) \in \mathcal{U}_{x_u(t-h),t-h}$. Define

$$\bar{u}(s) = \begin{cases} \tilde{u}(s) & 0 \leq s \leq t-h \\ u(s) & t-h \leq s \leq t. \end{cases}$$

Then $\tilde{u} \in \mathcal{U}_{x,t}$, and hence

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

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

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

Now let $\delta > 0$. Choose $(\tilde{x}_0, \tilde{u}) \in \mathcal{U}_{x,t}$ such that

$$\begin{aligned} W(x,t) + \delta &\geq S_0(x_{\tilde{u}}(0)) + \int_0^t L(x_{\tilde{u}}(s), \tilde{u}(s), s) ds \\ &\geq W(x_{\tilde{u}}(t-h), t-h) + \int_{t-h}^{t} L(x_{\tilde{u}}(s), \tilde{u}(s), s) ds \\ &\geq \inf_{(x_0,u) \in \mathcal{U}_{x,t}} \left\{ W(x_u(t-h), t-h) + \int_{t-h}^{t} L(x_u(s), u(s), s) ds \right\}. \end{aligned}$$

These inequalities imply (4.17). ∎

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

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

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

Lemma 4.2 *Fix $\epsilon > 0$. Then there exists $\eta > 0$ such that if $(x_0, u) \in \mathcal{U}_{x,t}^{\gamma}$ then $x_u(t-h) \in B_\epsilon$ 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

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

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 $\eta_{u_n} \to 0$ 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 $\eta > 0$ and the lemma is proved. ∎

Theorem 4.1 *The value function $W(x,t)$ defined by* (4.6) *is the unique viscosity solution of the Hamilton–Jacobi–Bellman equation* (4.7).

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

$$W(x,t) - \phi(x,t) \ \geq \ W(x',t') - \phi(x',t') \tag{4.18}$$

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

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 $0 < \delta \leq \epsilon$ such that $x_{\tilde{u}}(s) \in B_\epsilon$ for $\mid t - s \mid \leq \delta$. The Principle of Optimality (4.17) implies

$$W(x,t) \ \leq \ W(x_{\tilde{u}}(t-h), t-h) + \int_{t-h}^{t} L(x_{\tilde{u}}(s), \tilde{u}(s), s) ds. \tag{4.19}$$

If $0 \leq h \leq \delta$, then (4.18) gives

$$W(x,t) - \phi(x,t) \ \geq \ W(x_{\tilde{u}}(t-h), t-h) - \phi(x_{\tilde{u}}(t-h), t-h). \tag{4.20}$$

Combining (4.19) and (4.20) 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

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

But this holds for all $u \in U$, hence (4.15) and so $W(x,t)$ is a subsolution of (4.7).

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

$$W(x,t) - \phi(x,t) \ \leq \ W(x',t') - \phi(x',t',) \tag{4.21}$$

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

Suppose, contrary to (4.16), 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,

$$\tfrac{\partial}{\partial t}\phi(x',t') + \max_{u \in U}\{D\phi(x',t')\,(f(x') + u) - L(x',u,t')\} < -\theta < 0 \tag{4.22}$$

for all $x' \in B_\epsilon$, $|\, t - t' \,| \leq \epsilon$. Let $\gamma > W(x,t)$ and let η be given as in Lemma 3.2. By the Principle of Optimality (4.17) we have

$$W(x,t) \;=\; \inf_{(x_0,u)\,\in\,\mathcal{U}^\gamma_{x,t}} \{W(x_u(t-h),t-h) + \int_{t-h}^{t} L(x_u(s),u(s),s)ds\}. \tag{4.23}$$

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

$$W(x_u(t-h),t-h) + \int_{t-h}^{t} L(x_u(s),u(s),s)ds \;\leq\; W(x,t) + \frac{\theta h}{2}. \tag{4.24}$$

Since $x_u(t-h) \in B_\epsilon$, we have from (4.21)

$$W(x_u(t-h),t-h) - \phi(x_u(t-h),t-h) \;\geq\; W(x,t) - \phi(x,t). \tag{4.25}$$

Combining (4.24) and (4.25) we have

$$-\frac{\theta}{2} \;\leq\; \frac{\phi(x_u(t-h),t-h) - \phi(x,t)}{-h} - \frac{1}{h}\int_{t-h}^{t} L(x_u(s),u(s),s)ds. \tag{4.26}$$

However, for $t - h \leq s \leq t$, $x_u(s) \in B_\epsilon$ and $|\, t - s \,| < \epsilon$, and so from (4.22) 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

$$\frac{\phi(x,t) - \phi(x_u(t-h),t-h)}{h} - \frac{1}{h}\int_{t-h}^{t} L(x_u(s),u(s),s)ds \;<\; -\theta. \tag{4.27}$$

But (4.26) and (4.27) contradict each other, so we must have $\theta \leq 0$; proving (4.16). Thus $W(x,t)$ is a supersolution of (4.7).

The uniqueness assertion follows from Ishii [27], Theorem 1. In fact, since $S_0(x)$ is uniformly continuous, it follows from [27] that $W(x,t)$ is also uniformly continuous. ∎

4.3 Continuity

In this section we prove that the deterministic estimator, when well defined, depends continuously on the observation path.

Theorem 4.2 *There exists a positive constant C such that*

$$\sup_{(x,t)\in I\!R^n\times[0,T]} \mid W(x,t)[\eta]-W(x,t)[\zeta]\mid \leq C \parallel \eta-\zeta \parallel_{H^1}$$

for all $\eta,\ \zeta\in\Omega_0^{H^1}$.

Proof: Let $(x,t)\in I\!R^n\times[0,T]$, $\alpha>0$, and $\eta,\ \zeta\in\Omega_0^{H^1}$. Choose $(x_0,u)\in\mathcal{U}_{x,t}$ such that

$$J_t(x_0,u)[\eta] \leq W(x,t)[\eta]+\alpha.$$

Now

$$\begin{aligned}\mid J_t(x_0,u)[\eta]-J_t(x_0,u)[\zeta]\mid &= \mid \int_0^t(\dot\zeta_s-\dot\eta_s)h(x_u(s))ds \mid\\ &\leq \parallel \eta-\zeta \parallel_{H^1}\parallel h \parallel .\end{aligned}$$

Then

$$\begin{aligned}W(x,t)[\zeta] &\leq J_t(x_0,u)[\zeta]\\ &\leq J_t(x_0,u)[\eta]+C\parallel \eta-\zeta \parallel_{H^1}\\ &\leq W(x,t)[\eta]+C\parallel \eta-\zeta \parallel_{H^1}+\alpha.\end{aligned}$$

Since $\alpha>0$ was arbitrary,

$$W(x,t)[\zeta] \leq W(x,t)[\eta]+C\parallel \eta-\zeta \parallel_{H^1} .$$

Similarly,

$$W(x,t)[\eta] \leq W(x,t)[\zeta]+C\parallel \eta-\zeta \parallel_{H^1} .$$

This completes the proof. ∎

Corollary 4.1 *Assume that the deterministic estimator is well defined. Then the map* $\bar{\pi}_t$ *defined by* (4.10) *is continuous.*

Proof: Fix $0 \le t \le T$. It is enough to check that

$$\hat{x}(t)[\eta_i] \to \hat{x}(t)[\eta] \text{ in } I\!R^n \text{ as } \eta_i \to \eta \text{ in } \Omega_0^{H^1}.$$

Suppose not. Then there exists $\delta > 0$ such that

$$|\, \hat{x}(t)[\eta_i] - \hat{x}(t)[\eta] \,| \ge \delta$$

for all i sufficiently large. Since $W(\cdot, t)[\eta]$ has a unique minimum,

$$|\, x - \hat{x}(t)[\eta] \,| \ge \delta$$

implies

$$W(x,t)[\eta] - W(\hat{x}(t)[\eta], t)[\eta] \ge \gamma, \tag{4.28}$$

for some $\gamma > 0$. Also, we have

$$W(\hat{x}(t)[\eta_i], t)[\eta_i] \le W(x,t)[\eta_i]$$

for all $x \in I\!R^n$. From Theorem 4.2,

$$|\, W(x,t)[\eta_i] - W(x,t)[\eta] \,| \le \frac{\gamma}{4},$$

for all i sufficiently large.

Now

$$\begin{aligned} W(\hat{x}(t)[\eta_i], t)[\eta] &\le W(\hat{x}(t)[\eta_i], t)[\eta_i] + |\, W(\hat{x}(t)[\eta_i], t)[\eta] - W(\hat{x}(t)[\eta_i], t)[\eta_i] \,| \\ &\le W(\hat{x}(t)[\eta], t)[\eta_i] + \frac{\gamma}{4} \\ &\le W(\hat{x}(t)[\eta], t)[\eta] + \frac{\gamma}{2}; \end{aligned}$$

contradicting (4.28). ∎

4.4 Approximate Deterministic Estimation

In this section we wish to motivate the observer design appearing in Chapter 6. In Chapter 5 we will show that the deterministic filter is the limiting filter. Thus in terms of the asymptotic filtering approach to observer design, the deterministic estimator is a candidate observer. Unfortunately from a practical point of view, the deterministic estimator is generically infinite dimensional. Also, since $W(x,t)$ is in general only Lipshitz continuous, the deterministic estimate is difficult to compute. Thus an approximate finite dimensional method of deterministic estimation is desirable.

Suppose that $W(x,t)$ is a smooth solution of the HJB equation (4.7), and the deterministic estimator is well defined. Then (4.9) implies

$$DW(\hat{x}(t),t) = 0.$$

Differentiating, we get

$$D^2W(\hat{x}(t),t)\dot{\hat{x}}(t) + D\tfrac{\partial}{\partial t}W(\hat{x}(t),t) = 0.$$

Using (4.7) this gives

$$\begin{aligned} D^2W(\hat{x}(t),t)\dot{\hat{x}}(t) &= D^2W(\hat{x}(t),t)f(\hat{x}(t)) + Dh(\hat{x}(t))(\dot{y}(t) - h(\hat{x}(t))) \\ &\quad + DW(\hat{x}(t),t)(Df(\hat{x}(t)) + D^2W(\hat{x}(t),t)). \end{aligned}$$

Therefore

$$\begin{aligned} \dot{\hat{x}}(t) &= f(\hat{x}(t)) + \tilde{P}(t)^{-1}Dh(\hat{x}(t))'(\dot{y}(t) - h(\hat{x}(t)), \\ \hat{x}(0) &= m_0, \end{aligned}$$

where

$$\tilde{P}(t) = D^2W(\hat{x}(t),t).$$

A similar calculation yields

$$\begin{aligned}\dot{\tilde{P}}(t) &= -\tilde{P}(t)A(\hat{x}(t)) - A(\hat{x}(t))'\tilde{P}(t) - \tilde{P}(t)^2 + H(\hat{x}(t))'H(\hat{x}(t)) \\ &\quad + D^3W(\hat{x}(t),t)(\dot{\hat{x}}(t) - f(\hat{x}(t))) - D^2h(\hat{x}(t))(\dot{y}(t) - h(\hat{x}(t))),\end{aligned}$$

where

$$A(x) = Df(x), \quad H(x) = Dh(x).$$

For linear systems, this last equation reduces to the Riccati equation (4.14).

Continuing in this way one finds in general that $\hat{x}(t)$ cannot be computed by a finite set of ODEs. Moreover, in general D^2W does not exist (D^2W exists a.e. if W is convex in x).

If we omit the high order terms, we obtain the system

$$\begin{aligned}\dot{m}(t) &= f(m(t)) + P(t)^{-1}H(m(t))'(\dot{y}(t) - h(m(t))) \qquad (4.29)\\ m(0) &= m_0, \\ \dot{P}(t) &= -P(t)A(m(t)) - A(m(t))'P(t) - P(t)^2 + H(m(t))'H(m(t)), \\ P(0) &= P_0 \;=\; D^2S_0(m_0) > 0. \qquad (4.30)\end{aligned}$$

The system (4.29), (4.30) is called the *approximate deterministic estimator.* This is the deterministic analogue of the extended Kalman filter (c.f. Mitter [45]), and is *not* what one would obtain by linearising the system at an equilibrium position.

We can interpret this in terms of a HJB equation and hence an optimal control problem. Define

$$W^\alpha(x,t) = \frac{1}{2}(x - m(t))'P(t)(x - m(t)) + \int_0^t \frac{1}{2}h(m(s))^2 - \dot{y}(s)h(m(s))ds.$$

Here, α denotes a measure of the nonlinearity:

$$\alpha \;=\; \max\left(\| D^2f \|, \| D^2h \|\right).$$

One expects that $W^\alpha \to W$ as $\alpha \to 0$.

Then

$$m(t) = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} W^\alpha(x,t)$$

(this is well defined) and W^α solves the HJB equation

$$\begin{aligned} \tfrac{\partial}{\partial t}W^\alpha(x,t) + \tilde{H}^\alpha(x,t,DW^\alpha) &= 0 \\ W^\alpha(x,0) &= S_0(x) \end{aligned} \tag{4.31}$$

where

$$\begin{aligned} \tilde{H}^\alpha(x,t,\lambda) &= \tilde{H}(x,t,\lambda) + V^\alpha(x,t) \\ V^\alpha(x,t) &= (x-m_t)'P_t \int_0^1 \int_0^1 rD^2 f(m_t + rs(x-m_t))(x-m_t)^2 drds \end{aligned}$$

$$- \textstyle\int_0^1 \int_0^1 r \left[\mid H(m_t + rs(x-m_t)) \mid^2 + \right.$$

$$\left. H(m_t + rs(x-m_t))' D^2 h(m_t + rs(x-m_t)) \right] (x-m_t)^2 drds +$$

$$\tfrac{1}{2} \mid H(m_t)(x-m_t) \mid^2 + \textstyle\int_0^1 \int_0^1 rD^2 h(m_t + rs(x-m_t))(x-m_t)^2 drds\dot{y}_t.$$

The optimal control problem corresponding to (4.31) has Lagrangian $L^\alpha(x,u,t)$ $= 1/2 \mid u \mid^2 + 1/2h(x)^2 - \dot{y}(t)h(x) - V^\alpha(x,t)$. This may be useful in comparing $\hat{x}(t)$ and $m(t)$ when α is small. We do not pursue this further here.

Chapter 5
Small Parameter Asymptotics

We study the asymptotics as $\epsilon \to 0$ of the filtering problem introduced in Chapter 2. This entails studying the asymptotics of the corresponding Zakai equation, in both its strong and weak form. We obtain estimates showing $q^\epsilon(x,t) \asymp \exp\left(-\frac{1}{\epsilon}W(x,t)\right)$ in probability as $\epsilon \to 0$ for the unnormalised conditional density using PDE methods, where $W(x,t)$ is the value function arising in Mortensen's deterministic estimation for the limiting observation path. Also, we prove $\sigma_t^\epsilon(A) \asymp \exp\left(-\frac{1}{\epsilon}\inf_{x\in A} W(x,t)\right)$ in probability as $\epsilon \to 0$ for the unnormalised conditional measures, directly using probabilistic methods.

Hijab [25] has studied this asymptotic estimation problem, and he obtained a WKB expansion when $W(x,t)$ is smooth. He identified the limiting filter as Mortensen's deterministic estimator. In addition, Hijab [26] has proved a LDP for pathwise conditional measures, which we recall below.

We obtain a LDP in probability for the conditional measures, and prove that they converge weakly in probability to the Dirac measure concentrated on the deterministic trajectory. In the case that the deterministic estimate is well defined, we prove that the pathwise stochastic filters converge weakly to the deterministic estimator.

5.1 PDE Method

We consider the family of diffusion processes in $\mathbb{R}^n$ (2.2) with real valued observations (2.3). Assume as before that $f \in C_b^1(\mathbb{R}^n, \mathbb{R}^n)$, $h \in C_b^2(\mathbb{R}^n, \mathbb{R})$, and x_0^ϵ has density p_0^ϵ given by (2.1).

As $\epsilon \to 0$ the trajectories of (2.2) converge in probability to the trajectory $X = \{x(t) \,:\, 0 \leq t \leq T\}$ of the deterministic system (2.35) with initial condition x_0, and the trajectories of (2.3) converge in probability to $Z = \{z(t) \,:\, 0 \leq t \leq T\} \in \Omega_0$, where

$$z(t) = \int_0^t h(x(s))ds. \tag{5.1}$$

Let q^ϵ be the solution of the Zakai equation (2.27). Our objective in this section is to obtain the asymptotic formula

$$q^\epsilon(x,t) = \exp\left(-\frac{1}{\epsilon}\left(W(x,t) + o(1)\right)\right), \tag{5.2}$$

in probability as $\epsilon \to 0$, where $W(x,t)$ is the value function arising in deterministic estimation for the observation path $Z \in \Omega_0$ (see (5.29) below).

5.1.1 A Hamilton–Jacobi Equation

Fix $Y \in \Omega_0$ and denote by $p^\epsilon(x,t) = p^\epsilon(x,t)[Y]$ the solution of the robust Zakai equation (2.31). We will recover the probabilistic interpretation in Section 5.1.3.

Following Fleming and Mitter [18], who considered filtering problems with $\epsilon = 1$, we apply the logarithmic transformation

$$S^\epsilon(x,t) = -\epsilon \log p^\epsilon(x,t). \tag{5.3}$$

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

$$\begin{aligned} \tfrac{\partial}{\partial t} S^\epsilon(x,t) - \frac{\epsilon}{2}\Delta S^\epsilon(x,t) + H^\epsilon(x,t,DS^\epsilon(x,t)) &= 0, \\ S^\epsilon(x,0) &= S_0(x) - \epsilon \log C_\epsilon, \end{aligned} \tag{5.4}$$

where

$$H^\epsilon(x,t,\lambda) = \lambda \cdot g(x,t) + \frac{1}{2} \mid \lambda \mid^2 - V^\epsilon(x,t), \tag{5.5}$$

$$V^\epsilon(x,t) = V^\epsilon(x,y(t)) + \epsilon \mathrm{div} g(x,y(t)),$$
$$g(x,t) = g(x,y(t)).$$

Here $V^\epsilon(x,y)$, $g(x,y)$ are defined by (2.13), (2.16). Equation (5.4) is a nonlinear parabolic PDE, which can be interpreted as the Bellman equation for a stochastic control problem [18].

