Appl Math Optim 32:47–72 (1995)

Applied Mathematics & Optimization

Consistent Parameter Estimation for Partially Observed Diffusions with Small Noise*

M. R. James[1] and F. Le Gland[2]

[1] Department of Systems Engineering, Australian National University, Canberra, ACT 0200, Australia

[2] Institut de Recherche en Informatique et Systèmes Aléatoires, Institut National de Recherche en Informatique et en Automatique, Campus de Beaulieu, 35042 Rennes Cédex, France

Abstract. In this paper we provide a consistency result for the MLE for partially observed diffusion processes with small noise intensities. We prove that if the underlying deterministic system enjoys an identifiability property, then any MLE is close to the true parameter if the noise intensities are small enough. The proof uses large deviations limits obtained by PDE vanishing viscosity methods. A deterministic method of parameter estimation is formulated. We also specialize our results to a binary detection problem, and compare deterministic and stochastic notions of identifiability.

Key Words. Parameter estimation, Nonlinear filtering, Large deviations.

AMS Classification. 62F12, 93E10, 93E11, 60F10.

1. Introduction

In this paper we provide a consistency result for the Maximum Likelihood Estimator (MLE) for partially observed diffusions with small noise.

The problem of computing the MLE for partially observed diffusions has received recent attention. Dembo and Zeitouni [8] have investigated the EM

*This research was supported: by Systems Research Center, University of Maryland through NSF Grant CDR-85-00108 and AFOSR-URI Grant 87-0073; by Lefschetz Center for Dynamical Systems, Division of Applied Mathematics, Brown University, under ARO/MIT Grant DAAL-03-86-K-0171; by INRIA Sophia Antipolis, under ERO/INRIA Grant DAJA45-90-C-0008, and by the CNRS-GR *Automatique*.

algorithm, and Campillo and Le Gland [3] have compared this algorithm with a direct maximization approach. Of course, the goal of such efforts is to compute a good estimate of the unknown parameter. The success or otherwise of such algorithms depends on whether the MLE itself is a good approximation to the unknown parameter. The purpose of this paper is to address this question of consistency when the diffusion and observation noise intensities are *small.*

Our result was in part inspired by some large deviations limit results for nonlinear filtering in [14], [16], and [17]. The theory of large deviations for diffusions with small noise is presented in [12]. We exploit the fact that, on finite time intervals, the diffusion X with observations Y are *close* to a deterministic process x^α with observations y^α. We formulate a deterministic method of parameter estimation for this deterministic process.

We prove that if the underlying deterministic system is *identifiable* and if α is the true parameter, then any MLE $\hat{\theta}^\varepsilon$ is close to α if $\varepsilon > 0$ is small enough. Our proof uses PDE vanishing viscosity methods and Laplace's asymptotic method.

As an application of our results, we study a binary sequential detection problem, discussed in [1], when the noise intensities are small. Deterministic and stochastic notions of identifiability are compared in the context of threshold decision policies.

2. Maximum Likelihood Estimation

On a measurable space $(\Omega, \mathscr{F})$ we consider

- for each $\varepsilon > 0$, a family $\mathscr{M}^\varepsilon = \{P_{\theta,\varepsilon}, \theta \in \Theta\}$ of probability measures,
- a pair of stochastic processes $X \equiv \{X_t, 0 \le t \le T\}$ and $Y \equiv \{Y_t, 0 \le t \le T\}$ taking values in $\mathbf{R}^m$ and $\mathbf{R}^d$, respectively,

such that under $P_{\theta,\varepsilon}$

$$\begin{cases} dX_t = b_\theta(X_t)\,dt + dW_t^{\theta,\varepsilon}, & X_0 \sim p_0^{\theta,\varepsilon}(x)\,dx, \\ dY_t = h_\theta(X_t)\,dt + dV_t^{\theta,\varepsilon}, & Y_0 = 0, \end{cases}$$

where $\{W_t^{\theta,\varepsilon}, 0 \le t \le T\}$ and $\{V_t^{\theta,\varepsilon}, 0 \le t \le T\}$ are independent Wiener processes, with covariance matrices εI_m and εI_d, respectively, and X_0 is a random variable independent of the Wiener processes, with density of the form

$$p_0^{\theta,\varepsilon}(x) \triangleq C_{\theta,\varepsilon} \exp\left\{ -\frac{1}{\varepsilon} S_0^\theta(x) \right\}. \tag{2.1}$$

Note that the coefficients b_θ, h_θ as well as the initial condition depend on the unknown parameter θ.

The set of parameters $\Theta \subset \mathbf{R}^p$ is compact, and the coefficients satisfy the following hypotheses:

(A1) *For all* $\theta \in \Theta$, $b_\theta \in C_b^1(\mathbf{R}^m, \mathbf{R}^m)$ *and* $h_\theta \in C_b^2(\mathbf{R}^m, \mathbf{R}^d)$.

(A2) *For all* $\theta \in \Theta, S_0^\theta$ *is locally Lipschitz continuous, and, for some* $\bar{x}_0^\theta \in \mathbf{R}^m$, $S_0^\theta(\bar{x}_0^\theta) = 0$, $S_0^\theta(x) > 0$ *if* $x \neq \bar{x}_0^\theta$. *Assume also that*

$$C_1 + C_1|x|^2 \geq S_0^\theta(x) \geq C_2|x| - C_2',$$

for all $x \in \mathbf{R}^m$, $\theta \in \Theta$.

Further, the functions b_θ, h_θ, and S_0^θ depend continuously on the parameter θ in the sense that:

(A3) *For each* $\delta > 0$, $R > 0$, *there is* $\gamma > 0$ *such that* $|\theta' - \theta| < \gamma$ *implies*

$$\sup_{x \in \mathbf{R}^m} |b_{\theta'}(x) - b_\theta(x)| < \delta, \qquad \sup_{x \in \mathbf{R}^m} |h_{\theta'}(x) - h_\theta(x)| < \delta,$$

$$\sup_{x \in B(0,R)} |S_0^{\theta'}(x) - S_0^\theta(x)| < \delta.$$

There is no loss in generality in assuming that Ω is the *canonical space* $C([0,T]; \mathbf{R}^{m+d})$, in which case X and Y are the *canonical processes* on $C([0,T]; \mathbf{R}^m)$ and $C([0,T]; \mathbf{R}^d)$, respectively, and $P_{\theta,\varepsilon}$ is the probability law of (X,Y).

It is assumed that only Y is observed. Let $\mathscr{Y}_T$ denote the σ-algebra generated by the process Y on $C([0,T]; \mathbf{R}^d)$. The probability measures in $\mathscr{M}^\varepsilon$ are mutually absolutely continuous, and the *log-likelihood function* for estimating the parameter θ in the statistical model $\mathscr{M}^\varepsilon$ given $\mathscr{Y}_T$, can be expressed (note the minus sign) as

$$-l^\varepsilon(\theta) = \varepsilon \log \mathbf{E}_{\theta,\varepsilon}^\dagger(Z_T^{\theta,\varepsilon} | \mathscr{Y}_T).$$

Here $P_{\theta,\varepsilon}^\dagger$ is a probability measure on Ω, equivalent to $P_{\theta,\varepsilon}$ with Radon–Nikodym derivative

$$Z_T^{\theta,\varepsilon} \triangleq \frac{dP_{\theta,\varepsilon}}{dP_{\theta,\varepsilon}^\dagger} = \exp\left\{\frac{1}{\varepsilon}\int_0^T h_\theta^*(X_s)\, dY_s - \frac{1}{2\varepsilon}\int_0^T |h_\theta(X_s)|^2\, ds\right\},$$

so that under $P_{\theta,\varepsilon}^\dagger$

$$dX_t = b_\theta(X_t)\, dt + dW_t^{\theta,\varepsilon}, \qquad X_0 \sim p_0^{\theta,\varepsilon}(x)\, dx,$$

where $\{W_t^{\theta,\varepsilon}, t \geq 0\}$ and $\{Y_t, t \geq 0\}$ are independent Wiener processes, with covariance matrices εI_m and εI_d, respectively, and the random variable X_0 is independent of the Wiener processes, see [3].

The MLE of the parameter θ in the statistical model $\mathscr{M}^\varepsilon$ is defined on the canonical space $C([0,T]; \mathbf{R}^d)$ by

$$\hat{\theta}^\varepsilon \in \operatorname*{argmin}_{\theta \in \Theta} l^\varepsilon(\theta).$$

The likelihood function can be computed via the Duncan–Mortensen–Zakai (DMZ) equation

$$dp^{\theta,\varepsilon}(x,t) = L^*_{\theta,\varepsilon} p^{\theta,\varepsilon}(x,t)\,dt + \frac{1}{\varepsilon} h^*_\theta(x) p^{\theta,\varepsilon}(x,t)\,dY_t, \quad p^{\theta,\varepsilon}(x,0) = p_0^{\theta,\varepsilon}(x), \tag{2.2}$$

where $L^*_{\theta,\varepsilon}$ is the adjoint operator of the infinitesimal generator $L_{\theta,\varepsilon}$ of the diffusion process X under the probability measure $P_{\theta,\varepsilon}$:

$$L_{\theta,\varepsilon} \triangleq \tfrac{1}{2}\varepsilon \sum_{i,j=1}^{m} \frac{\partial^2}{\partial x_i\,\partial x_j} + \sum_{i=1}^{m} b^i_\theta \frac{\partial}{\partial x_i}.$$

Indeed

$$l^\varepsilon(\theta) = -\varepsilon \log \int_{\mathbf{R}^m} p^{\theta,\varepsilon}(x,T)\,dx. \tag{2.3}$$

The filtering problem is discussed in detail in [9] and [18]. The following lemma is proved in Appendix B.