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

$$\frac{\partial}{\partial t}S(x,t) + H(x,t,DS(x,t)) = 0, \tag{5.6}$$

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

where

$$H(x,t,\lambda) = \lambda \cdot g(x,t) + \frac{1}{2} \mid \lambda \mid^2 - V(x,t), \tag{5.7}$$

$$V(x,t) = \frac{1}{2}h(x)^2 + y(t)Dh(x)f(x) - \frac{1}{2}y(t)^2 \mid Dh(x) \mid^2 . \tag{5.8}$$

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

We shall interpret solutions of (5.6) in the viscosity sense. In fact, $S(x,t)$ can be interpreted as the value function for an optimal control problem. Consider the dynamics

$$\dot{x} = g(x,s) + u, \quad x(0) = x_0. \tag{5.9}$$

We wish to minimise

$$I_t(x_0,u) = S_0(x_0) + \int_0^t \left(\frac{1}{2} \mid u(s) \mid^2 + V(x_u(s),s)\right) ds. \tag{5.10}$$

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

$$S(x,t) = \inf_{(x_0,u) \in \mathcal{F}_{x,t}} I_t(x_0,u). \tag{5.11}$$

The arguments in Chapter 4 can be used to prove the following.

Theorem 5.1 *The value function* $S(x,t)$ *defined by* (5.11) *is the unique viscosity solution of the Hamilton–Jacobi equation* (5.6). *Further, if* $K \subset \Omega_0$ *and* $Q \subset I\!R^n$ *are compact, then there exists* $C > 0$ *such that*

$$\sup_{(x,t)\in Q\times[0,T]} \mid S(x,t)[\eta] - S(x,t)[\zeta] \mid \leq C \parallel \eta - \zeta \parallel$$

for all $\eta,\ \zeta \in K$.

Our main task is to prove that $S^\epsilon \to S$ as $\epsilon \to 0$ uniformly on compact subsets. From this the asymptotic formula (5.2) will follow (Theorem 5.3). Before proceeding, we define an extension of the deterministic estimator to all of Ω_0 as follows:

For $Y \in \Omega_0$, define

$$W(x,t)[Y] = S(x,t)[Y] - y(t)h(x). \tag{5.12}$$

If $W(\cdot,t)$ has a unique minimum for each $0 \leq t \leq T$ and each $Y \in \Omega_0$, we say as before that the deterministic estimate

$$\hat{x}(t)[Y] = \underset{x \in I\!R^n}{\operatorname{argmin}} W(x,t)[Y]$$

is well defined. Note that if $Y \in \Omega_0^{C^1}$, then $W[Y]$ satisfies the HJB equation (4.7). This defines a continuous map

$$\begin{aligned} \bar{\pi}_t : \Omega_{0,t} &\rightarrow \mathcal{P}(I\!R^n) \\ Y_t &\mapsto \delta_{\hat{x}(t)[Y_t]} \end{aligned} \tag{5.13}$$

which extends (4.10).

5.1.2 Some Estimates

Let $S^\epsilon(x,t)$ be the solution of (5.4). In this section we obtain estimates for $\mid S^\epsilon \mid$, $\mid DS^\epsilon \mid$ and for the Hölder norm of S^ϵ in $t \in [0,T]$ on compact subsets independent of the parameter ϵ. These estimates will be used in Section 5.1.3 to prove that $S^\epsilon \to S$.

Lemma 5.1 *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*

$$| S^\epsilon(x,t) | \leq K, \quad \text{for all } (x,t) \in Q, \tag{5.14}$$

$$| DS^\epsilon(x,t) | \leq K, \quad \text{for all } (x,t) \in Q. \tag{5.15}$$

To prove (5.14), 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 *(Maximum Principle,* Friedman [23]*) Define*

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

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 5.2 *(Comparison Theorem) Let S^ϵ be a solution of* (5.4)*, and define*

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

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

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

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

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

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 the maximum principle. ∎

Proof of Lemma 5.1: 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 [16]). Define

$$v(x,t) \;=\; \frac{1}{R^2 - \mid x \mid^2} + \mu t + M, \tag{5.16}$$

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

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

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

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 $\mid x \mid \to R$ 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

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

for all sufficiently small $\epsilon > 0$.

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

2. Next we estimate the gradient, using a variant of the techniques used in Evans and Ishii [16], as suggested to us by L. C. Evans. To simplify the notation we write $v = S^\epsilon$, which from (5.4) satisfies

$$v_t - \frac{\epsilon}{2}v_{ii} + \frac{1}{2}v_iv_i + v_ig^i - V^\epsilon \;=\; 0, \tag{5.17}$$

where we have used the summation convention. Let $Q \subset\subset Q' \subset\subset I\!R^n \times (0,T)$, where Q, Q' are open and "$\subset\subset$" means "compactly contained in". Choose $\zeta \in C_c^\infty(I\!R^n \times [0,T])$ such that $0 \leq \zeta \leq 1$, $\zeta \equiv 1$ on Q and $\zeta \equiv 0$ near $\partial Q'$, and define

$$z \;=\; \zeta^2 v_k v_k - \lambda v \tag{5.18}$$

where the constant $\lambda \geq 1$ is to be chosen.

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

$$z_i \;=\; 0 \text{ and} \tag{5.19}$$

$$0 \;\leq\; z_t - \frac{\epsilon}{2}z_{ii} \tag{5.20}$$

at the point (x_0,t_0). Then at this point, using (5.20),

$$\begin{aligned} 0 \;\leq\; & 2\zeta\zeta_t v_kv_k + 2\zeta^2 v_kv_{kt} - \lambda v_t \\ & -\epsilon\zeta_i\zeta_i v_kv_k - \epsilon\zeta\zeta_{ii}v_kv_k - 4\epsilon\zeta\zeta_i v_kv_{ki} \\ & -\epsilon\zeta^2 v_{ki}v_{ki} - \epsilon\zeta^2 v_kv_{kii} + \frac{\epsilon}{2}\lambda v_{ii} \end{aligned}$$

$$
\begin{aligned}
&\leq -\epsilon C\zeta^2 \mid D^2 v \mid^2 + 2\zeta^2 v_k \left(v_t - \frac{\epsilon}{2} v_{ii}\right)_k \\
&\quad + \lambda \left(-v_t + \frac{\epsilon}{2} v_{ii}\right) + C \mid Dv \mid^2
\end{aligned}
$$

for ϵ sufficiently small. Using (5.17) we find that

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

This together with (5.19) implies

$$
\frac{\lambda}{2} \mid Dv \mid^2 \leq C\zeta \mid Dv \mid^3 + C \mid Dv \mid^2 + \lambda C \mid Dv \mid + \lambda C. \tag{5.21}
$$

Let

$$
\lambda = \mu[\max(\zeta \mid Dv \mid) + 1], \tag{5.22}
$$

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

$$
\zeta \mid Dv \mid^3 \leq \mid Dv \mid^2 [\max(\zeta \mid Dv \mid) + 1],
$$

and thus from (5.21),

$$
\begin{aligned}
\frac{\mu}{2} \mid Dv \mid^2 &\leq C \mid Dv \mid^2 + [\max(\zeta \mid Dv \mid) + 1]^{-1} C \mid Dv \mid^2 + \mu C \mid Dv \mid + C\mu \\
&\leq C \mid Dv \mid^2 + \mu C[\mid Dv \mid + 1] \qquad (5.23)
\end{aligned}
$$

using the fact that $[\max(\zeta \mid Dv \mid) + 1]^{-1} \leq 1$. Choosing μ so large that $\mu/4 \leq \mu/2 - C$, we have from (5.23)

$$
\mid Dv \mid^2 \leq C[\mid Dv \mid + 1] \quad \text{at } (x_0, t_0).
$$

Then, because $\mu \geq 1$ and $0 \leq \zeta \leq 1$, at (x_0, t_0) we have

$$
\begin{aligned}
z &= \zeta^2 \mid Dv \mid^2 - \lambda v \\
&\leq \zeta C[\zeta \mid Dv \mid + 1] + \lambda C \\
&\leq \lambda C. \qquad (5.24)
\end{aligned}
$$

This implies

$$z \leq C\lambda \text{ in } Q'. \tag{5.25}$$

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 (5.25). But from (5.25),

$$\max\left(\zeta^2 \mid Dv \mid^2\right) \leq \max z + C\lambda \leq C\lambda,$$

and using the definition (5.22) we have

$$\max\left(\zeta^2 \mid Dv \mid^2\right) \leq C\mu[\max(\zeta \mid Dv \mid) + 1]$$

which implies

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

and hence

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

∎

The following Hölder estimate is a modification of Lemma 5.2 of Crandall and Lions [7].

Lemma 5.3 *For every compact subset $Q \subset I\!R^n$, there exists $\epsilon_0 > 0$ and $K > 0$ such that for $0 < \epsilon < \epsilon_0$ we have*

$$\mid S^\epsilon(x,t) - S^\epsilon(x,s) \mid \leq K\left(\sqrt{\epsilon} \mid t-s \mid^{1/2} + \mid t-s \mid\right) \tag{5.26}$$

for all $x \in Q;\ t, s \in [0,T]$.

Proof: Let $Q \subset\subset Q' \subset\subset \mathbb{R}^n$, and choose $\zeta \in C_c^\infty(\mathbb{R}^n)$ such that $\zeta \equiv 1$ on Q and $\zeta \equiv 0$ near $\partial Q'$. Choose $\epsilon_0 > 0$ such that $\text{dist}(\partial Q,\ \partial Q') > \sqrt{\epsilon_0 T}$ and so that (5.14), (5.15) hold on $Q \times [0,T]$ for $0 < \epsilon \le \epsilon_0$. Write

$$v \;=\; \zeta S^\epsilon.$$

Then

$$\begin{aligned} \mid v_t - \frac{\epsilon}{2} v_{ii} \mid \;&=\; \mid \zeta \left(-\frac{1}{2} S_i^\epsilon S_i^\epsilon - g^i S_i^\epsilon + V^\epsilon\right) + \zeta_t S^\epsilon - \epsilon \zeta_i S_i^\epsilon - \frac{\epsilon}{2} \zeta_{ii} S^\epsilon \mid \\ &\le\; K_0 \;\text{ on } \mathbb{R}^n \times [0,T]. \end{aligned}$$

Let ρ_α be a standard mollifier for $0 < \alpha \le \sqrt{\epsilon_0 T}$, and set $v^\alpha = \rho_\alpha * v$. Then as in [7], for $0 \le t \le T$,

$$\begin{aligned} \| v_t^\alpha \|_{L^\infty(\mathbb{R}^n)} \;&\le\; K_0 + \epsilon \| v_{ii}^\alpha \|_{L^\infty(\mathbb{R}^n)} \\ &\le\; K_0 + \frac{\epsilon C}{\alpha} \| Dv \|_{L^\infty(\mathbb{R}^n)} . \end{aligned}$$

Next, for $x \in \mathbb{R}^N,\ t, s \in [0,T]$,

$$\begin{aligned} \mid v(x,t) - v(x,s) \mid \;&\le\; \mid v(x,t) - v^\alpha(x,t) \mid + \mid v^\alpha(x,t) - v^\alpha(x,s) \mid \\ &\quad + \mid v^\alpha(x,s) - v(x,s) \mid \\ &\le\; \alpha K + \left(K_0 + \frac{\epsilon C}{\alpha}\right) \mid t - s \mid \\ &\le\; K\left(\epsilon^{1/2} \mid t - s \mid^{1/2} + \mid t - s \mid\right), \end{aligned}$$

on our setting $\alpha = \epsilon^{1/2} \mid t - s \mid^{1/2}$.

Finally note that $v = S^\epsilon$ on Q. ∎

5.1.3 Asymptotic Result

We first present an asymptotic result for the robust Zakai equation for a fixed observation record.

Theorem 5.2 *Fix $Y \in \Omega_0$. Let $p^\epsilon, S^\epsilon, S$ denote the corresponding solutions of* (2.31), (5.4), (5.6). *Then under the above assumptions, we have*

$$-\lim_{\epsilon \to 0} \epsilon \log p^\epsilon(x,t) \;=\; \lim_{\epsilon \to 0} S^\epsilon(x,t) \;=\; S(x,t) \tag{5.27}$$

uniformly on compact subsets of $\mathbb{R}^n \times [0,T]$.

Proof: Lemmas 5.1 and 5.3 imply that S^ϵ are uniformly bounded and equicontinuous on compact subsets. By 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 [6], $\tilde{S}$ is a viscosity solution of (5.6). By uniqueness, $\tilde{S} = S$. In fact, $S^\epsilon \to S$ as $\epsilon \to 0$. ∎

Recall that the solutions X^ϵ, Y^ϵ of (2.2), (2.3) converge in probability uniformly on $[0,T]$ to the solution of the corresponding deterministic system, and $Z \in \Omega_0$ denotes the limit of the process Y^ϵ. We write $S^\epsilon[Y]$ for the solution of (5.4) for $Y \in \Omega_0$, and similarly $S[Y]$.

Let $q^\epsilon(x,t)$ denote the solution of the Zakai equation (2.27). Then by (2.29), (5.3):

$$-\epsilon \log q^\epsilon(x,t) \;=\; S^\epsilon[Y^\epsilon](x,t) - y^\epsilon(t)h(x). \tag{5.28}$$

As in (5.12), define

$$W(x,t) \;=\; S[Z](x,t) - z(t)h(x). \tag{5.29}$$

Note that Z is continuously differentiable and so W is the value function for deterministic estimation with observation record $Z_t = \{z(s), 0 \le s \le t\} \in \Omega_0^{C^1}$.

Lemma 5.4 *Fix $Z \in \Omega_0$, and let Q be a compact subset of $\mathbb{R}^n$. Then for all*

$\beta > 0$ *there exists* $\gamma > 0$, $\epsilon_0 > 0$ *such that if* $\| \eta - Z \| \leq \gamma$, $0 < \epsilon \leq \epsilon_0$, *we have*

$$\sup_{x \in Q,\ 0 \leq t \leq T} | S^{\epsilon}[\eta](x,t) - S^{\epsilon}[Z](x,t) | \leq \beta. \tag{5.30}$$

Proof: Let $\| \cdot \|_Q$ denote the supremum norm on $Q \times [0,T]$. Suppose the assertion is false. Then there exists $\beta > 0$, a subsequence $\epsilon_j \to 0$, sequences $\gamma_j \to 0$, $\eta_j \in \Omega_0$ with $\| \eta_j - Z \| \leq \gamma_j$ such that

$$\| S^{\epsilon_j}[\eta_j] - S^{\epsilon_j}[Z] \|_Q \geq \beta > 0, \tag{5.31}$$

for $j = 1, 2, \ldots$ Note that $\eta_j \to Z$ uniformly on $[0,T]$ as $j \to \infty$.

An inspection of the proofs of Lemmas 5.1, 5.3 reveals that the sequence $\{S^{\epsilon_j}[\eta_j]\}_{j \geq 1}$ is uniformly bounded and equicontinuous on $Q \times [0,T]$. These proofs did not require differentiation of the coefficients with respect to t, and so the estimates can be made uniform in the parameter η over a compact subset in Ω_0. Thus, as in the proof of Theorem 5.2, there is a subsequence j_k such that

$$\lim_{k \to \infty} S^{\epsilon_{j_k}}[\eta_{j_k}] = S[Z] = \lim_{k \to \infty} S^{\epsilon_{j_k}}[Z].$$

Letting $j = j_k \to \infty$ in (5.31) yields a contradiction. ∎

Corollary 5.1 *Let* $K \subset \Omega_0$ *and* $Q \subset \mathbb{R}^n$ *be compact. Then for all* $\beta > 0$ *there exists* $\epsilon_0 > 0$ *such that if* $0 < \epsilon \leq \epsilon_0$ *then*

$$\sup_{(x,t) \in Q \times [0,T]} | S^{\epsilon}(x,t)[\eta] - S(x,t)[\eta] | < \beta$$

for all $\eta \in K$.

Proof: Using the fact that $S[\eta]$ depends continuously on $\eta \in \Omega_0$, we can argue by contradiction as above. ∎

Theorem 5.3 *Under the above assumptions, we have:*

$$\lim_{\epsilon \to 0} \epsilon \log q^\epsilon(x,t) = -W(x,t) \tag{5.32}$$

in probability uniformly on compact subsets of $\mathbb{R}^n \times [0,T]$.

Proof: Since $Y^\epsilon \to Z$ in probability, in view of (5.28) it is enough to show that $\rho(S^\epsilon[Y^\epsilon], S[Z]) \to 0$ in probability, where ρ denotes a metric on $C(\mathbb{R}^n \times [0,T]; \mathbb{R})$ corresponding to uniform convergence on compact subsets.

Fix $\delta > 0, \beta > 0$. From Lemma 5.4, we can choose $\gamma > 0, \epsilon_1 > 0$ such that

$$\rho(S^\epsilon[Y], S^\epsilon[Z]) < \beta/2 \tag{5.33}$$

for all $0 < \epsilon < \epsilon_1$ and all $\| Y - Z \| \leq \gamma$.

Choose $\epsilon_2 > 0$ such that

$$P(\| Y^\epsilon - Z \| > \gamma) < \delta \tag{5.34}$$

for all $0 < \epsilon < \epsilon_2$. From Theorem 5.2, choose $\epsilon_3 > 0$ such that

$$\rho(S^\epsilon[Z], S[Z]) < \beta/2 \tag{5.35}$$

for all $0 < \epsilon < \epsilon_3$. Set $0 < \epsilon_0 < \epsilon_1 \wedge \epsilon_2 \wedge \epsilon_3$.

Then if $0 < \epsilon \leq \epsilon_0$,

$$\begin{aligned} P(\rho(S^\epsilon[Y^\epsilon], S[Z]) > \beta) &\leq P(\rho(S^\epsilon[Y^\epsilon], S^\epsilon[Z]) > \beta/2; \ \| Y^\epsilon - Z \| > \gamma) \\ &\leq \delta. \end{aligned}$$

∎

5.2 Probabilistic Method

Once again we are interested in the filtering problem (2.2), (2.3), and we make the same assumptions as above.

Let σ_t^ϵ be the solution of the weak form of the Zakai equation (2.26). Complementing formula (5.2), we prove in this section

$$\sigma_t^\epsilon(A) \asymp e^{-\frac{1}{\epsilon}\inf_{x\in A} W(x,t)} \tag{5.36}$$

in probability as $\epsilon \to 0$, where $W(x,t)$ is defined by (5.29). The proof uses the representation formula (2.28), rather than the results of the previous section.

5.2.1 Varadhan's Theorem "in Probability"

In order to use the representation formula (2.28) to prove (5.36), we need an extension of Varadhan's Theorem to random functions F^ϵ. Recall the notation and terminology used in Section 2.2. We follow the development in Varadhan [55].

Theorem 5.4 *Let $\{\mathsf{P}^\epsilon\}$ obey the LDP with action function I. Let $F^\epsilon : \mathsf{X} \times \Omega \to I\!R$ be a family of random functions such that for each $\omega \in \Omega$, $F^\epsilon(\cdot,\omega) : \mathsf{X} \to I\!R$ is continuous. Let $F : \mathsf{X} \to I\!R$ be bounded and continuous. Assume that for all $\alpha > 0, \beta > 0$ there exists $\epsilon_1 > 0$ such that $0 < \epsilon \le \epsilon_1$ implies*

$$P\left(\sup_{\mathsf{x}\in\mathsf{X}} \mid F^\epsilon(\mathsf{x}) - F(\mathsf{x}) \mid \le \alpha\right) \;\ge\; 1-\beta. \tag{5.37}$$

Then if $C \subset \mathsf{X}$ is closed, $G \subset \mathsf{X}$ is open, we have: for all $\gamma > 0$, $\delta > 0$ there exists $\epsilon_0 > 0$ such that $0 < \epsilon \le \epsilon_0$ implies

$$P\left(\epsilon \log \int_C \exp\left[-\frac{1}{\epsilon}F^\epsilon(\mathsf{x})\right] \mathsf{P}^\epsilon(d\mathsf{x}) \;\le\; -\inf_{\mathsf{x}\in C}[I(\mathsf{x}) + F(\mathsf{x})] + \delta\right) \;\ge\; 1-\gamma, \tag{5.38}$$

$$P\left(\epsilon \log \int_G \exp\left[-\frac{1}{\epsilon}F^\epsilon(\mathsf{x})\right] \mathsf{P}^\epsilon(d\mathsf{x}) \;\ge\; -\inf_{\mathsf{x}\in G}[I(\mathsf{x}) + F(\mathsf{x})] - \delta\right) \;\ge\; 1-\gamma. \tag{5.39}$$

To prove this theorem, we use the following lemma. This lemma is more general than required.

Lemma 5.5 *Let F^ϵ be a family of random functions on* X. *Let F be a lower semicontinuous function on* X. *Assume:*

(i) *If $I(\mathsf{x}) < \infty$, then there exists $\epsilon' > 0$, $r > 0$ such that if $N_\mathsf{x} = \{\mathsf{y} \in \mathsf{X} : d(\mathsf{x},\mathsf{y}) < r\}$ and $0 < \epsilon \le \epsilon'$ then*

$$P\left(\sup_{\mathsf{y}\in N_\mathsf{x}} (-F^\epsilon(\mathsf{y})) \le -F(\mathsf{x}) + \alpha\right) \ \ge \ 1 - \beta. \tag{5.40}$$

(ii) *For all $\rho > 0$ there is an $L \ge 0$, $\epsilon'' > 0$ such that $0 < \epsilon \le \epsilon''$ implies*

$$P\left(\inf_{\mathsf{x}\in\mathsf{X}} F^\epsilon(\mathsf{x}) \ge -L\right) \ \ge \ 1 - \rho. \tag{5.41}$$

Then for all $\gamma > 0$, $\delta > 0$ there exists $\epsilon_0 > 0$ such that $0 < \epsilon \le \epsilon_0$ implies

$$P\left(\epsilon \log \int_\mathsf{X} \exp\left[-\frac{1}{\epsilon}F^\epsilon(\mathsf{x})\right] \mathsf{P}^\epsilon(d\mathsf{x}) \ \le \ -\inf_{\mathsf{x}\in\mathsf{X}}[I(\mathsf{x}) + F(\mathsf{x})] + \delta\right) \ \ge \ 1 - \gamma. \tag{5.42}$$

Proof: Choose $\gamma > 0$, $\delta > 0$; $0 < \rho < \gamma/2$. From hypothesis (ii), let L, ϵ'' be such that (5.41) holds. Write

$$\Lambda \ = \ \inf_{\mathsf{x}\in\mathsf{X}}[I(\mathsf{x}) + F(\mathsf{x})].$$

Choose $M \ge 0$ such that $L - M \le -\Lambda$, and let $K = \{\mathsf{x} \in \mathsf{X} : I(\mathsf{x}) \le M\}$. Let $\{\mathsf{x}_i\}_{i=1}^\infty$ be a dense subset of K. Set $\alpha = \delta/5$, $\beta = \gamma/2$.

By hypothesis (i), for each $i = 1, 2, \ldots$, there exists ϵ_i', r_i, $N_i = \{\mathsf{y} : d(\mathsf{x}_i, \mathsf{y}) < r_i\}$, such that if $0 < \epsilon \le \epsilon_i'$ then

$$P(A_{i,\epsilon}) \ \ge \ 1 - \beta/2^i$$

for

$$A_{i,\epsilon} = \left\{ \sup_{y \in N_i} (-F^\epsilon(y)) \le -F(x_i) + \alpha \right\}.$$

Since I is lower semicontinuous, reducing r_i if necessary, if $y \in N_i$ then

$$I(y) \ge I(x_i) - \alpha.$$

Let $G_i = \{y : d(x_i, y) < r_i/2\}$. Since $\{x_i\}$ is dense in K, and since K is compact, $K \subset \bigcup_{i=1}^N G_i \equiv G$ for some positive integer N. Fix $i \in \{1, \ldots, N\}$.

Now $\{P^\epsilon\}$ obey the LDP with action function I, so there is an $0 < \epsilon_i \le \epsilon_i'$ such that $0 < \epsilon \le \epsilon_i$ implies

$$\begin{aligned} \epsilon \log P^\epsilon(\bar{G}_i) &\le - \inf_{y \in \bar{G}_i} I(y) + \alpha \\ &\le -I(x_i) + 2\alpha. \end{aligned}$$

Then on $A_{i,\epsilon}$, if $0 < \epsilon \le \epsilon_i$,

$$\begin{aligned} Q^\epsilon(G_i) &= \int_{G_i} \exp\left[-\frac{1}{\epsilon} F^\epsilon(x)\right] P^\epsilon(dx) \\ &\le \exp\left[-\frac{1}{\epsilon} \inf_{y \in G_i} F^\epsilon(y)\right] P^\epsilon(G_i) \\ &\le \exp\left[-\frac{1}{\epsilon}(F^\epsilon(x_i) - \alpha)\right] P^\epsilon(\bar{G}_i), \end{aligned}$$

and hence

$$\begin{aligned} \epsilon \log Q^\epsilon(G_i) &\le -F(x_i) + \alpha + \epsilon \log P^\epsilon(\bar{G}_i) \\ &\le -\Lambda + 3\alpha. \end{aligned}$$

Let $A_{K,\epsilon} = \bigcap_{i=1}^N A_{i,\epsilon}$. Then there exists $\epsilon_K > 0$ such that if $0 < \epsilon \le \epsilon_K$, then

$$P(A_{K,\epsilon}) \ge 1 - \sum_{i=1}^N \beta/2^i \ge 1 - \gamma/2,$$

and $\epsilon \log N \le \alpha$. So on $A_{K,\epsilon}$, if $0 < \epsilon \le \epsilon_K$,

$$\begin{aligned} \epsilon \log Q^\epsilon(G) &\le \epsilon \log N + \max_{i=1,\ldots,N} \epsilon \log Q^\epsilon(G_i) \\ &\le -\Lambda + 4\alpha. \end{aligned}$$

Let

$$B_{K,\epsilon} \;=\; A_{K,\epsilon} \cap \left\{ \inf_{\mathsf{x}\in\mathsf{X}} F^\epsilon(\mathsf{x}) \geq -L \right\}.$$

Then there exists $\epsilon_0 > 0$ such that $0 < \epsilon \leq \epsilon_0$ implies

$$\begin{aligned} P(B_{K,\epsilon}) \;&\geq\; 1 - \gamma/2 - \rho \\ &\geq\; 1 - \gamma, \end{aligned}$$

and $\epsilon \log 2 \leq \alpha$. Thus on $B_{K,\epsilon}$, if $0 < \epsilon \leq \epsilon_0$,

$$\begin{aligned} \epsilon \log \mathsf{Q}^\epsilon(\mathsf{X}) \;&\leq\; \epsilon \log 2 + \max\left(\epsilon \log \mathsf{Q}^\epsilon(G),\ \epsilon \log \mathsf{Q}^\epsilon(G^c)\right) \\ &\leq\; \alpha + \max\left(-\Lambda + 4\alpha,\ L - M\right) \\ &\leq\; -\Lambda + 5\alpha. \end{aligned}$$

Consequently

$$\begin{aligned} P\left(\epsilon \log \mathsf{Q}^\epsilon(\mathsf{X}) \leq -\Lambda + \delta\right) \;&\geq\; P(B_{K,\epsilon}) \\ &\geq\; 1 - \gamma \end{aligned}$$

as required. ∎

Proof of Theorem 5.4: *Upper bound.* Let $C \subset \mathsf{X}$ be closed. Define $\bar{F}^\epsilon$, $\bar{F}$ to equal F^ϵ, F on C respectively, and to equal $+\infty$ off C. Then $\bar{F}^\epsilon$, $\bar{F}$ are lower semicontinuous functions.

Let $\mathsf{x} \in C$. Then (5.37) and the continuity of $\bar{F}$ on C imply (5.40). If $\mathsf{x} \notin C$, (5.40) is obvious. Since $\bar{F}^\epsilon \geq F^\epsilon$ on X, (5.37) and the boundedness of F imply (5.41). Then applying Lemma 5.5 we obtain (5.38).

Lower bound. Let $G \subset \mathsf{X}$ be open. Assume

$$\Lambda_G \;=\; \inf_{\mathsf{x}\in G}[I(\mathsf{x}) + F(\mathsf{x})] \;<\; \infty.$$

Set $\alpha = \delta/3$, $\beta = \gamma$. Choose $\mathsf{z} \in G$ such that

$$I(\mathsf{z}) + F(\mathsf{z}) \;\le\; \Lambda_G + \alpha.$$

Our hypotheses imply there exist $\epsilon_0 > 0$, $r > 0$ such that if $N_\mathsf{z} = \{\mathsf{y} : d(\mathsf{z}, \mathsf{y}) < r\}$ and $0 < \epsilon \le \epsilon_0$, then

$$P(A_{\mathsf{z},\epsilon}) \;\ge\; 1 - \beta$$

for

$$A_{\mathsf{z},\epsilon} \;=\; \left\{ \sup_{\mathsf{y} \in N_\mathsf{z}} (-F^\epsilon(\mathsf{y})) \ge -F(\mathsf{z}) - \alpha \right\}.$$

Reducing ϵ_0 if necessary,

$$\begin{aligned} \epsilon \log \mathsf{P}^\epsilon(N_\mathsf{z}) \;&\ge\; -\inf_{\mathsf{x} \in N_\mathsf{z}} I(\mathsf{x}) - \alpha \\ &\ge\; -I(\mathsf{z}) - 2\alpha \end{aligned}$$

provided $0 < \epsilon \le \epsilon_0$.

Then on $A_{\mathsf{z},\epsilon}$, if $0 < \epsilon \le \epsilon_0$,

$$\begin{aligned} \mathsf{Q}^\epsilon(G) \;&=\; \int_G \exp\left[-\frac{1}{\epsilon} F^\epsilon(\mathsf{x})\right] \mathsf{P}^\epsilon(d\mathsf{x}) \\ &\ge\; \exp\left[-\frac{1}{\epsilon}(F(\mathsf{z}) + \alpha)\right] \mathsf{P}^\epsilon(N_\mathsf{z}), \end{aligned}$$

and hence

$$\begin{aligned} \epsilon \log \mathsf{Q}^\epsilon(G) \;&\ge\; -F(\mathsf{z}) - \alpha + \epsilon \log \mathsf{P}^\epsilon(N_\mathsf{z}) \\ &\ge\; -[I(\mathsf{z}) + F(\mathsf{z})] - 2\alpha \\ &\ge\; -\Lambda_G - 3\alpha. \end{aligned}$$

Hence

$$\begin{aligned} P\left(\epsilon \log \mathsf{Q}^\epsilon(G) \ge -\Lambda_G - \delta\right) \;&\ge\; P(A_{\mathsf{z},\epsilon}) \\ &\ge\; 1 - \gamma \end{aligned}$$

as required. ∎

5.2.2 Asymptotic Result

Recall that $Y^\epsilon = \{y_s^\epsilon : 0 \le s \le T\}$ converges in probability to $Z = \{z_s : 0 \le s \le T\}$. Fix $0 \le t \le T$, define the random functions

$$F^\epsilon(\theta, \omega) = \frac{1}{2}\int_0^t h(\theta_s)^2 ds - \int_0^t h(\theta_s) dy_s^\epsilon(\omega), \quad (\theta \in \Omega^n,\ \omega \in \Omega) \tag{5.43}$$

and the function

$$F(\theta) = \frac{1}{2}\int_0^t h(\theta_s)^2 ds - \int_0^t h(\theta_s)\dot{z}_s ds. \tag{5.44}$$

Lemma 5.6 *For all $\alpha > 0$, $\beta > 0$ there exists $\epsilon_1 > 0$ such that $0 < \epsilon \le \epsilon_1$ implies*

$$P\left(\sup_{\theta \in \Omega^n} \mid F^\epsilon(\theta) - F(\theta) \mid > \alpha\right) < \beta. \tag{5.45}$$

Proof: From (2.3),

$$\int_0^t h(\theta_s) dy_s^\epsilon = \int_0^t h(\theta_s) h(x_s^\epsilon) ds + \sqrt{\epsilon}\int_0^t h(\theta_s) dv_s.$$

Then using the Chebeshev inequality,

$$\begin{aligned} &P\left(\sup_{\theta\in\Omega^n} \mid \textstyle\int_0^t h(\theta_s)h(x_s^\epsilon)ds - \int_0^t h(\theta_s)\dot{z}_s ds \mid > \alpha\right) \\ &\le \tfrac{1}{\alpha^2} E \sup_{\theta\in\Omega^n} \mid \textstyle\int_0^t h(\theta_s)[h(x_s^\epsilon) - h(x_s)]ds \mid^2 \\ &\le \tfrac{1}{\alpha^2} \parallel h \parallel^2 \mathrm{Lip}(h)^2 T\epsilon C, \end{aligned}$$

where the last inequality follows from Lemma 2.1.

Again using the Chebeshev inequality,

$$P\left(\sup_{\theta\in\Omega^n} \sqrt{\epsilon} \mid \int_0^t h(\theta_s) dv_s \mid > \alpha\right) \le \frac{\epsilon}{\alpha^2} E \sup_{\theta\in\Omega^n} \mid \int_0^t h(\theta_s) dv_s \mid^2 . \tag{5.46}$$

Choose a sequence of random variables θ^i so that

$$\sup_{\theta\in\Omega^n} \mid \int_0^t h(\theta_s)dv_s \mid^2 = \lim_{i\to\infty} \mid \int_0^t h(\theta_s^i)dv_s \mid^2 \leq \infty \quad \text{a.s.}$$

Then the RHS of (5.46) is bounded by

$$\frac{\epsilon}{\alpha^2} \liminf_{i\to\infty} E \int_0^t h(\theta_s^i)^2 ds \;\leq\; \frac{\epsilon}{\alpha^2} T \parallel h \parallel^2 .$$

∎

Let $W(x,t)$ be Mortensen's value function for the observation path Z, defined by (5.29).