Lemma 2.1. *The log-likelihood function* $-l^\varepsilon(\theta)$ *depends continuously on the parameter* $\theta \in \Theta$ *a.s.*

Now let θ be fixed. When $\varepsilon \downarrow 0$, the following weak convergence result holds on $C([0,T];\mathbf{R}^{m+d})$:

$$P_{\theta,\varepsilon} \overset{\varepsilon \downarrow 0}{\Longrightarrow} \delta_{(x^\theta, y^\theta)},$$

where (x^θ, y^θ) is given by the deterministic differential system

$$(\Sigma^\theta) \quad \begin{cases} \dot{x}^\theta_t = b_\theta(x^\theta_t), & x^\theta_0 = \bar{x}^\theta_0, \\ \dot{y}^\theta_t = h_\theta(x^\theta_t), & y^\theta_0 = 0. \end{cases}$$

In particular, for all $\theta \in \Theta$, $\delta > 0$,

$$P_{\theta,\varepsilon}\left(\sup_{0 \le t \le T} |Y_t - y^\theta_t| > \delta \right) \xrightarrow{\varepsilon \downarrow 0} 0, \tag{2.4}$$

see [12].

Remark 2.2. As long as $\varepsilon > 0$, the probability measures in $\mathscr{M}^\varepsilon$ are mutually absolutely continuous, which allows us to define the log-likelihood function $-l^\varepsilon(\theta)$. On the other hand, asymptotically when $\varepsilon \downarrow 0$, these probability measures look more and more mutually singular, which, together with an identifiability property of the underlying deterministic system, indicates that the MLE may be consistent. Actually, this result is proved below.

The purpose of the next section is to consider the problem of estimating the unknown parameter θ in the deterministic model $\mathscr{M}^0 = \{(\Sigma^\theta), \theta \in \Theta\}$.

3. Deterministic Parameter Estimation

Consider the family $\mathscr{M}^0 = \{(\Sigma^\theta, \theta \in \Theta\}$ of deterministic differential systems

$$(\Sigma^\theta) \quad \begin{cases} \dot{x}_t^\theta = b_\theta(x_t^\theta), & x_0^\theta = \bar{x}_0^\theta, \\ \dot{y}_t^\theta = h_\theta(x_t^\theta), & y_0^\theta = 0. \end{cases} \tag{3.1}$$

Note that, for all $\theta \in \Theta$, (Σ^θ) describes the weak limit as $\varepsilon \downarrow 0$ of the family of probability measures $\{P_{\theta,\varepsilon}, \varepsilon > 0\}$.

The problem is to estimate the unknown parameter θ on the basis of an observation record, which is supposed to be the output of some deterministic differential system in $\mathscr{M}^0$. Introduce the following definition:

Definition 3.1. The model $\mathscr{M}^0$ is *identifiable* on $[0,T]$ if, for all $\theta' \neq \theta$ in Θ, there is $t \in [0,T]$ such that

$$y_t^{\theta'} \neq y_t^\theta.$$

In other words, the mapping $\theta \mapsto \{y_t^\theta, 0 \le t \le T\}$ is injective. The deterministic parameter estimation problem consists of inverting this mapping. This can be expressed in terms of the following variational problems.

A natural idea is to define on Θ the following criterion:

$$d_\alpha(\theta) \triangleq \tfrac{1}{2}\int_0^T |\dot{y}_t^\alpha - \dot{y}_t^\theta|^2 \, dt \tag{3.2}$$

and to define the *deterministic estimate* (DPE) of the unknown parameter θ in the model $\mathscr{M}^0$ on the basis of the observation record $\{y_t^\alpha, 0 \le t \le T\}$, by

$$\hat{\theta}_\alpha \in M_\alpha \triangleq \operatorname*{argmin}_{\theta \in \Theta} d_\alpha(\theta). \tag{3.3}$$

Obviously

$$M_\alpha = \{\theta \in \Theta : y_t^\alpha = y_t^\theta, \text{ for all } t \in [0,T]\},$$

i.e., M_α is the set of those parameter values that we cannot distinguish, on the basis of the available observation record, from the true value $\alpha \in \Theta$. Notice that $\alpha \in M_\alpha$. If in addition the model $\mathscr{M}^0$ is identifiable, then by definition

$$M_\alpha = \{\alpha\}$$

for all $\alpha \in \Theta$.

However, we define on Θ another criterion $l_\alpha(\theta)$, with the following two properties:

(i) $-l_\alpha(\theta)$ is the limit in some sense of the log-likelihood function $-l^\varepsilon(\theta)$ when $\varepsilon \downarrow 0$, see Lemma 4.2 below.

(ii) The DPE based on $l_\alpha(\theta)$ coincides exactly with the DPE M_α introduced above, which was based on the original criterion $d_\alpha(\theta)$, see Proposition 3.2 below.

Define first the following functional on $C([0,T];\mathbf{R}^m)$:

$$J_\alpha^\theta(\xi,t) \triangleq S_0^\theta(\xi_0) + \tfrac{1}{2}\int_0^t |\dot{\xi}_s - b_\theta(\xi_s)|^2\,ds + \tfrac{1}{2}\int_0^t |\dot{y}_s^\alpha - h_\theta(\xi_s)|^2\,ds - \tfrac{1}{2}\int_0^t |\dot{y}_s^\alpha|^2\,ds, \tag{3.4}$$

if ξ is absolutely continuous, $J_\alpha^\theta(\xi,t) = +\infty$ otherwise. For all $x \in \mathbf{R}^m$ set

$$W_\alpha^\theta(x,t) \triangleq \inf\{J_\alpha^\theta(\xi,t) \colon \xi_t = x\}. \tag{3.5}$$

The value function $W_\alpha^\theta(x,t)$ is continuous in (x,t) and is the unique *viscosity solution* of the Hamilton–Jacobi equation [17]

$$\frac{\partial}{\partial t}W_\alpha^\theta(x,t) + H_\alpha^\theta(x,t,DW_\alpha^\theta(x,t)) = 0, \qquad W_\alpha^\theta(x,0) = S_0^\theta(x), \tag{3.6}$$

where the Hamiltonian $H_\alpha^\theta(x,t,\lambda)$ is defined by

$$\begin{aligned} H_\alpha^\theta(x,t,\lambda) &\triangleq \max_{u\in\mathbf{R}^m}\left\{\lambda^*(b_\theta(x)+u) - \tfrac{1}{2}|u|^2\right\} - \tfrac{1}{2}|\dot{y}_t^\alpha - h_\theta(x)|^2 + \tfrac{1}{2}|\dot{y}_t^\alpha|^2 \\ &= b_\theta^*(x)\lambda + \tfrac{1}{2}|\lambda|^2 + h_\theta^*(x)\dot{y}_t^\alpha - \tfrac{1}{2}|h_\theta(x)|^2. \end{aligned} \tag{3.7}$$

For definitions and an introduction to viscosity solutions of Hamilton–Jacobi equations, the reader is refered to [5] and [6].

Consider then the following alternate criterion, defined on Θ by

$$l_\alpha(\theta) \triangleq \inf_{x\in\mathbf{R}^m} W_\alpha^\theta(x,T) = \inf\{J_\alpha^\theta(\xi,T) \colon \xi \in C([0,T];\mathbf{R}^m)\}. \tag{3.8}$$

The main result of this section is the following:

Proposition 3.2. *For all $\alpha \in \Theta$,*

$$\operatorname*{argmin}_{\theta\in\Theta} l_\alpha(\theta) = M_\alpha,$$

where M_α has been defined in (3.3).

Proof. From (3.4) we have

$$J_\alpha^\theta(\xi,T) \ge c_\alpha \triangleq -\tfrac{1}{2}\int_0^T |\dot{y}_s^\alpha|^2\,ds,$$

for all $\xi \in C([0,T];\mathbf{R}^m)$, and therefore from (3.8) we have

$$l_\alpha(\theta) \geq c_\alpha.$$

From (3.4) we have

$$J_\alpha^\theta(x^\theta, T) = d_\alpha(\theta),$$

for the particular choice $\xi = x^\theta$, and, therefore, from (3.8) we have

$$c_\alpha \leq l_\alpha(\theta) \leq d_\alpha(\theta),$$

for all $\theta \in \Theta$. Therefore

$$M_\alpha \subset \operatorname*{argmin}_{\theta \in \Theta} l_\alpha(\theta).$$

To prove the reverse inclusion, assume that $l_\alpha(\hat{\theta}) = c_\alpha$. Since $J_\alpha^{\hat{\theta}}(\cdot, T)$ is lower semicontinuous

$$l_\alpha(\hat{\theta}) = \inf\{J_\alpha^{\hat{\theta}}(\xi, T) \colon \xi \in C([0,T];\mathbf{R}^m)\} = J_\alpha^{\hat{\theta}}(\hat{\xi}, T),$$

for some $\hat{\xi} \in C([0,T];\mathbf{R}^m)$. Then from (3.4):

(i) $S_0^{\hat{\theta}}(\hat{\xi}_0) = 0$.
(ii) $\dot{\hat{\xi}}_s = b_{\hat{\theta}}(\hat{\xi}_s)$, $0 \leq s \leq T$.
(iii) $\dot{y}_s^\alpha = h_{\hat{\theta}}(\hat{\xi}_s)$, $0 \leq s \leq T$.