Theorem 5.5 *Under the above assumptions, for each closed set $C \subset I\!R^n$ and each open set $G \subset I\!R^n$, we have:*

$$\limsup_{\epsilon\to 0} \epsilon \log \sigma_t^\epsilon(C) \;\leq\; -\inf_{x\in C} W(x,t) \;\text{ in probability,} \tag{5.47}$$

$$\liminf_{\epsilon\to 0} \epsilon \log \sigma_t^\epsilon(G) \;\geq\; -\inf_{x\in G} W(x,t) \;\text{ in probability.} \tag{5.48}$$

Proof: Let p_t denote the cannonical projection $\Omega^n \to I\!R^n$, for $0 \leq t \leq T$. Let $A \in \mathcal{B}(I\!R^n)$. Then from (2.8)

$$\begin{aligned} \sigma_t^\epsilon(A)(\omega) &= \Sigma^\epsilon\left(\mathsf{p}_t^{-1}A\right) \\ &= \int_{\mathsf{p}_t^{-1}A} \exp\left[-\frac{1}{\epsilon}F^\epsilon(\theta,\omega)\right] P_X^\epsilon(d\theta). \end{aligned} \tag{5.49}$$

Recall that $\{P_X^\epsilon\}$ obey the LDP with action function I_X defined by (2.38). Now

$$\begin{aligned} \inf_{\theta\in\mathsf{p}_t^{-1}A} [I_X(\theta)+F(\theta)] &= \inf_{x\in A} \inf_{\theta:\theta_t=x} [I_X(\theta)+F(\theta)] \\ &= \inf_{x\in A} W(x,t). \end{aligned}$$

Then apply Theorem 5.4 to (5.49) for the sets $\mathsf{p}_t^{-1}C$, $\mathsf{p}_t^{-1}G$. ∎

5.3 Large Deviations

We have seen that the optimal control problem associated with deterministic minimum energy estimation plays a key role in studying the asymptotics of the Zakai equation. In this section we shall see that this control problem is exactly the variational problem governing a large deviation principle for certain conditional measures. In addition we study the weak limit of the conditional measures, and identify the limiting filter.

5.3.1 Hijab's Result

We present Hijab's [26] "pathwise" LDP for certain conditional measures, employing our notation, etc. Hijab proved his result for the case that (2.2) has fixed initial conditions $x_0^\epsilon = x_0$, so the result we state is a slight extension of his; here x_0^ϵ is random with density $p_0^\epsilon(x)$ defined by (2.1).

Recall the measures $\check{\Sigma}^\epsilon$, $\check{\Pi}^\epsilon$ defined by (2.17), (2.18) respectively, parameterised by $\eta \in \Omega_0$. For each $\eta \in \Omega_0$, the measures $\{\check{P}^\epsilon_{[\eta]}\}$ obey the LDP with action function

$$\check{I}(\theta)[\eta] \;=\; \inf_u \left\{ S_0(\check{x}_u(0)) + \frac{1}{2}\int_0^T \mid u(t) \mid^2 dt \;:\; \check{x}_u = \theta \right\} \tag{5.50}$$

where here $\check{x}_u$ denotes the solution of (5.9) with initial condition $\check{x}_u(0) = \theta(0)$. Define

$$\begin{aligned} F^\epsilon(\theta,\eta) &= -\eta_T h(\theta_T) + \int_0^T V^\epsilon(\theta_s,\eta_s)ds \quad (\theta \in \Omega^n,\ \eta \in \Omega_0) & (5.51)\\ F(\theta,\eta) &= -\eta_T h(\theta_T) + \int_0^T V(\theta_s,\eta_s)ds & (5.52)\end{aligned}$$

where V^ϵ is defined by (2.13) and

$$V(x,y) \;=\; \frac{1}{2}h(x)^2 + yDh(x)f(x) - \frac{1}{2}y^2 \mid Dh(x) \mid^2 \quad (x \in \mathbb{R}^n,\ y \in \mathbb{R}). \tag{5.53}$$

Next define

$$\begin{aligned}\check{J}(\theta)[\eta] &= \check{I}(\theta)[\eta] + F(\theta,\eta) \qquad (5.54)\\ &= \inf_u \left\{ S_0(\check{x}_u(0)) + \int_0^T \frac{1}{2} \mid u(t) \mid^2 + V(\check{x}_u(t),\eta_t)dt - \eta_T h(\check{x}_u(T)) \ : \ \check{x}_u = \theta \right\};\end{aligned}$$

compare with (5.10), (5.11). If $\eta \in \Omega_0^{H^1}$, this expression reduces to

$$\inf_u \left\{ S_0(x_u(0)) + \int_0^T \frac{1}{2} \mid u(t) \mid^2 + \frac{1}{2}h(x_u(t))^2 - \dot{\eta}_t h(x_u(t))dt \ : \ x_u = \theta \right\};$$

where x_u is the solution of (2.36). Compare with (4.3), (4.6), (4.11). Write

$$J(\theta)[\eta] = \check{J}(\theta)[\eta] - \inf_{\theta' \in \Omega^n} \check{J}(\theta')[\eta]. \qquad (5.55)$$

Theorem *(Hijab [26]) For each $\eta \in \Omega_0$, $\{\bar{\Pi}^\epsilon[\eta]\}$ obey the LDP with action function $J(\theta)[\eta]$.*

Proof: Apply Varadhan's Theorem to

$$\bar{\Sigma}^\epsilon(A)[\eta] = \int_A \exp\left[-\frac{1}{\epsilon}F^\epsilon(\theta,\eta)\right] \check{P}^\epsilon_{[\eta]}(d\theta) \quad (A \in \mathcal{B}(\Omega^n)).$$

∎

5.3.2 A LDP "in Probability" for Conditional Measures

We now prove large deviation results for the conditional measures Π^ϵ, π^ϵ_t defined by (2.7), (2.9) respectively. Recall Definition 2.1. Let F^ϵ, F be given by (5.43), (5.44) respectively and let I_X be the action function for $\{P^\epsilon_X\}$ defined by (2.38). Define

$$\begin{aligned}\check{J}(\theta) &= I_X(\theta) + F(\theta) \qquad (5.56)\\ &= \inf_u \left\{ S_0(x_u(0)) + \int_0^T \frac{1}{2} \mid u(t) \mid^2 + \frac{1}{2}h(x_u(t))^2 - \dot{z}_t h(x_u(t))dt \ : \ x_u = \theta \right\}\end{aligned}$$

where x_u is the solution of (2.36). Set

$$J(\theta) = \breve{J}(\theta) - \inf_{\theta' \in \Omega^n} \breve{J}(\theta'), \tag{5.57}$$
$$= J(\theta)[Z]$$

If $\theta \in \Omega^n$ is absolutely continuous, then

$$J(\theta) = S_0(\theta_0) + \frac{1}{2}\int_0^T \mid \dot{\theta}_s - f(\theta_s) \mid^2 + (\dot{z}_s - h(\theta_s))^2 ds.$$

Theorem 5.6 $\{\Pi^\epsilon\}$ *obey the LDPP with action function* $J(\theta)$.

Proof: First note that the upper and lower bounds (iii), (iv) of Definition 2.1 follow as in the proof of Theorem 5.5 using Theorem 5.4.

By definition, $J \geq 0$ and is lower semicontinuous. Choose $L \geq 0$ such that $F \geq -L$. Then

$$\left\{\theta : \tilde{J}(\theta) \leq M\right\} \subset \{\theta : I(\theta) \leq M + L\}$$

which is compact. This proves (i) and (ii) of Definition 2.1. ∎

For each $0 \leq t \leq T$, define

$$J_t(x) = W(x,t) - \inf_{x' \in R^n} W(x',t), \tag{5.58}$$
$$= \inf_{\theta \in \Omega^n} \{J_t(\theta) \ : \ \theta_t = x\},$$

where $J_t(\theta)$ is defined by (5.57) with T replaced by t.

Corollary 5.2 $\{\pi_t^\epsilon\}$ *obey the LDPP with action function* $J_t(x)$.

5.3.3 Limiting Measure and Filter

Recall that the solution $X = \{x(t) : 0 \leq t \leq T\}$ of (2.35) with initial condition x_0 is the limit in probability of the solution of (2.2), and that $Z = \{z(t) : 0 \leq t \leq T\}$ given by (5.1) is the limit in probability of the solution of (2.3). $W(x,t)$ is the corresponding value function defined by (5.29), and $J_t(x)$ is defined by (5.58).

Lemma 5.7 *For each $0 \leq t \leq T$, $x(t)$ is the unique minimiser of $W(\cdot,t)$. That is, $J_t(x(t)) = 0$, $\hat{x}(t) = x(t)$, and for each $\alpha > 0$*

$$\inf_{x:|x-x(t)|\geq\alpha} J_t(x) \; > \; 0 \quad (0 \leq t \leq T).$$

Proof: Clearly $J_t(x) \geq 0$ for all $x \in \mathbb{R}^n$. Suppose $J_t(x^*) = 0$ for some $x^* \in \mathbb{R}^n$. Then since $\{\theta : \theta_t = x^*\}$ is closed, there exists $\theta^* \in \Omega_t^n$ such that $J_t(\theta^*) = 0$ and $\theta_t^* = x^*$. Then $\theta_0^* = x_0$ and

$$\dot{\theta}_s^* = f(\theta_s^*), \quad \dot{z}_s = h(\theta_s^*), \quad \text{a.e. } 0 \leq s \leq t.$$

Now $X \in \Omega_t^n$ is the unique solution of (2.35), hence $\theta^* = X$. Consequently, $x^* = x(t)$.

Now suppose that there exists $\alpha > 0$ such that

$$\inf_{x:|x-x(t)|\geq\alpha} J_t(x) \; = \; 0.$$

Let $\{x_k\}_{k=1}^{\infty}$ be a sequence such that

$$\lim_{k\to\infty} J_t(x_k) \; = \; 0.$$

Since J_t is an action function, we can assume $x_k \to x^*$ for some $x^* \in \mathbb{R}^n$ and $J_t(x^*) = 0$. Then $| x^* - x(t) | \geq \alpha > 0$ and the above argument shows $x^* = x(t)$; a contradiction. This completes the proof. ∎

The proof of the following theorem uses an extension of results in Varadhan [55].

Theorem 5.7 *Under the above assumptions, for each* $0 \leq t \leq T$, *we have:*

$$\pi_t^\epsilon \stackrel{P}{\Rightarrow} \delta_{x(t)} \tag{5.59}$$

as $\epsilon \to 0$.

Proof: Let $0 \leq t \leq T$. We must show that for each $\alpha > 0$, $\beta > 0$ there exists $\epsilon_0 > 0$ such that $0 < \epsilon \leq \epsilon_0$ implies

$$P\left(\varrho(\pi_t^\epsilon, \delta_{x(t)}) > \alpha\right) \; < \; \beta. \tag{5.60}$$

Define

$$B_\alpha \; = \; \{x \in I\!R^n : \mid x - x(t) \mid \geq \alpha\}.$$

By hypothesis,

$$\inf_{x \in B_\alpha} \{J_t(x)\} \; = \; \gamma \; > \; 0.$$

Since $\{\pi_t^\epsilon\}$ obey the LDPP, there exists $\epsilon_1 > 0$ such that $0 < \epsilon \leq \epsilon_1$ implies

$$P(A_\epsilon) \; \geq \; 1 - \beta$$

for

$$A_\epsilon \; = \; \left\{\epsilon \log \pi_t^\epsilon(C) \leq - \inf_{x \in C} J_t(x) + \frac{\gamma}{2}\right\}.$$

Choose $0 < \epsilon_0 \leq \epsilon_1$ such that

$$e^{-\frac{\gamma}{2\epsilon}} \; \leq \; \alpha$$

provided $0 < \epsilon \leq \epsilon_0$.

Let $C \subset I\!R^n$ be closed. Then on A_ϵ, if $0 < \epsilon \leq \epsilon_0$,

$$\pi_t^\epsilon(B_\alpha) \; \leq \; e^{-\frac{\gamma}{2\epsilon}},$$

hence

$$\pi_t^\epsilon(C) \;\leq\; \delta_{x(t)}(C^\alpha) + \alpha,$$

since: if $C \subset B_\alpha$, then $\delta_{x(t)}(C^\alpha) = 0$ and $\pi_t^\epsilon(C) \leq \pi_t^\epsilon(C^\alpha) \leq \alpha$; if $C \not\subset B_\alpha$, then $\delta_{x(t)}(C^\alpha) = 1$.

Consequently, (5.60) holds. ∎

We turn now to the pathwise filters. Recall the maps

$$\bar{\pi}_t^\epsilon, \bar{\pi}_t : \Omega_0 \longrightarrow \mathcal{P}(\mathbb{R}^n),$$

where $\bar{\pi}_t^\epsilon$ is the pathwise filter (2.20), (2.21) and $\bar{\pi}_t$ is the deterministic estimator (5.13); assumed well defined.

Theorem 5.8 *Under the above assumptions, we have*

$$\bar{\pi}_t^\epsilon[\eta] \Rightarrow \bar{\pi}_t[\eta] \tag{5.61}$$

as $\epsilon \longrightarrow 0$ *uniformly on compact subsets of* Ω_0.

Proof: 1. From the results of earlier sections, we see that for each $\eta \in \Omega_0$, $\{\bar{\pi}_t^\epsilon[\eta]\}$ obey the LDP with action function

$$J_t(x)[\eta] \;=\; W(x,t)[\eta] - W(\hat{x}(t),t)[\eta]. \tag{5.62}$$

In fact, we claim the following: Let $K \subset \Omega_0$ be compact, and $C \subset \mathbb{R}^n$ closed. For all $\beta > 0$, there exists $\epsilon_1 > 0$ such that

$$\epsilon \log \bar{\sigma}_t^\epsilon(C)[\eta] \;\leq\; -\inf_{x\in C} W(x,t)[\eta] + \beta \tag{5.63}$$

for all $0 < \epsilon \leq \epsilon_1$, and all $\eta \in K$. Here, $\bar{\sigma}_t^\epsilon$ is the unnormalised measure defined by (2.19).

Suppose that (5.63) is false. Then there exists $\beta > 0$, and sequences $\epsilon_i \to 0$, $\{\eta_i\} \subset K$ such that

$$\epsilon_i \log \bar{\sigma}_t^{\epsilon_i}(C)[\eta_i] \;\geq\; -\inf_{x \in C} W(x,t)[\eta_i] + \beta$$

for all $i = 1, 2, \ldots$ Since K is compact, we may assume that $\eta_i \to \bar{\eta} \in K$ as $i \to \infty$. This implies

$$\liminf_{i \to \infty} \epsilon_i \log \bar{\sigma}_t^{\epsilon_i}(C)[\eta_i] \;\geq\; -\inf_{x \in C} W(x,t)[\bar{\eta}] + \beta. \tag{5.64}$$

Now employing the notation used in the proof of Proposition 2.1,

$$\begin{aligned}
\bar{\sigma}_t^{\epsilon_i}[\eta_i](C) &= EI_{x_t^{\epsilon_i} \in C} \exp\left(\frac{1}{\epsilon_i} G(X^{\epsilon_i}, \eta_i)\right) \\
&= EI_{x_t^{\epsilon_i} \in C} \exp\left(\frac{1}{\epsilon_i} G(X^{\epsilon_i}, \bar{\eta})\right) \exp\left(\frac{1}{\epsilon_i} F(X^{\epsilon_i}, \eta_i - \bar{\eta})\right) \\
&\leq \left(EI_{x_t^{\epsilon_i} \in C} \exp\left(\frac{2}{\epsilon_i} G(X^{\epsilon_i}, \bar{\eta})\right)\right)^{\frac{1}{2}} \left(E \exp\left(\frac{2}{\epsilon_i} F(X^{\epsilon_i}, \eta_i - \bar{\eta})\right)\right)^{\frac{1}{2}} \\
&\leq \left(\bar{\sigma}_t^{\frac{\epsilon_i}{2}}(C)[\bar{\eta}]\right)^{\frac{1}{2}} \exp\left(\frac{M \parallel \eta_i - \bar{\eta} \parallel}{\epsilon_i}\right).
\end{aligned}$$

The last inequality uses the arguement from part 2 of the proof of Proposition 2.1. Hence

$$\limsup_{i \to \infty} \epsilon_i \log \bar{\sigma}_t^{\epsilon_i}(C)[\eta_i] \;\leq\; -\inf_{x \in C} W(x,t)[\bar{\eta}].$$

Combining this last inequality with (5.64) we deduce $\beta \leq 0$, a contradiction. This proves (5.63).

2. We next claim that for all $\alpha > 0$, there exists $\gamma > 0$ such that

$$\inf_{\eta \in K,\; x : |x - \hat{x}(t)[\eta]| \geq \alpha} J_t(x)[\eta] \;\geq \gamma \;> 0. \tag{5.65}$$

Suppose not. Then there exists $\alpha > 0$, $\gamma_i \to 0$, $\eta_i \to \bar{\eta} \in K$ and $\{x_i\} \subset I\!R^n$ with

$$|\; x_i - \hat{x}(t)[\eta_i] \;| \geq \;\alpha$$

and

$$0 \leq J_t(x_i)[\eta_i] \leq \gamma_i. \tag{5.66}$$

From Proposition 2.2, it follows that there exists $C' > 0$ such that

$$W(x,t)[\eta] \geq C' \mid x \mid - C'$$

for all $(x,t) \in I\!R^n \times [0,T]$, $\eta \in K$. Now $\eta_i \to \bar{\eta}$ implies $\hat{x}(t)[\eta_i] \to \hat{x}(t)[\bar{\eta}]$, so there exists $M > 0$ such that

$$C' \mid x_i \mid - C' \leq W(x_i,t)[\eta_i] \leq M,$$

which implies

$$\mid x_i \mid \leq \frac{M + C'}{C'}.$$

Therefore we can assume $x_i \to \bar{x} \in I\!R^n$ with

$$\mid \bar{x} - \hat{x}(t)[\bar{\eta}] \mid \geq \alpha.$$

Sending $i \to \infty$ in (5.66) we deduce

$$W(\bar{x},t)[\bar{\eta}] = W(\hat{x}(t),t)[\bar{\eta}].$$

Since the deterministic estimator is well defined, this equality forces $\bar{x} = \hat{x}(t)[\bar{\eta}]$, a contradiction. This proves (5.65).