From (i) $\hat{\xi}_0 = \bar{x}_0^{\hat{\theta}}$, and therefore, by (ii) and (3.1), $\hat{\xi}_s = x_s^{\hat{\theta}}$, $0 \leq s \leq T$. Then (iii) and (3.1) imply

$$\dot{y}_s^{\hat{\theta}} = h_{\hat{\theta}}\left(x_s^{\hat{\theta}}\right) = \dot{y}_s^\alpha, \qquad 0 \leq s \leq T.$$

Therefore $\hat{\theta} \in M_\alpha$, and the proposition is proved. □

The following lemma is proved in Appendix C.

Lemma 3.3. *For all* $\alpha \in \Theta$:

(i) *There are constants* $C > 0$, $C' > 0$ *such that, for all* $x \in \mathbf{R}^m$, $\theta \in \Theta$,

$$C_1|x|^2 + C \geq W_\alpha^\theta(x,T) \geq C|x| - C',$$

where the constant C_1 *was defined in* (A1).

(ii) *For all* $R > 0$, $\delta > 0$ *there is* $\gamma > 0$ *such that* $|\theta' - \theta| < \gamma$ *implies*

$$\sup_{x \in B(0,R)} |W_\alpha^{\theta'}(x,T) - W_\alpha^\theta(x,T)| < \delta.$$

(iii) *The mapping* $\theta \mapsto l_\alpha(\theta)$ *is continuous.*

Remark 3.4. The notion of identifiability is reminiscent of a notion of observability for nonlinear systems, which also has a variational characterization, see [15] and [16].

For instance, if b_θ and h_θ do not depend on θ, and the mapping $\theta \mapsto \bar{x}_0^\theta$ is injective, then the identifiability of the model $\mathscr{M}^0$ is equivalent to the observability of the system

$$(\Sigma) \quad \begin{cases} \dot{x}_t = b(x_t), \\ \dot{y}_t = h(x_t), \qquad y_0 = 0. \end{cases}$$

4. Consistency Result for MLE

The main result of this paper is the following:

Theorem 4.1. *For all $\alpha \in \Theta$:*

(i) *Any MLE sequence $\{\hat{\theta}^\varepsilon, \varepsilon > 0\}$ converges in $P_{\alpha,\varepsilon}$-probability to the deterministic set M_α: for all $\delta > 0$,*

$$P_{\alpha,\varepsilon}\left(d\left(\hat{\theta}^\varepsilon, M_\alpha\right) > \delta\right) \xrightarrow{\varepsilon \downarrow 0} 0.$$

(ii) *If the deterministic model $\mathscr{M}^0$ is identifiable, then any MLE sequence $\{\hat{\theta}^\varepsilon, \varepsilon > 0\}$ converges in $P_{\alpha,\varepsilon}$-probability to the* true *parameter: for all $\delta > 0$,*

$$P_{\alpha,\varepsilon}(|\hat{\theta}^\varepsilon - \alpha| > \delta) \xrightarrow{\varepsilon \downarrow 0} 0.$$

The proof of this theorem depends on a technical extension of large deviations limit results for nonlinear filtering contained in [15] and [17]. We need to show that certain limits are uniform in the parameter $\theta \in \Theta$. The key technical lemma is the following:

Lemma 4.2. *The sequence $\{l^\varepsilon(\theta), \varepsilon > 0\}$ converges in $P_{\alpha,\varepsilon}$-probability, uniformly in $\theta \in \Theta$, to $l_\alpha(\theta)$: for all $\delta > 0$,*

$$P_{\alpha,\varepsilon}\left(\sup_{\theta \in \Theta} |l^\varepsilon(\theta) - l_\alpha(\theta)| > \delta\right) \overset{\varepsilon \downarrow 0}{\to} 0.$$

We next prove Theorem 4.1 using Lemma 4.2. The remainder of this section is concerned with proving Lemma 4.2.

Proof of Theorem 4.1. By Lemma A.1, for all $\delta > 0$ there is $\gamma > 0$ such that

$$\left\{\sup_{\theta \in \Theta} |l^\varepsilon(\theta) - l_\alpha(\theta)| < \gamma\right\} \subset \left\{d\left(\hat{\theta}^\varepsilon, M_\alpha\right) < \delta\right\}.$$

Therefore, by Lemma 4.2,

$$P_{\alpha,\varepsilon}\Big(d\big(\hat{\theta}^{\varepsilon}, M_\alpha\big) > \delta\Big) \le P_{\alpha,\varepsilon}\Big(\sup_{\theta\in\Theta} |l^{\varepsilon}(\theta) - l_\alpha(\theta)| > \gamma\Big) \xrightarrow{\varepsilon\downarrow 0} 0,$$

which proves (i).