3. To prove (5.61), we must show that for all $\delta > 0$ there exists $\epsilon_0 > 0$ such that

$$\varrho(\bar{\pi}_t^\epsilon[\eta], \bar{\pi}_t[\eta]) \leq \delta \tag{5.67}$$

for all $0 < \epsilon \leq \epsilon_0$ and all $\eta \in K$; where ϱ is the Prohorov metric on $\mathcal{P}(I\!R^n)$.

Let $\delta > 0$ and define

$$B_\delta = \{x : \mid x - \hat{x}(t)[\eta] \mid \geq \delta\}.$$

From (5.63), (5.65), there exists $\epsilon_1 > 0, \gamma > 0$ such that

$$\epsilon \log \bar{\pi}_t^\epsilon(B_\delta)[\eta] \;\leq\; -\inf_{x \in B_\delta} J_t(x)[\eta] + \frac{\gamma}{2},$$

and hence

$$\bar{\pi}_t^\epsilon(B_\delta)[\eta] \;\leq\; e^{-\frac{\gamma}{2\epsilon}}$$

for all $0 < \epsilon \leq \epsilon_1$ and all $\eta \in K$. Choose $\epsilon_0 > 0$ such that

$$e^{-\frac{\gamma}{2\epsilon}} \;\leq\; \delta.$$

Then we deduce

$$\bar{\pi}_t^\epsilon(C)[\eta] \;\leq\; \bar{\pi}_t(C^\delta)[\eta] + \delta$$

for all $0 < \epsilon \leq \epsilon_0$, $\eta \in K$. Hence (5.67) follows. ∎

Chapter 6
Observer Designs

In this chapter we present observer designs for the nonlinear control system

$$\begin{aligned} \dot{x} &= f(x,u), \qquad x(0) = x_0, \\ y &= h(x) \end{aligned} \tag{6.1}$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, $|u_i| \leq 1$ $i = 1, \ldots, m$ and $y \in \mathbb{R}^p$. The initial condition x_0 is unknown.

We first consider a finite dimensional observer, and prove the following result for this design: provided that we have some knowledge of x_0 (in the form $|x_0 - m_0| < \rho$, where m_0 is the initial condition of the observer) and assuming that (6.1) satisfies a *detectability* condition, then the observer trajectory $m(t)$ converges exponentially to the system trajectory $x(t)$ as $t \to \infty$. The radius of convergence ρ depends on the nonlinearities in the dynamics and observations as well as on certain design parameters. For a certain class of systems, no knowledge of x_0 is required.

This design is a development of the design given in Baras, Bensoussan and James [2], a paper which treats systems with uncontrolled nonlinear dynamics and linear observations. The main contributions here are the results for nonlinear observations and controlled dynamics. We remark that these designs do not involve coordinate transformations, canonical forms, local linearization, etc., and seem robust when compared with other designs. However, the designs do involve solving Riccati equations and computing certain matrices and constants.

We also give an infinite dimensional observer design based on the deterministic estimator. Assuming that (6.1) satisfies an *observability* condition, the observer computes exactly the unknown state after a finite time has elapsed.

6.1 An Observer Design

We assume that f, h are smooth with bounded derivatives of orders 1 and 2. Assume $t \mapsto u(t)$ is continuous. Let $N \in L(I\!R^n, I\!R^n)$, $R \in L(I\!R^p, I\!R^p)$ and assume rank $N = n$ and $R > 0$.

Write $A(x,u) = Df(x,u)$, $H(x) = R^{-1}Dh(x)$, where D denotes gradient in the x variable. Set

$$\|A\| = \sup\{\|A(x,u)\| : \ x \in I\!R^n, \ |u_i| \le 1\}$$

and similarly define $\|H\|$, and so on.

Consider the coupled system

$$\begin{aligned}
\dot{m}(t) &= f(m(t),u(t)) + Q(t)H(m(t))'R^{-1}\left(y(t) - h(m(t))\right) && (6.2)\\
m(0) &= m_0 \\
\dot{Q}(t) &= A(m(t),u(t))Q(t) + Q(t)A(m(t),u(t))' \\
& \quad -Q(t)H(m(t))'H(m(t))Q(t) + NN' && (6.3)\\
Q(0) &= Q_0 > 0.
\end{aligned}$$

This is our finite dimensional observer for (6.1). It is essentially a modification of the deterministic or minimum energy estimator, as discussed in Baras, Bensoussan and James [2]. In Section 4.4, this system was interpreted as an approximate deterministic minimum energy estimator. Note in particular that the Riccati differential equation (6.3) depends on the control. This is not necessary when $f(x,u) = f(x) + Bu$: set $A(x) = Df(x)$.

We will assume that the system (6.1) is *uniformly detectable, or uniformly of full rank*; see Section 3.2. Since N has rank n and $\|A\| < \infty$, the pair $(A(x,u), N)$ is *uniformly stabilisable* (refer to Section 6.2.1), and $NN' \ge r_0 I$ for some $r_0 > 0$.

Let $P_0 = Q_0^{-1}$, $P(t) = Q(t)^{-1}$, and let p, q be the bounds for $\|P(t)\|$, $\|Q(t)\|$ (given in Section 6.2.2).

Regard A_0, N, R as design parameters. Define $\rho = \rho(Q_0, N, R)$ by

$$\rho = \frac{r_0}{q^2\|P_0^{1/2}\|}\left(\sqrt{p}\|D^2 f\| + \sqrt{q}\|R^{-1}\|^2\,\|Dh\|\,\|D^2 h\|\right)^{-1} \tag{6.4}$$

Our main theorem is the following convergence result, similar to Theorem 8 in [2].

Theorem 6.1 *Assume there exist Q_0, N, R such that*

$$|x_0 - m_0| < \rho(Q_0, N, R). \tag{6.5}$$

Then the system (6.2), (6.3) *is an observer for the nonlinear control system* (6.1) *provided that* (6.1) *is uniformly detectable or uniformly of full rank, and the above assumptions hold. That is, there exist contants $K > 0, \gamma > 0$ such that*

$$|x(t) - m(t)| \leq K|x_0 - m_0|e^{-\gamma t} \tag{6.6}$$

for all $t \geq 0$.

Remark There is a trade-off between the decay rate $\gamma = \gamma(Q_0, N, R)$ and the radius of convergence ρ. The designer can optimize his choice of γ, ρ by varying the design parameters. ///

By using different estimates for the nonlinearities, we obtain an observer for (6.1) without any contraints on the initial conditions x_0, m_0 for a class of systems. Included in this class are systems for which $A(x, u)$ is uniformly negative definite (see the example in Section 6.4.2).

Define $\delta = \delta(Q_0, N, R)$ by

$$\delta = \frac{r_0}{q^2} - 4p\|Df\| - 4\|R^{-1}\|^2\,\|Dh\|^2. \tag{6.7}$$

If $D^2 f$ or $D^2 h$ is zero, we omit the corresponding term from (6.7).

Corollary 6.1 *Assume there exist Q_0, N, R such that*

$$0 < \delta(Q_0, N, R). \tag{6.8}$$

Then the system (6.2), (6.3) *is an observer for the control system* (6.1) *provided that* (6.1) *is uniformly detectable or uniformly of full rank, and the above assumptions hold. That is, there exist contants $K > 0, \gamma > 0$ such that*

$$|x(t) - m(t)| \leq K|x_0 - m_0|e^{-\gamma t} \tag{6.9}$$

for all $t \geq 0$ and all $x_0, m_0 \in I\!R^n$.

Remark Our design can readily be extended to time varying systems. One has to modify appropriately the detectability conditions. ///

6.2 Riccati Equations

Write $X = I\!R^n \times [-1,1]^m$ and $\xi = (x, u) \in X$. If $t \mapsto \xi_t = (x_t, u_t)$ is a continuous curve, we write $A_t = A(\xi_t) = A(x_t, u_t)$, etc.

Consider the Riccati differential equations

$$\begin{aligned} \dot{Q}_t &= A_t Q_t + Q_t A_t' - Q_t H_t' H_t Q_t + NN', && (6.10) \\ \dot{P}_t &= -P_t A_t - A_t' P_t - P_t NN' P_t + H_t' H_t, && (6.11) \\ Q_0 &= P_0^{-1} > 0. \end{aligned}$$

Existence and uniqueness for (6.10), (6.11) are standard, and note that $P_t = Q_t^{-1}$.

In this section, we obtain uniform bounds for the solutions of these Riccati equations.

6.2.1 Uniform Detectability and Stabilisability

The bound for $\| Q_t \|$ requires a detectability condition, such as uniform detectability or uniform full rank.

To obtain a uniform bound for $\| P_t \|$, we assume that $\mathrm{rank} N = n$ and use the following *uniform stabilisation* result, based on Kalman [35]. Let $\Phi_F(t, t_0)$ denote the fundamental transition matrix corresponding to a time varying matrix F_t. Recall $NN' \geq r_0 I$.

Lemma 6.1 *Assume* $\mathrm{rank} N = n$. *Consider the control system*

$$\dot{z}_t = -A_t z_t - N u_t, \quad z(0) = z, \tag{6.12}$$

where $A_t = A(\xi_t)$ *for some curve* $t \mapsto \xi_t$. *Then there exists a feedback control* $u_t^o = \Gamma_t z_t$ *such that*

$$\| \Phi_{\tilde{A}}(t, t_0) \| \leq \sqrt{\frac{\beta_1}{\beta_0}} \exp\left(-\frac{1}{2\beta_1}(t - t_0)\right), \tag{6.13}$$

for $t \geq t_0 \geq 0$, *where* $\tilde{A}_t = -A_t - N\Gamma_t$ *and for any* $\sigma > 0$,

$$\beta_0(\sigma) = \sigma e^{-2\sigma\|A\|} \left(1 + \sigma^2 e^{2\sigma\|A\|} \| N \|^2\right)^{-1},$$

$$\beta_1(\sigma) = \sigma e^{4\sigma\|A\|} \left(1 + \frac{\| N \|^2}{r_0 \sigma}\right),$$

$$\| \Gamma_t \| \leq \| N \| \beta_1(\sigma) \equiv \| \Gamma \| .$$

Proof: Consider the optimal control problem of minimising the cost

$$J_u(t_0, z) = \int_{t_0}^{\infty} \| z_t \|^2 + \| u_t \|^2 \, dt$$

over controls u for the system (6.12). Define the controllability grammian

$$C(t_0, t_0 + \sigma) = \int_{t_0}^{t_0+\sigma} \Phi_{-A}(t_0, t) N N' \Phi'_{-A}(t_0, t) dt.$$

Note that $\| \Phi_{-A}(t,t_0) \|^2 \leq e^{2\|A\|}$, and

$$r_0 \sigma e^{-2\|A\|} I \leq \mathcal{C}(t_0, t_0+\sigma) \leq \| N \|^2 \sigma e^{2\|A\|} I$$

so that the system (6.12) is uniformly completely controllable [35].

The value function is given by

$$V(t_0, z) = \frac{1}{2} z' Z_{t_0} z,$$

where Z_t solves the Riccati equation

$$\dot{Z}_t = A_t Z_t + Z_t A_t' + Z_t N N' Z_t - I.$$

The optimal control is

$$u_t^o = -N' Z_t z_t.$$

Now

$$\begin{aligned} V(t_0, z) &\geq \int_{t_0}^{t_0+\sigma} \| z_t \|^2 \, dt \\ &\geq \beta_0(\sigma) \mid z \mid^2 . \end{aligned}$$

Also, using the control

$$u_t^1 = N' \Phi_{-A}(t_0, t) \mathcal{C}(t_0, t_0+\sigma)^{-1} z,$$

we obtain

$$V(t_0, z) \leq \beta_1(\sigma) \mid z \mid^2 .$$

Finally, note that $V(t,z)$ is a Lyapunov function for

$$\dot{z}_t^o = -(A_t + N\Gamma_t) z_t^o;$$

whence (6.13). ∎

6.2.2 Bounds

Theorem 6.2 *Assume that $\xi \mapsto A(\xi), H(\xi)$ are continuous and bounded,* (6.1) *is uniformly detectable, and rank$N = n$. Then we have*

$$\|Q_T\| \leq \left(\|Q_0\| + \frac{\|N\|^2 + \|\Lambda\|^2}{2\alpha_0}\right) \equiv q < \infty, \tag{6.14}$$

$$\|P_T\| \leq \left(\frac{\beta_0}{\beta_1}\|P_0\| + \frac{\|H\|^2 + \|\Gamma\|^2}{2\beta_0}\right) \equiv p < \infty. \tag{6.15}$$

These bounds are independent of $T > 0$.

Note The bound q depends on the choice of Λ, while p depends on σ. To obtain the best bound, one can optimise over these parameters. For linear time–invariant detectable systems, one can also obtain a bound for $\| Q_t \|$. ///

Proof: We modify an argument due to A. Bensoussan in [2]. To prove (6.14), consider the following linear optimal control problem with time-varying coefficients:

$$-\dot{\eta}_t = A_t'\eta_t + H_t'v_t, \qquad \eta_T = h, \tag{6.16}$$

where $h \in I\!R^n$ is given and v is the control. The cost functional is

$$J_1(v,T) = \eta_0'Q_0\eta_0 + \int_0^T (v_t'v_t + \eta_t'NN'\eta_t)dt. \tag{6.17}$$

Define a value function

$$V_1(h,T) = \inf\{J_1(v,T) : \eta_T = h\}$$

The Hamilton-Jacobi-Bellman (HJB) equation is

$$\frac{\partial}{\partial T}V_1 + \max_v[D_\eta V_1(-A_t' - H_t'v) - v^2 - \eta'NN_\eta'] = 0$$

Let Q_t be the solution of (6.10). Then

$$V(\eta, t) = \eta'Q_t\eta$$

is the unique (viscosity) solution of (6.11) with $V(\eta, 0) = \eta' Q_0 \eta$.

Consider the (suboptimal) feedback control law

$$v(t) = \Lambda_t' \eta_t.$$

Then by (6.16),

$$-\dot{\eta}_t = (A_t' + H_t' \Lambda_t')\eta_t, \quad \eta_T = h. \tag{6.18}$$

Then we have

$$V(\eta, T) = h' Q_T h \le \eta_0 Q_0 \eta_0 + \int_0^T \eta_t' (NN' + \Lambda_t \Lambda_t')\eta_t dt \tag{6.19}$$

Now using (6.18),

$$|\eta_0|^2 - 2\int_0^T \eta_t'(A_t' + H_t' \Lambda_t')\eta_t dt = |h|^2$$

Hence using uniform detectability (3.7), $|\eta_0|^2 \le |h|^2$ and

$$\int_0^T |\eta_t|^2 dt \le \frac{|h|^2}{2\alpha_0}.$$

Combining this with (6.19) we obtain

$$h' Q_T h \le h' \left(\|Q_0\| + \frac{\|N\|^2 + \|\Lambda\|^2}{2\alpha_0} \right) h$$

which proves (6.14).

Similarly, to prove (6.15), consider the optimal control problem

$$\dot{\lambda}_t = A_t \lambda_t + N v_t, \quad \lambda_T = h \tag{6.20}$$

with cost

$$J_2(v, T) = \lambda_0' P_0 \lambda + \int_0^T (v_t' v_t + \lambda_t' H_t' H_t \lambda_t) dt.$$

Define the value function

$$V_2(h, T) = \inf\{J_2(v, T) : \quad \lambda_T = h\}.$$

Then the HJB equation

$$\frac{\partial}{\partial T}V_2 + \max_v [D_\lambda V_2(A_t\lambda + Nv) - v^2 - \lambda' H_t' H_t \lambda] = 0,$$

with $V_2(\lambda, 0) = \lambda' P_0 \lambda$ and solution $V_2(\lambda, t) = \lambda' P_t \lambda$, where P_t is the solution of (6.11). By Lemma 6.1, set

$$v(t) = \Gamma_t \lambda_t,$$

then by (6.20)

$$\dot{\lambda}_t = (A_t + N\Gamma_t); \qquad \lambda_T = h. \tag{6.21}$$

Thus

$$h' P_T h \leq \lambda_0' P_0 \lambda_0 + \int_0^t \lambda_t' (\Gamma_t' \Gamma_t + H_t' H_t) \lambda_t \, dt \tag{6.22}$$

Now

$$\lambda_t = \Phi_{\tilde{A}}(-t, -T)h,$$

and hence $| \lambda_0 |^2 \leq \frac{\beta_0}{\beta_1} | h |^2$ and

$$\int_0^T |\lambda_t|^2 dt \leq \frac{|h|^2}{2\beta_0}.$$

This together with (6.22) yields (6.15). ∎

Corollary 6.2 *Assume that $\xi \mapsto A(\xi)$, $H(\xi)$ are bounded and continuous, and that* (6.1) *is uniformly of full rank. Then*

$$\| Q_T \| \leq \left(\frac{\alpha_0}{\alpha_1} \| Q_0 \| + \frac{\| N \|^2 + \| \Lambda \|^2}{2\alpha_0} \right) \equiv q \leq \infty, \tag{6.23}$$

for all $T > 0$, where for any $\tau > 0$,

$$\alpha_0(\tau) = \tau e^{-2\tau\|A\|} \left(1 + \tau^2 e^{2\tau\|A\|} \| N \|^2\right)^{-1},$$

$$\alpha_1(\tau) = \tau e^{4\tau\|A\|} \left(1 + \frac{\| N \|^2}{r_0 \tau}\right),$$

$$\| \Lambda_t \| \leq \| H \| \alpha_1(\tau) \equiv \| \Gamma \| .$$

Proof: Consider the control system

$$\dot{z}_t = A'_t z_t + H'_t u_t, \quad z(0) = z, \tag{6.24}$$

where $A_t = A(\xi_t)$ for some curve $t \mapsto \xi_t$. Define the grammian

$$\mathcal{O}(t_0, t_0 + \tau) = \int_{t_0}^{t_0+\tau} \Phi_{A'}(t_0, t) H'_t H_t \Phi'_{A'}(t_0, t) dt.$$

Now $\| \Phi_{A'}(t, t_0) \|^2 \leq e^{2\|A\|}$, and

$$s_0 \tau e^{-2\|A\|} I \leq \mathcal{O}(t_0, t_0 + \tau) \leq \| H \|^2 \tau e^{2\|A\|} I$$

so that the system (6.24) is uniformly completely controllable [35]. Now proceed as in the proof of Lemma 6.1. ∎

6.3 Asymptotic Convergence

Using the bounds (6.14), (6.15) we prove Theorem 6.1, and Corollary 6.1. The proofs are modifications of a proof due to A. Bensoussan in [2].