The proof of (ii) follows at once from (i). □

Remark 4.3. Note that the log-likelihood function $-l^{\varepsilon}(\theta)$ (resp. the DPE functional $l_\alpha(\theta)$) can be defined as an integral (resp. an infimum) over either the state space $\mathbf{R}^m$ or the path space $C([0,T];\mathbf{R}^m)$. Actually

$$-l^{\varepsilon}(\theta) = \varepsilon \log \int_{\mathbf{R}^m} p^{\theta,\varepsilon}(x,T)\,dx = \varepsilon \log \mathbf{E}^{\dagger}_{\theta,\varepsilon}(Z^{\theta,\varepsilon}|\mathscr{Y}_T),$$

and

$$l_\alpha(\theta) = \inf_{x\in\mathbf{R}^m} W^{\theta}_{\alpha}(x,T) = \inf\{J^{\theta}_{\alpha}(\xi,T)\colon \xi \in C([0,T];\mathbf{R}^m)\}.$$

To prove the large deviations result we are interested in, we could adopt the path space formulation and employ Varadhan's theorem, which relies on probabilistic arguments only. Instead we use the Laplace's asymptotic method on the state space, which entails use of the PDE vanishing viscosity method. Our choice is motivated by the fact that PDEs provide the most efficient way for computing both the log-likelihood function and the DPE functional.

As in [15] and [17] we employ the vanishing viscosity method of Evans and Ishii [10]. We proceed by a logarithmic change of variables used by Fleming and Mitter [11]. Define

$$W^{\theta,\varepsilon}(x,t) \triangleq -\varepsilon \log p^{\theta,\varepsilon}(x,t). \tag{4.1}$$

The $\mathscr{Y}_t$-measurable random variable $[W^{\theta,\varepsilon}(x,t) + h^*_\theta(x)Y_t]$ can be extended to a continuous function defined on the whole canonical space $\Omega_0 \equiv \{\eta \in C([0,T];\mathbf{R}^d)\colon \eta_0 = 0\}$, which we denote by $u^{\theta,\varepsilon}[\eta](x,t)$, see [11] and [17]. For any fixed $\eta \in \Omega_0$,

$$u^{\theta,\varepsilon}[\eta] \in C^{2,1}(\mathbf{R}^m \times [0,T];\mathbf{R})$$

is the unique solution of the Hamilton–Jacobi–Bellman equation

$$\frac{\partial}{\partial t}u^{\theta,\varepsilon}[\eta](x,t) - \tfrac{1}{2}\varepsilon\,\Delta u^{\theta,\varepsilon}[\eta](x,t) + H^{\theta,\varepsilon}[\eta](x,t,Du^{\theta,\varepsilon}[\eta](x,t)) = 0$$

$$u^{\theta,\varepsilon}[\eta](x,0) = S^{\theta}_0(x) - \varepsilon \log C_{\theta,\varepsilon}, \tag{4.2}$$

where the Hamiltonian $H^{\theta,\varepsilon}[\eta](x,t,\lambda)$ is defined by

$$
\begin{aligned}
&H^{\theta,\varepsilon}[\eta](x,t,\lambda) \triangleq g_\theta^*(x,\eta_t)\lambda + \tfrac{1}{2}|\lambda|^2 - V^{\theta,\varepsilon}(x,\eta_t),\\
&V^{\theta,\varepsilon}(x,\eta) \triangleq V^\theta(x,\eta) + \tfrac{1}{2}\varepsilon\eta^*\Delta h_\theta(x) + \varepsilon \operatorname{div} g_\theta(x,\eta),\\
&V^\theta(x,\eta) \triangleq \tfrac{1}{2}|h_\theta(x)|^2 + b_\theta^*(x)\eta^* Dh_\theta(x) - \tfrac{1}{2}Dh_\theta^*(x)\eta\eta^* Dh_\theta(x),\\
&g_\theta(x,\eta) \triangleq b_\theta(x) - \eta^* Dh_\theta(x).
\end{aligned}
\tag{4.3}
$$

Next, for $\eta \in \Omega_0$, let

$$u^\theta[\eta] \in C(\mathbf{R}^m \times [0,T]; \mathbf{R})$$

denote the unique viscosity solution of the Hamilton–Jacobi equation

$$\frac{\partial}{\partial t}u^\theta[\eta](x,t) + H^\theta[\eta](x,t,Du^\theta[\eta](x,t)) = 0, \qquad u^\theta[\eta](x,0) = S_0^\theta(x), \tag{4.4}$$

where the Hamiltonian $H^\theta[\eta](x,t,\lambda)$ is defined by

$$H^\theta[\eta](x,t,\lambda) \triangleq g_\theta^*(x,\eta_t)\lambda + \tfrac{1}{2}|\lambda|^2 - V^\theta(x,\eta_t). \tag{4.5}$$

The following estimate is obtained as in [15] and [17] using methods introduced in [10].

Lemma 4.4. *Let* $R > 0$, *and* $K \subset \Omega_0$ *be* bounded. *Then, if* $\varepsilon > 0$ *is sufficiently small, we have*

$$|u^{\theta,\varepsilon}[\eta](x,t)| \le C,$$

for some constant $C > 0$ *and for all* $\theta \in \Theta$, $\eta \in K$, *and* $(x,t) \in B(0,R) \times [0,T]$.