Proof of Theorem 6.1 The error $e(t) = x(t) - m(t)$ satisfies

$$\begin{aligned}
\dot{e}(t) &= f(x(t), u(t)) - f(m(t), u(t)) - Q(t) H(m_t)' R^{-1} (y(t) - h(m(t))) \\
&= [A(m(t), u(t)) - Q(t) H(m(t))' H(m(t))] e(t) \\
&\quad + [f(x(t), u(t)) - f(m(t), u(t)) - Df(m(t), u(t)) e(t)] \\
&\quad - Q(t) H(m(t))' R^{-1} [h(x(t)) - h(m(t)) - Dh(m(t)) e(t)]
\end{aligned}$$

Therefore using the Riccati equation (6.11) for $P(t)$,

$$\begin{aligned}\frac{d}{dt}e(t)'\,P(t)\,e(t) &= -e(t)'\,P(t)NN'P(t)e(t) - e(t)'H(m(t))'H(m(t))e(t) \\ &\quad +2e(t)'P(t)\int_0^1\int_0^1 rD^2f(m(t)+rse(t),u(t))e(t)^2drds \\ &\quad -2e(t)'H(m(t))'R^{-1}\int_0^1\int_0^1 rD^2h(m(t)+rse(t))e(t)^2drds\end{aligned}$$

$$\begin{aligned}\frac{d}{dt}|P(t)^{\frac{1}{2}}e(t)|^2 &\le e(t)'\left(-r_0/q^2 + |P(t)^{\frac{1}{2}}e(t)|[\sqrt{p}\|D^2f\|\right. \\ &\quad \left. +\sqrt{q}\|R^{-1}\|^2\,\|Dh\|\,\|D^2h\|]\right)e(t)\end{aligned} \tag{6.25}$$

Let $C = (\sqrt{p}\|D^2f\| + \sqrt{q}\|R^{-1}\|^2\,\|Dh\|\,\|D^2h\|)$. By hypothesis (6.5) we have

$$-\frac{r_0}{q^2} + |P_0^{\frac{1}{2}}e_0|C < 0.$$

Since $P(t)^{\frac{1}{2}}e(t)$ is continuous, there is an interval $[0,t_0)$ such that

$$-\frac{r_0}{q^2} + |P(t)^{\frac{1}{2}}e(t)|C < 0 \quad \text{on} \quad [0,t_0).$$

But then (6.25) implies

$$\frac{d}{dt}|P(t)^{\frac{1}{2}}e(t)|^2 < 0 \quad \text{on} \quad [0,t_0),$$

and thus

$$|P(t)^{\frac{1}{2}}e(t)| \le |P_0^{\frac{1}{2}}e_0|$$

for $t \in [0,t_0)$. By continuity this inequality holds for $t \in [0,t_0]$. Hence we can proceed from t_0 on. Thus there exists $\delta > 0$ such that

$$|P(t)^{\frac{1}{2}}e(t)| \le \frac{1}{C}\left(\frac{r_0}{q^2} - \delta\right)$$

for all $t \ge 0$. So (6.25) implies

$$\frac{d}{dt}|P(t)^{\frac{1}{2}}e(t)|^2 \le -\delta|e(t)|^2.$$

But from (6.15)

$$e(t)'P(t)e(t) \le \|P(t)\| \; |e(t)|^2 \le p|e_t|^2,$$

so that

$$\frac{d}{dt}e(t)'P(t)e(t) \le -\frac{\delta}{p}e(t)'P(t)e(t),$$

which implies

$$e(t)'\,P(t)e(t) \le e(0)'P_0e(0)e^{-\frac{\delta}{p}t}, \;\; t \ge 0.$$

Therefore, using (6.14), we have

$$\begin{aligned} |e(t)|^2 &\le q\,e(t)'P(t)e(t) \\ &\le q\,e(0)'P_0e(0)e^{-\frac{\delta}{p}t}, \;\; t \ge 0, \end{aligned}$$

which implies (6.6). ∎

Proof of Corollary 6.1 We have

$$\begin{aligned} \frac{d}{dt}e(t)'P(t)e(t) &= -e(t)'P(t)NN'P(t)e(t) - e(t)'H(m(t))'H(m(t))e(t) \\ &\quad +2e(t)'P(t)(f(x(t),u(t)) - f(m(t),u(t)) - Df(m(t),u(t)))e(t)) \\ &\quad -2e(t)'(R^{-1}Dh(m(t))'R^{-1}(h(x(t)) - h(m(t)) - Dh(m(t)e(t)) \\ &\le \left(-\frac{r_0}{q^2} + 4p\|Df\| + 4\|R^{-1}\|^2\,\|Dh\|^2\right)\,|e(t)|^2 \end{aligned}$$

By assumption (6.8) there is a $\delta > 0$ such that

$$-\frac{r_0}{q^2} + 4p\|Df\| + 4\|R^{-1}\|^2\,\|Dh\|^2 = -\delta < 0.$$

Therefore

$$\frac{d}{dt}|P(t)^{\frac{1}{2}}e(t)|^2 \le -\delta|e(t)|^2$$

for all $t \ge 0$ and all $e_0 \in I\!R^n$. This implies (6.9). ∎

6.4 Examples

We now give some simple examples to illustrate the applicability of our design.

6.4.1 Bilinear Dynamics, Linear Observation

Consider the general bilinear system

$$\begin{aligned} \dot{x} &= \left(A + \sum_{i=1}^{m} u_i B_i\right) x, \qquad x(0) = x_0, \\ y &= Cx. \end{aligned} \tag{6.26}$$

We assume $|u_i| \leq 1$, $p = 1$, and here $\xi = u \in [-1, 1]^m = X$. Write

$$A(u) = A + \sum_{i=1}^{m} u_i B_i.$$

Define, for $\tau > 0$, the observability grammian

$$\mathcal{O}(t_0, t_0 + \tau) = \int_{t_0}^{t_0+\tau} \Phi_{A'}(t_0, t)C'C\Phi'_{A'}(t_0, t)dt,$$

where $A_t = A(u(t))$. Assume that (6.26) is *uniformly observable* in the sense that there exists $\tau > 0$ such that for all $t_0 \geq 0$

$$\gamma_0(\tau)I \leq \mathcal{O}(t_0, t_0 + \tau) \leq \gamma_1(\tau)I$$

for constants $\gamma_0(\tau)$, $\gamma_1(\tau) > 0$, independent of the control. The we can bound $\| Q_t \|$ as in Corollary 6.2.

Then the following system is an observer for (6.26), with no contraints on the initial conditions.

$$\begin{aligned} \dot{m}(t) &= A(u(t))m(t) + Q(t)C'(y(t) - Cm(t)), \; m(0) = m_0, \\ \dot{Q}(t) &= A(u(t))Q(t) + Q(t)A(m(t))' - Q(t)C'CQ(t) + I, \;\; Q_0 = I. \end{aligned} \tag{6.27}$$

For simplicity we have taken Q_0, N, R to be identity matrices. To improve the decay rate γ, one could try other values for Q_0, N, R.

Compare this design with the design for linear time–varying systems in Willems and Mitter [57], and O' Reilly [49].

Now in $I\!R^2$ consider

$$
\begin{aligned}
A(u) &= \begin{pmatrix} a_{11}(u) & a_{12}(u) \\ a_{21}(u) & a_{22}(u) \end{pmatrix}, \\
C &= (1,0),
\end{aligned}
$$

and assume that

$$a_{22}(u) \leq -\alpha_0, \quad \alpha_0 > 0,$$

for all $| u | \leq 1$. Then setting

$$\Lambda(u) = \begin{pmatrix} -\alpha_0 - a_{11}(u) \\ -a_{12}(u) - a_{21}(u) \end{pmatrix}$$

gives

$$\eta' \left(A(u) + \Lambda(u)C\right)\eta = -\alpha_0 | \eta |^2 .$$

Then $(C, A(u))$ is uniformly detectable and (6.27) is an observer for (6.26).

6.4.2 Stable Linear Dynamics, Nonlinear Observation

Consider the system

$$
\begin{aligned}
\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} &= \begin{pmatrix} 1 & \sqrt{2} \\ \sqrt{2} & -4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} u, \\
y &= \sin x_1.
\end{aligned} \tag{6.28}
$$

This system is controllable and observable. However, the pair of matrices $(Dh(x), A)$ is not observable for $x_1 = k\frac{\pi}{2}$, where k is an odd integer. The system has eigenvalues $-1, -2$ and A is symmetric, hence $(Dh(x), A)$ is automatically uniformly detectable, with $\alpha_0 = 1$, $\Lambda(x) \equiv 0$. Let $R = rI$, $N = \sqrt{r_0}I$, $Q_0 = \gamma^2 I$. Here, $H(x) = \frac{1}{r}(\cos x_1, 0)$. Now

$$\delta = r_0(\gamma^2 + r_0/2)^{-2} - 4r^{-2}.$$

Set $r = 3$, $r_0 = 0.2$, $\gamma = 0.1$. Then $\delta = 7.82$.

The observer for (6.28) is

$$\begin{aligned} \dot{m}(t) &= Am(t) + Bu(t) + \frac{1}{3}Q(t)\, H(m(t))'(y(t) - \sin m_1(t)), \qquad (6.29) \\ \dot{Q}(t) &= AQ(t) + Q(t)A' - Q(t)H(m(t))'H(m(t))Q(t) + 0.2I. \end{aligned}$$

By Corollary 6.1, $m(t)$ converges exponentially to $x(t)$ for all $x_0, m_0 \in \mathbb{R}^n$.

6.4.3 A Special Case

We now give an example of a class of systems for which observer design is easy. Consider

$$\begin{aligned} \dot{x}(t) &= f(x(t), u(t)), \quad x(0) = x_0, \qquad (6.30) \\ y(t) &= Cx(t), \end{aligned}$$

assuming that C is invertable:

$$C'C \geq s_0 I > 0.$$

Then the system

$$\dot{m}(t) = f(m(t), u(t)) + \frac{1}{\alpha}C'(y(t) - Cm(t)), \quad m(0) = m_0 \qquad (6.31)$$

is an observer for the control system (6.30) for all initial values provided that given $\gamma > 0$, α is chosen so that

$$\alpha \leq \frac{s_0}{\gamma + \| A \|}.$$

To see this, we have

$$\begin{aligned}
\frac{d}{dt} \mid e(t) \mid^2 &= 2e(t)\left(f(x(t),u(t)) - f(m(t),u(t)) - \frac{1}{\alpha}C'Ce(t)\right) \\
&\leq 2\left(\parallel A \parallel - \frac{s_0}{\alpha}\right) \mid e(t) \mid^2 \\
&\leq -2\gamma \mid e(t) \mid^2 .
\end{aligned}$$

Thus

$$\mid x(t) - m(t) \mid \leq \mid x_0 - m_0 \mid e^{-\gamma t}$$

for all $x_0, m_0 \in I\!R^n$.

Of course one can compute $x(t)$ directly by inverting C. The above dynamical procedure seems more robust. In fact, one could use it to invert matrices!

The rate of convergence can be made arbitrarily fast by adjusting $\gamma > 0$.

6.5 An Exact Observer

In this section we shift our perspective somewhat and obtain an infinite dimensional observer for uncontrolled systems. The observer recovers exactly the unknown state under a natural observability condition.

Consider the dynamical system

$$\begin{aligned}
\dot{x} &= f(x), \quad x(0) = x_0, \\
y &= h(x),
\end{aligned} \tag{6.32}$$

and assume $f \in C_b^1(I\!R^n, I\!R^n)$, $h \in C_b(I\!R^n, I\!R^p)$. Let $x(t) = \gamma(t)x_0$ denote the solution to (6.32).

Definition 6.1 *The system* (6.32) *is* <u>*observable in finite time*</u> *if there exists* $0 \leq T < \infty$ *such that for any* $x_0^1 \neq x_0^2$ *there is a* $t \in [0,T]$ *such that*

$$h(\gamma(t)x_0^1) \neq h(\gamma(t)x_0^2).$$

Compare with the definition of observability given in Chapter 3, and also the definition for time–varying linear systems. Our design also applies to controlled systems provided that Definition 6.1 holds for each control function.

We modify the deterministic estimator (4.7) by setting $S_0 \equiv 0$, and obtain the following system:

$$
\begin{aligned}
\tfrac{\partial}{\partial t}m(x,t) + f(x)\cdot Dm(x,t) + \frac{1}{2} \mid Dm(x,t) \mid^2 - \frac{1}{2}h(x)^2 + y(t)h(x) &= 0, \\
m(x,0) &= 0, \\
\hat{x}(t) &= \underset{x \in \mathbb{R}^n}{\operatorname{argmin}}\ m(x,t).
\end{aligned} \tag{6.33}
$$

This system is of the form (3.2) with state space $\mathcal{M} = C(\mathbb{R}^n, \mathbb{R})$, and is controlled by the observation path $y(\cdot)$ of (6.32).

Theorem 6.3 *If the system* (6.32) *is observable in finite time, then*

$$\hat{x}(t) = x(t)$$

for all $t > T$ *and all* $x_0 \in \mathbb{R}^n$. *Therefore* (6.33) *is an observer for* (6.32).

Proof: We have

$$m(x,t) = \inf_{\theta \in \Omega^n} \left\{ \frac{1}{2} \int_0^t \mid \dot{\theta}_s - f(\theta_s) \mid^2 + \mid y_s - h(\theta_s) \mid^2 - \mid y_s \mid^2 ds \ : \theta_t = x \right\}.$$

Then, for $t > T$, $m(\cdot, t)$ has a unique minimum $\hat{x}(t) = x(t)$ if and only if (6.32) is observable on $[0,T]$. Note that (3.3) is satisfied a fortiori. ∎

Remark This type of observer can be defined for systems evolving on manifolds with manifold–valued observations. This and other developments will be treated elsewhere. ///

Chapter 7
Miscellaneous Topics

We collect in this final chapter some miscellaneous comments and results. A large deviation principle for the distributions of the conditional measures is presented in Section 7.1. Thus there is a second level of large deviation behaviour associated with our asymptotic filtering problem. The Beneš class of systems, for which finite dimensional filters exist, is treated in Section 7.2. Then in Section 7.3 we obtain the different limits that result when the signal to noise ratio is altered. Finally in Section 7.4 we make some concluding remarks.

7.1 Distributions of Conditional Measures

Let π_t^ϵ be the conditional probability measures defined by (2.9), (2.10), viewed as random elements of $\mathcal{P}(\mathbb{R}^n)$ (2.11). The distribution of π_t^ϵ is a probability measure $P_{\pi_t}^\epsilon$ on $\mathcal{P}(\mathbb{R}^n)$ defined by

$$P_{\pi_t}^\epsilon(A) \;=\; P\left(\pi_t^\epsilon \in A\right) \quad (A \in \mathcal{B}\left(\mathcal{P}(\mathbb{R}^n)\right)). \tag{7.1}$$

In this section we establish a LDP for $\{P_{\pi_t}^\epsilon\}$.

Recall the maps

$$\bar{\pi}_t^\epsilon, \bar{\pi}_t : \Omega_0 \to \mathcal{P}(\mathbb{R}^n),$$

where $\bar{\pi}_t^\epsilon$ is the pathwise filter (2.20), (2.21) and $\bar{\pi}_t$ is the deterministic estimator (5.13); assumed well defined. Recall that $\{P_Y^\epsilon\}$ obey the LDP with action function I_Y defined by (2.40). For $\nu \in \mathcal{P}(\mathbb{R}^n)$, define

$$I_{\pi_t}(\nu) \;=\; \inf\left\{I_Y(\eta) : \eta \in \bar{\pi}_t^{-1}(\nu)\right\}. \tag{7.2}$$

Theorem 7.1 *Under the above assumptions, $\{P^{\epsilon}_{\pi_t}\}$ obey the LDP with action function $I_{\pi_t}(\nu)$.*

Proof: Noting that

$$P^{\epsilon}_{\pi_t} = P^{\epsilon}_{Y} \circ (\bar{\pi}^{\epsilon}_t)^{-1},$$

and since $\bar{\pi}^{\epsilon}_t$, $\bar{\pi}_t$ are continuous and $\lim_{\epsilon \to 0} \bar{\pi}^{\epsilon}_t = \bar{\pi}_t$ exists uniformly on compact subsets of Ω_0, we can apply the contraction principle to deduce the assertion. ∎

Notice that

$$I_{\pi_t}(\nu) = \inf \left\{ W(\hat{x}(t), t)[\eta] + \frac{1}{2} \int_0^t \dot{\eta}_s^2 ds : \eta \in \Omega_0^{H^1} \cap \bar{\pi}_t^{-1}(\nu) \right\}. \tag{7.3}$$

7.2 The Beneš Class

In [3], Beneš introduced a class of systems, with nonlinear drift and linear observations, for which the conditional density can be explicitly computed by a finite set of ordinary and stochastic differential equations. That is, the stochastic filter is finite dimensional. Hijab [25] has studied this class, as well as the analogue for deterministic estimation. In this section we apply the asymptotic filtering method to the Beneš class of systems. Under certain (restrictive) conditions, an observer results in the limit.