Proof. Consider the function

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

Then, for sufficiently small $\varepsilon > 0$, we have

$$\frac{\partial}{\partial t}v(x,t) - \tfrac{1}{2}\varepsilon\,\Delta v(x,t) + H^{\theta,\varepsilon}[\eta](x,t,Dv(x,t)) \le 0 \qquad \text{in} \quad B(0,R) \times [0,T],$$

for all $\eta \in K$, $\theta \in \Theta$, provided $\mu > 0$ is chosen sufficiently large. Choose $M > 0$ so large that

$$S_0^\theta(x) - \varepsilon \log C_{\theta,\varepsilon} \le M \qquad \text{in} \quad B(0,R),$$

for all $\theta \in \Theta$. The comparison theorem implies

$$u^{\theta,\varepsilon}[\eta](x,t) \leq v(x,t) \quad \text{in} \quad B(0,R) \times [0,T],$$

and hence, for some constant $C > 0$,

$$u^{\theta,\varepsilon}[\eta](x,t) \leq C \quad \text{in} \quad B(0,\tfrac{1}{2}R) \times [0,T].$$

Similarily, we can find a lower bound for $u^{\theta,\varepsilon}[\eta]$ in $B(0,\frac{1}{2}R) \times [0,T]$, which proves the result. □

The comparison theorem used in this proof, and in the proof of the next lemma, is classical, see, for instance, Lemma 4.2 of [17].

Remark 4.5. It follows from the proof of Lemma 3.3 that a similar estimate holds for the solution of (4.4). Indeed, let $R > 0$ and $z \in \Omega_0$. Then

$$|u^{\theta}[z](x,t)| \leq C,$$

for some constant $C > 0$ and for all $\theta \in \Theta$, and $(x,t) \in B(0,R) \times [0,T]$.

Lemma 4.6. *Fix $z \in \Omega_0$. For all $\beta > 0$, there are $\gamma > 0$, $\varepsilon_0 > 0$ such that if $\|\eta - z\| \leq \gamma$ and $0 < \varepsilon \leq \varepsilon_0$, then*

$$\sup_{\theta \in \Theta} \sup_{(x,t) \in Q_R} \left| u^{\theta,\varepsilon}[\eta](x,t) - u^{\theta}[z](x,t) \right| \leq \beta,$$

where $Q_R = B(0,R) \times [0,T]$.

Proof. We claim first that

$$\lim_{\varepsilon \to 0,\, \theta' \to \theta,\, \eta \to z,\, x' \to x,\, t' \to t} u^{\theta',\varepsilon}[\eta](x',t') = u^{\theta}[z](x,t) \quad \text{in} \quad \mathbf{R}^m \times [0,T]. \tag{4.6}$$

Following Barles and Perthame [2], if we let $\bar{u}$ denote the upper limit

$$\bar{u}(x,t) = \limsup_{\varepsilon \to 0,\, \theta' \to \theta,\, \eta \to z,\, x' \to x,\, t' \to t} u^{\theta',\varepsilon}[\eta](x',t')$$

and $\underline{u}$ denote the corresponding lower limit (these functions are well-defined in view of the uniform bound in $C(Q_R)$ ensured by Lemma 4.4), then as in Theorem A.2 of [2] we deduce

$$\left.\begin{aligned} &\frac{\partial}{\partial t}\bar{u}(x,t) + H^{\theta}[z](x,t,D\bar{u}(x,t)) \leq 0, \\ &\frac{\partial}{\partial t}\underline{u}(x,t) + H^{\theta}[z](x,t,D\underline{u}(x,t)) \geq 0, \end{aligned}\right\} \quad \text{in} \quad \mathbf{R}^m \times [0,T],$$

in the viscosity sense. To see this, we proceed as follows. Let $\varphi \in C^{2,1}(\mathbf{R}^m \times [0,T])$ and assume, without loss of generality [5, Theorem 1.4], [2, Theorem A2 and Lemma

A3], that $\bar{u} - \varphi$ attains a strict local maximum at $(x_0, t_0) \in \mathbf{R}^n \times (0, T)$. Then there is a sequence indexed by i such that $u^{\theta_i, \varepsilon_i}[\eta_i] - \varphi$ attains a local maximum at a point $(x_i', t_i'), (x_i', t_i') \to (x_0, t_0), u^{\theta_i, \varepsilon_i}[\eta_i](x_i', t_i') \to \bar{u}(x_0, t_0)$, $\varepsilon_i \to 0$, $\theta_i \to \theta$, $\eta_i \to z$ as $i \to \infty$. Let us write $u^i(x, t) = u^{\theta_i, \varepsilon_i}[\eta_i](x, t)$ and $H^i(x, t, p) = H^{\theta_i, \varepsilon_i}[\eta_i](x, t, p)$. Then

$$\begin{aligned}&\frac{\partial}{\partial t}\varphi(x_i', t_i') - \tfrac{1}{2}\varepsilon_i\,\Delta\varphi(x_i', t_i') + H^i(x_i', t_i', D\varphi(x_i', t_i'))\\&\qquad\le \frac{\partial}{\partial t}u^i(x_i', t_i') - \tfrac{1}{2}\varepsilon_i\,\Delta u^i(x_i', t_i') + H^i(x_i', t_i', Du^i(x_i', t_i')) = 0.\end{aligned}$$