We are interested in systems of the form

$$\begin{aligned} \dot{x} &= f(x), \quad x(0) = x_0, \\ \dot{z} &= Cx, \quad z(0) = 0, \end{aligned} \tag{7.4}$$

where $x(t) \in I\!R^n$ and $z(t) \in I\!R^p$. Consider the filtering problem

$$dx^{\epsilon}(t) = f(x^{\epsilon}(t))dt + \sqrt{\epsilon} N dw(t), \quad x^{\epsilon}(0) = x^{\epsilon}_0, \tag{7.5}$$

$$dy^\epsilon(t) \;=\; Cx^\epsilon(t)dt + \sqrt{\epsilon}Rdv(t), \quad y^\epsilon(0) = 0,$$

where $NN' \geq r_0 I, \; R > 0$.

The *Beneš class* of drifts $f \in C^\infty(I\!R^n, I\!R^n)$ are those satisfying:

(i) for some $F \in C^\infty(I\!R^n, I\!R)$,

$$f(x) \;=\; -NN'DF(x)';$$

(ii) for all $\epsilon \geq 0$,

$$f(x)'(NN')^{-1}f(x) + \epsilon \mathrm{div} f(x) \;=\; x'\Sigma^\epsilon x + 2\Xi x + 2\varphi^\epsilon,$$

where $\Sigma^\epsilon \to \Sigma, \; \Xi^\epsilon \to \Xi, \; \varphi^\epsilon \to \varphi$ as $\epsilon \to 0$, and Σ^ϵ, Σ are symmetric.

Two additional *detectability* conditions are required:

(iii) $$D^2F(x) \;\geq\; 0;$$

(iv) for all $\epsilon \geq 0$,

$$C'(RR')^{-1}C + \Sigma^\epsilon \;\geq\; \alpha_0 I, \quad \alpha_0 > 0.$$

The initial condition x_0 is unknown, so we let x_0^ϵ be random with density $p_0^\epsilon(x) = C_\epsilon \exp(-\frac{1}{\epsilon}S_0(x))$, where

$$S_0(x) \;=\; \frac{1}{2}(x-\mu_0)'P_0(x-\mu_0) + F(x),$$

for some $\mu_0 \in I\!R^n, P_0 > 0$. As before, P_0, N, R are design parameters, although N is restricted according to (i) and (ii) above.

Let $q^\epsilon(x,t)$ denote the unnormalised conditional density solving the Zakai equation. We use the exponential representation:

$$q^\epsilon(x,t) \;=\; e^{-\frac{1}{\epsilon}W^\epsilon(x,t)}. \tag{7.6}$$

Using Itô's rule, $W^\epsilon(x,t)$ solves

$$\begin{aligned}
\partial W^\epsilon - \frac{\epsilon}{2}\mathrm{tr}(NN'D^2W^\epsilon)dt + \frac{1}{2} \mid DW^\epsilon N \mid^2 dt + DW^\epsilon \cdot f(x)dt & \\
-\frac{1}{2} \mid R^{-1}Cx \mid^2 dt + x'C'(RR')^{-1}dy^\epsilon(t) - \epsilon \mathrm{div} f(x)dt &= 0, \qquad (7.7)\\
W^\epsilon(x,0) &= S_0(x) - \epsilon \log C_\epsilon.
\end{aligned}$$

We now find an explicit solution for $q^\epsilon(x,t)$ via (7.6), (7.7) using a gauge transformation.

Define

$$U^\epsilon(x,t) = W^\epsilon(x,t) - F(x).$$

Then we have

$$\begin{aligned}
\partial U^\epsilon - \frac{\epsilon}{2}\mathrm{tr}(NN'D^2U^\epsilon)dt + \frac{1}{2} \mid DU^\epsilon N \mid^2 dt - \Xi^\epsilon x dt + x'C'(RR')^{-1}dy^\epsilon(t) & \\
-\frac{1}{2}x'(C'(RR')^{-1}C + \Sigma^\epsilon)x dt - \varphi^\epsilon dt &= 0,\\
U^\epsilon(x,0) &= \frac{1}{2}(x-\mu_0)'P_0(x-\mu_0) - \epsilon \log C_\epsilon.
\end{aligned}$$

If we look for solutions of the form

$$U^\epsilon(x,t) = \frac{1}{2}(x-\mu^\epsilon(t))'P^\epsilon(t)(x-\mu^\epsilon(t)) + \psi^\epsilon(t),$$

we obtain, after some manipulation,

$$\begin{aligned}
d\mu^\epsilon(t) &= -P^\epsilon(t)^{-1}(\Xi^\epsilon + \Sigma^\epsilon \mu^\epsilon(t))dt \qquad (7.8)\\
&\quad + P^\epsilon(t)^{-1}C'(RR')^{-1}(dy^\epsilon(t) - C\mu^\epsilon(t)dt),\\
\dot{P}^\epsilon(t) &= -P^\epsilon(t)NN'P^\epsilon(t) + C'(RR')^{-1}C + \Sigma^\epsilon, \qquad (7.9)\\
\psi^\epsilon(t) &= -\frac{1}{2}\mu^\epsilon(t)'P^\epsilon(t)\mu^\epsilon(t) + \frac{1}{2}\mu_0'P_0\mu_0 - \epsilon \log C_\epsilon \qquad (7.10)\\
&\quad + \int_0^t \left(\frac{\epsilon}{2}\mathrm{tr}(NN'P^\epsilon(s)) - \frac{1}{2}\mu^\epsilon(s)'P^\epsilon(s)NN'P^\epsilon(s)\mu^\epsilon(s) + \varphi^\epsilon\right) ds,
\end{aligned}$$

with initial conditions

$$\mu^\epsilon(0) = \mu_0, \quad P^\epsilon(0) = P_0. \qquad (7.11)$$

Then one can prove:

Under assumptions (i)–(iv), the unnormalised conditional density $q^\epsilon(x,t)$ has the explicit solution (7.6), *where*

$$W^\epsilon(x,t) \;=\; \frac{1}{2}(x-\mu^\epsilon(t))'P^\epsilon(t)(x-\mu^\epsilon(t))+\psi^\epsilon(t)+F(x),$$

and $P^\epsilon(t),\mu^\epsilon(t),\psi^\epsilon(t)$ are the solutions of (7.8)–(7.11).

A filter for the *maximum likelihood estimate*

$$m^\epsilon(t) \;=\; \underset{x\,\in\,\mathbb{R}^n}{\operatorname{argmin}}\; W^\epsilon(x,t)$$

can be obtained by differentiating the equality: $DW^\epsilon(m^\epsilon(t),t)=0$ (c.f. Chapter 4). In general $m^\epsilon(t)$ does not equal the conditional mean $\hat{x}^\epsilon(t)$.

We now turn to the limiting filter: the deterministic estimator for the observation record $Z=\{z(t):0\le t\le T\}$. Here, Z is the limit in probability of Y^ϵ, given by (7.4), and is viewed as an observation record of the auxilliary system

$$\begin{aligned}\dot{\tilde{x}} &= f(\tilde{x})+Nu, \quad \tilde{x}(0)=\tilde{x}_0, \\ \dot{y} &= C\tilde{x}+Rv, \quad y(0)=0,\end{aligned} \tag{7.12}$$

with cost

$$\begin{aligned}J_t(\tilde{x}_0,u) \;=\;& \frac{1}{2}(\tilde{x}_0-\mu_0)'P_0(\tilde{x}_0-\mu_0) \\ &+\int_0^t \frac{1}{2}\mid u(s)\mid^2+\frac{1}{2}\mid R^{-1}C\tilde{x}(s)\mid^2-\tilde{x}(s)'C'(RR')^{-1}\dot{y}(s)ds.\end{aligned} \tag{7.13}$$

Conditions (ii) and (iv) reduce to

(ii)' $$f(x)'(NN')^{-1}f(x) \;=\; x'\Sigma x+2\Xi x+2\varphi;$$

(iv)' $$C'(RR')^{-1}C+\Sigma \;\ge\; \alpha_0 I.$$

The value function $W(x,t)$ satisfies

$$\frac{\partial}{\partial t}W + \frac{1}{2} \mid DWN \mid^2 + DW \cdot f(x) - \frac{1}{2} \mid R^{-1}Cx \mid^2 + x'C'(RR')^{-1}\dot{z}(t) = 0,$$
$$W(x,0) \;=\; S_0(x), \tag{7.14}$$

and is given explicitly by

$$W(x,t) \;=\; \frac{1}{2}(x-\mu(t))'P(t)(x-\mu(t)) + \psi(t) + F(x), \tag{7.15}$$

where $P(t), \mu(t)$ and $\psi(t)$ are the solutions of

$$\dot{\mu}(t) \;=\; -P(t)^{-1}(\Xi + \Sigma\mu(t)) + P(t)^{-1}C'(RR')^{-1}(\dot{z}(t) - C\mu(t)), \tag{7.16}$$
$$\dot{P}(t) \;=\; -P(t)NN'P(t) + C'(RR')^{-1}C + \Sigma, \tag{7.17}$$
$$\begin{aligned}\psi(t) \;=\; & -\frac{1}{2}\mu(t)'P(t)\mu(t) + \frac{1}{2}\mu_0'P_0\mu_0 \\ & + \int_0^t \left(-\frac{1}{2}\mu(s)'P(s)NN'P(s)\mu(s) + \varphi\right) ds,\end{aligned} \tag{7.18}$$

with initial conditions

$$\mu(0) \;=\mu_0, \quad P(0) \;=\; P_0. \tag{7.19}$$

Then $W^\epsilon(x,t) \to W(x,t)$ in probability as $\epsilon \to 0$.

Recall the deterministic estimate

$$m(t) \;=\; \underset{x \in I\!R^n}{\operatorname{argmin}}\; W(x,t)$$

(Previously we used the notation $\hat{x}(t)$.)

Theorem 7.2 *Under assumptions (i), (ii)', (iii) and (iv)', the deterministic estimator for the system* (7.4) *is well defined and can be computed by the finite dimensional filter*

$$\dot{m}(t) \;=\; f(m(t)) + \tilde{P}(t)^{-1}C'(RR')^{-1}(\dot{z}(t) - Cm(t)), \quad m(0) = m_0, \tag{7.20}$$
$$\tilde{P}(t) \;=\; P(t) + D^2F(m(t)), \tag{7.21}$$

where $P(t)$ is the solution of (7.17). Further, the value function can be explicitly computed as above.

Proof: Let $Q(t) = P(t)^{-1}$, which solves

$$\dot{Q}(t) = -Q(t)\left(C'(RR')^{-1}C + \Sigma\right)Q(t) + NN'.$$

Using the methods in Section 6.2, we have

$$\begin{aligned} \| P(t) \| &\leq \| P_0 \| + \frac{\| N \|^2 + \| R^{-1}C \|^2}{r_0} \equiv p, \\ \| Q(t) \| &\leq \| Q_0 \| + \frac{\| N \|^2 + \| C'(RR')^{-1}C + \Sigma \|}{2\alpha_0} \equiv q, \end{aligned}$$

for all $t \geq 0$. To obtain the bound for $P(t)$, one uses the control $v(t) = N\lambda_t$ in (6.20) with $A_t \equiv 0$. The bound for $Q(t)$ follows from the detectability condition (iv)': setting $\Lambda = -(C'(RR')^{-1}C + \Sigma)^{\frac{1}{2}}$, we have

$$\eta'\left(\Lambda(C'(RR')^{-1}C + \Sigma)\right)\eta \leq -\alpha_0 \mid \eta \mid^2.$$

Now

$$\begin{aligned} D^2W(x,t) &= P(t) + D^2F(x) \\ &\geq \frac{1}{q}I > 0. \end{aligned}$$

Hence $W(\cdot, t)$ is strictly convex, and therefore has a unique minimum. ∎

Finally, we present a limited result on whether or not the deterministic estimator is an observer for (7.4). Assume further

(iii)' $\qquad D^2F(x) \geq \beta_0 I$, for some $\beta_0 > 0$.

Define

$$\begin{aligned} \delta(P_0, N, R) &= \frac{r_0}{q^2} - 2p \| Df \| - \| \Sigma \| \\ &\quad -2q \| C \|^2 \| R^{-1} \|^2 \left(q + \| (D^2F)^{-1} \|\right). \end{aligned}$$

Theorem 7.3 *Under assumptions (i), (ii)'-(iv)', if there exists* P_0, N, R *such that*

$$\delta(P_0, N, R) \; > \; 0, \tag{7.22}$$

then the system (7.20), (7.21), (7.17) *is an observer for the dynamical system* (7.4)*. That is, there exists constants* $K > 0, \gamma > 0$ *such that*

$$| \, x(t) - m(t) \, | \leq \; K \, | \, x_0 - m_0 \, | \, e^{-\gamma t}$$

for all $t \geq 0$ *and all* $x_0, m_0 \in I\!R^n$*.*

Proof: We have

$$\begin{aligned} \frac{d}{dt} e(t)' P(t) e(t) \;&=\; 2e(t)'P(t) \left(f(x(t)) - f(m(t)) - \tilde{P}(t)^{-1} C'(RR')^{-1} C e(t) \right) \\ &\quad + e(t)' \left(-P(t) N N' P(t) + C'(RR')^{-1} C + \Sigma \right) e(t). \end{aligned}$$

Now using Kailath [32], p656, we have

$$\begin{aligned} \tilde{P}(t)^{-1} \;&=\; [P(t) + D^2 F(m(t))]^{-1} \\ &=\; Q(t) - Q(t) \left(Q(t) + (D^2 F(m(t)))^{-1} \right) Q(t). \end{aligned}$$

Using this identity in the above equality and estimating the various terms, one proves the theorem. ∎

Remark In general, it is not clear that condition (7.22) is verified. Notice that the bounds p, q are bounded functions of $\| \; R \; \| \rightarrow \infty$, so the last terms in $\delta(P_0, N, R)$ can be made small. From another point of view, under our assumptions (7.4) is asymptotically stable, and the observer (7.20), (7.21) can be viewed as a perturbation. ///

7.3 Limits for Other Signal to Noise Ratios

In the Introduction we introduced the family of filtering problems (1.4) with noise scaling $\sqrt{\epsilon}$ in the state equation and $\sqrt{\delta}$ in the observation equation. The quantity ϵ/δ is a measure of the signal to noise ratio or the relative rate at which the noise goes to zero. We now describe the different limiting behaviour that occurs as this ratio is altered.

Specifically, consider the filtering problems

$$\begin{aligned} dx^\epsilon(t) &= f(x^\epsilon(t))dt + \epsilon^\gamma dw(t), \quad x^\epsilon(0) = x_0^\epsilon, \\ dy^\epsilon(t) &= h(x^\epsilon(t))dt + \epsilon^{1-\gamma} dv(t), \quad y^\epsilon(0) = 0, \end{aligned} \tag{7.23}$$

where $0 < \gamma < 1$. The signal to noise ratio (SNR) is $\epsilon^{2\gamma-1}$, and note that $-1 < 2\gamma - 1 < 1$. Bensoussan [4] treated the case $0 \leq \gamma < 1/2$, Picard [51] and Ji [31] the case $\gamma = 0$, and Hijab [25], [26] the case $\gamma = 1/2$. In [4], [51] and [31] their main concern was the construction of approximate filters.

In Chapter 2 we defined the notion of large deviations for the specific normalisation factor ϵ^{-1}. More generally, as in Freidlin and Wentzell [20], we say that $\{\mathsf{P}^\epsilon\}$ obey the LDP with action function I and normalisation coefficient $\lambda(\epsilon)$ provided (i) and (ii) hold in Varadhan's definition (Section 2.2.1) and

$$\lim_{\epsilon \to 0} \lambda(\epsilon)^{-1} \log \mathsf{P}^\epsilon(A) \asymp e^{-\lambda(\epsilon) \inf_{\mathrm{x} \in A} I(\mathrm{x})}$$

in the sense of (iii) and (iv) in that definition. Here, $\lambda(\epsilon) \to \infty$ as $\epsilon \to 0$. Similarly we can modify the definition for random measures $\{\mathsf{Q}^\epsilon\}$.

We take x_0^ϵ to have density $p_0^\epsilon(x) = C_\epsilon \exp(-\frac{1}{\epsilon^{2\gamma}} S_0(x))$, with S_0 defined as in Section 2.1. The trajectories of (7.23) converge in probability to the trajectories of the corresponding deterministic system (2.35), (5.1). It is clear that the distributions $\{P_X^\epsilon\}$ obey the LDP with action function $I_X(\theta)$ defined by (2.38) and normalisation coefficient $\lambda(\epsilon) = \epsilon^{-2\gamma}$. It turns out that three types of limiting behaviour occur for the filtering problem, depending on the value of the parameter γ:

(a) $0 < \gamma < 1/2$ (high SNR),

(b) $\gamma = 1/2$ (balanced SNR),

(c) $1/2 < \gamma < 1$ (low SNR).

The Kallianpur–Streibel formula for the conditional measures takes the form

$$\begin{aligned}
\Pi^{\epsilon,\gamma}(A) &= \frac{\Sigma^{\epsilon,\gamma}(A)}{\Sigma^{\epsilon,\gamma}(\Omega^n)} \quad P\text{–a.s.}, \\
\Sigma^{\epsilon,\gamma}(A) &= \int_A \exp\left(-\frac{1}{\epsilon^{2-2\gamma}}\left[\frac{1}{2}\int_0^T h(\theta_s)^2 ds - \int_0^T h(\theta_s) dy_s^\epsilon\right]\right) P_X^\epsilon(d\theta) \quad P\text{–a.s.},
\end{aligned}$$

for $A \in \mathcal{B}(\Omega^n)$. For $\theta \in \Omega^n$ define $F(\theta)$ by (5.44). Define

$$\lambda^\gamma(\epsilon) = \begin{cases} \frac{1}{\epsilon^{2-2\gamma}} & \text{if } 0 < \gamma \le 1/2, \\ \frac{1}{\epsilon^{2\gamma}} & \text{if } 1/2 \le \gamma < 1; \end{cases}$$

$$\breve{J}^\gamma(\theta) = \begin{cases} F(\theta) & \text{if } 0 < \gamma < 1/2, \\ I(\theta) + F(\theta) & \text{if } \gamma = 1/2, \\ I(\theta) & \text{if } 1/2 < \gamma < 1. \end{cases}$$

Also, we define

$$J^\gamma(\theta) = \breve{J}^\gamma(\theta) - \inf_{\theta' \in \Omega^n} \breve{J}^\gamma(\theta').$$

If $\theta \in \Omega^n$ is absolutely continuous,

$$J^\gamma(\theta) = \begin{cases} \frac{1}{2}\int_0^T (h(\theta_s) - \dot{z}_s)^2 ds & \text{if } 0 < \gamma < 1/2, \\ S_0(\theta_0) + \frac{1}{2}\int_0^T \mid \dot{\theta}_s - f(\theta_s) \mid^2 + (h(\theta_s) - \dot{z}_s)^2 ds & \text{if } \gamma = 1/2, \\ S_0(\theta_0) + \frac{1}{2}\int_0^T \mid \dot{\theta}_s - f(\theta_s) \mid^2 ds & \text{if } 1/2 < \gamma < 1. \end{cases}$$

Theorem 7.4 $\{\Pi^{\epsilon,\gamma}\}$ *obey the LDPP with action function $J^\gamma(\theta)$ and normalisation coefficient $\lambda^\gamma(\epsilon)$.*

The proof of this theorem depends on the following lemma.

Lemma 7.2 *If $C \subset \Omega^n$ is closed and $G \subset \Omega^n$ is open, we have:*

$$\begin{aligned}
\limsup_{\epsilon\to 0} \lambda^\gamma(\epsilon) \log \Sigma^{\epsilon,\gamma}(C) &\leq -\inf_{\theta\in C} \breve{J}^\gamma(\theta) \quad \text{in probability,} \\
\liminf_{\epsilon\to 0} \lambda^\gamma(\epsilon) \log \Sigma^{\epsilon,\gamma}(G) &\geq -\inf_{\theta\in G} \breve{J}^\gamma(\theta) \quad \text{in probability.}
\end{aligned}$$

Proof: The proof is similar to that of Theorem 5.4, so we only sketch some of the changes required.

The basic idea is to notice that

$$\epsilon^{2-2\gamma} \log \Sigma^{\epsilon,\gamma}(C) \;\leq\; -\inf_{\theta\in C} F(\theta) + \epsilon^{2-4\gamma}\epsilon^{2\gamma} \log P_X^\epsilon(C) + \delta,$$

and if $0 < \gamma < 1/2$ then the second term on the right goes to zero. Similarly,

$$\epsilon^{2\gamma} \log \Sigma^{\epsilon,\gamma}(C) \;\leq\; -\epsilon^{4\gamma-2} \inf_{\theta\in C} F(\theta) + \epsilon^{2\gamma} \log P_X^\epsilon(C) + \delta,$$

and if $1/2 < \gamma < 1$, then the first term goes to zero.

These ideas can be made precise along the lines of Theoerm 5.4 and Lemma 5.5. ∎

If $\gamma = 0$, then the state equation is independent of ϵ. Then the resulting action function is *random*, given by, for $\theta \in \Omega^n$ absolutely continuous,

$$J^0(\theta)(\omega) \;=\; \frac{1}{2}\int_0^T \left(h(\theta_s) - h(x_s^1(\omega))\right)^2 ds,$$

with normalisation coefficient $\lambda^0(\epsilon) = \epsilon^{-2}$. This result was obtained by Ji [31].

At the other extreme, if $\gamma = 1$, then $y^\epsilon(t) \to y^0(t)$, where

$$y^0(t)(\omega) \;=\; \int_0^t h(x(s))ds + v(t)(\omega).$$

Here, $x(t)$ is the deterministic trajectory. In this case, Theorem 7.4 remains valid.

In the cases $1/2 \leq \gamma \leq 1$ we have:

$$\Pi^{\epsilon,\gamma} \overset{P}{\Rightarrow} \delta_X,$$

as $\epsilon \to 0$, where $X \in \Omega^n$ denotes the deterministic trajectory. If $0 < \gamma < 1/2$ and h is injective, then

$$\Pi^{\epsilon,\gamma} \overset{P}{\Rightarrow} \delta_X,$$

as $\epsilon \to 0$. However, when $\gamma = 0$ and h is injective the limit is random:

$$\Pi^{\epsilon,0} \overset{P}{\Rightarrow} \delta_{X^1}$$

as $\epsilon \to 0$. This injectivity condition was used in Picard [51] and Ji [31].

7.4 Concluding Remarks

We have seen that the asymptotic filtering approach leads to the suggestion that the limiting filter, identified as the deterministic estimator, is a candidate observer (Baras and Krishnaprasad [1]). Indeed, this approach has motivated two observer designs: a finite dimensional observer for uniformly detectable systems, and an infinite dimensional observer for observable systems. The former estimates the state asymptotically as $t \to \infty$, while the latter computes the state exactly after a finite time has elapsed.

A gap in the theory is the lack of simpler conditions implying detectability and observability. The principal disadvantages of our finite dimensional design involve the computation of the matrix–valued function $\Lambda(x, u)$ and the sensitive dependence of the radius of convergence and decay rate on the design parameters. The infinite dimensional observer presents some interesting computational problems, akin to those arising in nonlinear filtering.

There are a number of interesting questions regarding the interactions between the limits as $\epsilon \to 0$ and $t \to \infty$ that might have a bearing on our problem. Kunita [36] has studied the large time behaviour of nonlinear filtering errors in the case that the signal state space X is compact. Ji [31] has extended some of Kunita's results to noncompact state spaces. The conditional measures $\{\pi_t\}$ are viewed as a Markov process with values in $\mathcal{P}(\mathsf{X})$, and the filtering error is computed in terms of invariant measures of the processes $\{x_t\}$, $\{\pi_t\}$. Large deviation results for invariant measures of diffusions in $I\!R^n$ have been obtained by Freidlin and Wentzell [20] (for example Theorem 4.3, page 129). In view of the LDP for the distributions of $\{\pi_t^\epsilon\}$ obtained in Section 7.1, one might find some large deviation result for invariant measures of the filtering process $\{\pi_t^\epsilon\}$. Detectability and observability may have implications in filtering. Detectability is connected with the positive definiteness of the Hessian of the value function. In the context of extended Kalman filtering, Mitter [45] alluded to this. These remain unresolved issues in our investigations.

It is our opinion that "approximate" estimators, an example of which was presented in Chapter 6, are more likely to be useful in practice than "exact" estimators. Further, in practical situations the structure and symmetry of the problem at hand should be fully exploited. Like the Kalman filter, both of our designs seem to enjoy robustness properties. An approximation to the infinite dimensional observer may be useful in providing an initial estimate for the finite dimensional asymptotic observer.

It seems difficult to build a general and easily usable theory of observer design

design for nonlinear control systems. The asymptotic filtering approach provides insight and suggests potentially useful observer designs.

Bibliography

[1] J. S. Baras and P. S. Krishnaprasad, *Dynamic Observers as Asymptotic Limits of Recursive Filters*, IEEE Proc. 21st CDC, Orlando, Florida, Dec., 1982, 1126–1127.

[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. To appear, *SIAM J. Applied Math.*

[3] V. E. Beneš, *Exact Finite–Dimensional Filters for Certain Diffusions with Nonlinear Drift*, Stochastic, 5 (1981), 65–92.

[4] A. Bensoussan, *On Some Approximation Techniques in Nonlinear Filtering*, preprint.

[5] P. Billingsley, *Weak Convergence of Measures: Application in Probability*, CBMS–NSF Reg. Conf. Series in Appl. Math., SIAM, Philadelphia, 1971.

[6] M. G. Crandall and P. L. Lions, *Viscosity Solutions of Hamilton–Jacobi Equations*, Trans. AMS, 277(1) (1983), 1–42.

[7] M. G. Crandall and P. L. Lions, *Two Approximations of Solutions of Hamiltion– Jacobi Equations*, Math. Comput., 43(167) (1984), 1–19.

[8] M. G. Crandall and P. L. Lions, *Remarks on the Existence and Uniqueness of Unbounded Viscosity Solutions of Hamilton–Jacobi equations*, MRC Technical Summary Report 2875, University of Wisconsin–Madison, October 1985.

[9] 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.

[10] M. H. A. Davis, *Linear Estimation and Stochastic Control*, Chapman and Hall, London, 1977.

[11] M. H. A. Davis, *On a Multiplicative Functional Transformation Arising in Nonlinear Filtering Theory*, Z. Wahrsch. verw. Gebiete, 54, (1980) 125–139.

[12] I. Derese, P. Stevens and E. Noldus, *The design of state observers for bilinear systems*, Journal A, 20(4) (1979) 193–202.

[13] R. J. Elliott, *Stochastic Calculus and Applications*, Springer Verlag, New York, 1982.

[14] R. S. Ellis, *Entropy, Large Deviations, and Statistical Mechanics*, Springer Verlag, New York, 1985.

[15] S. N. Ethier and T. G. Kurtz, *Markov Processes: Characterization and Convergence*, Wiley, New York, 1986.

[16] 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(1) (1985), 1–20.

[17] W. H. Fleming, *Exit Probabilities and Optimal Stochastic Control*, Appl. Math. Optim., 4 (1978), 329–346.

[18] W. H. Fleming and S. K. Mitter, *Optimal Control and Nonlinear Filtering for Nondegenerate Diffusion Processes*, Stochastics, 8 (1982), 63–77.

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

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

[21] M. I. Freidlin, *Functional Integration and Partial Differential Equations*, Princeton University Press, 1985.

[22] M. I. Freidlin, *Limit Theorems for Large Deviations and Reaction–Diffusion Equations*, Annals of Prob., 13(3) (1985), 639–675.

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

[24] R. Hermann and A. J. Krener, *Nonlinear Controllability and Observability*, IEEE Trans. AC., 22(5) (1977) 728–740.

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

[26] O. Hijab, *Asymptotic Bayesian Estimation of a First Order Equation with Small Diffusion*, Annals of Probability, 12 (1984), 890–902.

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

[28] A. Isidori, *Nonlinear Control Systems: An Introduction*, Springer Verlag, 1985.

[29] M. R. James and J. S. Baras, *Nonlinear Filtering and Large Deviations: A PDE–Control Theoretic Approach*, Stochastics, 23(3) (1988), 391–412.

[30] M. R. James and J. S. Baras, *An Observer Design for Nonlinear Control Systems*, Technical Report SRC–TR–87–193, Systems Research Center, University of Maryland, October 1987. To appear, *8th Int. Conf. Analysis and Optimization of Systems*, Antibes, June 1988.

[31] D. Ji, *Asymptotic Analysis of Nonlinear Filtering Problems*, Ph.D. Dissertation, Brown University, May 1987.

[32] T. Kailath, *Linear Systems*, Prentice–Hall, Englewood Cliffs, NJ, 1980.

[33] G. Kallianpur and C. Striebel, *Estimation of Stochastic Systems: Arbitrary System Process with Additive White Noise Observation Errors*, Annals Math. Stat., 39(3) (1968) 785–801.

[34] G. Kallianpur, *Stochastic Filtering Theory*, Springer Verlag, New York, 1980.

[35] R. E. Kalman, *Contributions to the Theory of Optimal Control*, Bol. Soc. Mat. Mex., (1960), 102–119.

[36] H. Kunita, *Asymptotic Behaviour of the Nonlinear Filtering Errors of Markov Processes*, J. Multivariate Anal., 1 (1971), 365–393.

[37] D. L. Kleinman, *An Easy Way to Stabilize a Linear Constant System*, IEEE Trans. Aut. Control, AC–15 (1970), 692.

[38] A. J. Krener, *Minimum Covariance, Minimax and Minimum Energy Estimators*, in *Stochastic Control Theory and Stochastic Differential systems*, M. Kohlmann and W. Vogel (eds.), Berlin, Springer–Verlag, 1979, 490–495.

[39] A. J. Krener and W. Respondek, *Nonlinear Observers with Linearizable Error Dynamics*, SIAM J. Cont. Opt., 23 (2) (1985) 197–216.

[40] S. R. Kuo, D. L. Elliott, and T. J. Tarn, *Exponential Observers for Nonlinear Dynamic Systems*, Inform. Cont., 29 (1975) 204–216.

[41] H. J. Kushner, *A Robust Discrete State Approximation to the Optimal Nonlinear Filter for a Diffusion*, Stochastics, 3 (1979), 75–83.

[42] J. Levine and R. Marino, *Nonlinear system immersion, observers and finite dimensional filters*, System and Control Letters, 7 (1986), 133–142.

[43] R. S. Liptser and A. N. Shiryayev, *Statistics of Random Processes I: General Theory*, Springer Verlag, New York, 1977.

[44] D. G. Luenberger, *Observers for Multivariable Systems*, IEEE Trans. Aut. Control, AC–11 (1966), 190–199.

[45] S. K. Mitter, *Approximations for Nonlinear Filtering*, NATO Adv. Studies Inst. on Nonlinear Stochastic Systems, Algarve, Portugal, 1982.

[46] S. K. Mitter, *Nonlinear Filtering and Stochastic Systems: The Mathematics of Filtering and Identification and Applications*, Proc. NATO Adv. Studies Inst., Les Arcs, France, 1980. D. Reidel, Dordrecht.

[47] S. K. Mitter, *On the Analogy Between Mathematical Problems of Nonlinear Filtering and Quantum Physics*, Ricerche di Automatica, 10 (2) (1979) 163–216.

[48] R. E. Mortensen, *Maximum–Likelihood Recursive Nonlinear Filtering*, J. Opt. Theory and Appl., 2(6) (1968), 386–394.

[49] J. O' Reilly, *Observers for Linear Systems*, Academic Press, London, 1983.

[50] E. Pardoux, *Équations du Filtrage non Linéaire de la Prédiction et du Lissage*, Stochastics, 6 (1982), 193–231.

[51] J. Picard, *Filtrage de Diffusions Vectorielles Faiblement Bruitees*, Proc. 7th Int. Conf. on Analysis and Optimisation of Systems, Antibes, 1986, Springer Verlag.

[52] D. W. Stroock and S. R. S. Varadhan, *Multidimensional Diffusion Processes*, Springer Verlag, 1979.

[53] H. J. Sussman, *Continuous Version of the Conditional Statistics of Nonlinear Filtering*, in: *Stochastic Differential Systems: Filtering and Control*, Eds. M. Metivier and E. Pardoux, Springer Verlag, 1985.

[54] S. R. S. Varadhan, *Large Deviations and Applications*, CBMS–NSF Regional Conf. Series in Appl. Math., SIAM, Philadelphia, 1984.

[55] S. R. S. Varadhan, *Asymptotic Probabilities and Differential Equations*, Comm. Pure and Appl. Math., 19 (1966), 261–286.

[56] A. D. Wentzell and M. I. Freidlin, *On Small Random Perturbations Of Dynamical Systems*, Russian Math. Surveys, 25(1) (1970), 1–55.

[57] J. C. Willems and S. K. Mitter, *Controllability, Observability, Pole Allocation, and State Reconstruction*, IEEE Trans. Aut. Cont., AC–16(6) (1971), 582–595.

[58] D. Williamson, *Observation of Bilinear Systems with Application to Biological Control*, Automatica, 13 (1977), 243–254.

[59] W. M. Wonham, *Linear Multivariable Control: a Geometric Approach*, Second Edition, Springer Verlag, 1979.

CURRICULUM VITAE

Name: Matthew Ronald James.

Address: 2303 Lewisdale Drive, Hyattsville, MD 20783.

Degree and date to be conferred: Doctor of Philosophy, 1988.

Date of birth: February 27, 1960.

Place of birth: Sydney, Australia.

Secondary education: Westfields High School, Sydney, Australia.

Collegiate institutions attended:

	Dates	Degree	Date of Degree
University of Maryland, College Park MD.	8/84–5/88	Ph.D.	May 1988.
University of New South Wales, Sydney, Australia.	2/78–11/82	B.E. (Hon. I)	April 1983.
		B.Sc.	April 1981.

Majors: Applied Mathematics and Electrical Engineering.

Professional publications:

L. C. Evans and M. R. James, *The Hamilton–Jacobi–Bellman Equation for Time–Optimal Control*, in preparation.

M. R. James and J. S. Baras, *An Observer Design for Nonlinear Control Systems*, to appear, Proc. 8th Int. Conf. on Analysis and Optimization of Systems, Antibes, June 1988.

M. R. James and J. S. Baras, *Nonlinear Filtering and Large Deviations: A PDE–Control Theoretic Approach*, Stochastics, 23(3) (1988), 391–412.

J. S. Baras, A. Bensoussan and M. R. James, *Dynamic Observers as Asymptotic Limits of Recursive Filters: Special Cases*, to appear, SIAM J. Applied Math.

Professional positions held:

Brown University, Lefschetz Center for Dynamical Systems, Providence RI; Visiting Assistant Professor, from 9/88.

University of Maryland, Systems Research Center, College Park MD; Fellow, 6/85–5/88.

University of Maryland, Department of Mathematics, College Park MD; Teaching Assistant, 9/84–5/85.

Electricity Commission of New South Wales, Sydney, Australia; Electrical Engineer, 12/82–6/84; Cadet Engineer, 2/78–11/82.