Since $\varphi \in C^{2,1}$ and, by our continuity assumptions, sending $i \to \infty$ gives

$$\lim_{i\to\infty}\frac{\partial}{\partial t}\varphi(x_i', t_i') = \frac{\partial}{\partial t}\varphi(x_0, t_0), \qquad \lim_{i\to\infty}\Delta\varphi(x_i', t_i') = \Delta\varphi(x_0, t_0),$$

and

$$\lim_{i\to\infty} H^i(x_i', t_i', D\varphi(x_i', t_i')) = H^\theta[z](x_0, t_0, D\varphi(x_0, t_0));$$

hence

$$\frac{\partial}{\partial t}\varphi(x_0, t_0) + H^\theta[z](x_0, t_0, D\varphi(x_0, t_0)) \le 0,$$

which proves that $\bar{u}$ satisfies the first inequality in the viscosity sense. Similarly, it can be shown that $\underline{u}$ satisfies the second inequality in the viscosity sense.

By definition, $\underline{u} \le \bar{u}$, and by the comparison theorem [6], [2], $\bar{u} \le \underline{u}$, thus

$$\begin{aligned}\lim_{\varepsilon\to 0,\,\theta'\to\theta,\,\eta\to z,\,x'\to x,\,t'\to t} u^{\theta',\varepsilon}[\eta](x', t') &= \bar{u}(x, t) = \underline{u}(x, t)\\&= u^\theta[z](x, t) \qquad \text{in} \quad \mathbf{R}^m \times [0, T].\end{aligned}$$

This proves (4.6).

Suppose the assertion of the lemma is false. Then there are $\beta > 0$, sequences $\varepsilon_j \downarrow 0$, $\gamma_j \downarrow 0$, $\eta_j \in \Omega_0$ with $\|\eta_j - z\| \le \gamma_j$, $\theta_j \to \theta$ in Θ, and $(x_j, t_j) \to (x, t)$ in Q_R, such that

$$\left|u^{\theta_j, \varepsilon_j}[\eta_j](x_j, t_j) - u^{\theta_j}[z](x_j, t_j)\right| > \beta. \tag{4.7}$$

However, (4.6) implies that

$$\lim_{j\to\infty} u^{\theta_j, \varepsilon_j}[\eta_j](x_j, t_j) = u^\theta[z](x, t). \tag{4.8}$$

In addition, it follows from Remark 4.5 above, that the set $\{u^{\theta_j}[z], j \in \mathbf{N}\}$ is uniformly bounded in $C(Q_R)$. Repeating exactly the same argument as above, we see that, for $(x,t) \in \mathbf{R}^m \times [0,T]$,

$$\lim_{j \to \infty} u^{\theta_j}[z](x_j, t_j) = u^{\theta}[z](x,t). \tag{4.9}$$

Clearly, (4.8) and (4.9) contradict (4.7), which proves the result. □

Note that

$$W^{\theta,\varepsilon}(x,t) = u^{\theta,\varepsilon}[Y](x,t) - h_\theta^*(x)Y_t$$

and

$$W_\alpha^\theta(x,t) = u^\theta[y^\alpha](x,t) - h_\theta^*(x)y_t^\alpha.$$

Lemma 4.7. *We have*

$$\lim_{\varepsilon \downarrow 0} W^{\theta,\varepsilon}(x,t) = W_\alpha^\theta(x,t)$$

in $P_{\alpha,\varepsilon}$-probability, uniformly in $\theta \in \Theta$, $t \in [0,T]$, and uniformly on compact subsets of $x \in \mathbf{R}^m$.

Proof. Let ρ denote a metric on $C(\mathbf{R}^m \times [0,T]; \mathbf{R})$ corresponding to uniform convergence on compact subsets. By (2.4), it is enough to show that, for each $\delta > 0$,

$$P_{\alpha,\varepsilon}\left(\sup_{\theta \in \Theta} \rho(u^{\theta,\varepsilon}[Y], u^\theta[y^\alpha]) > \delta\right) \xrightarrow{\varepsilon \downarrow 0} 0.$$

From Lemma 4.6, for all $\beta > 0$ there are $\gamma > 0$, $\varepsilon_0 > 0$ such that if $0 < \varepsilon \le \varepsilon_0$, then

$$\{\|Y - y^\alpha\| \le \gamma\} \subset \left\{\sup_{\theta \in \Theta} \rho(u^{\theta,\varepsilon}[Y], u^\theta[y^\alpha]) \le \beta\right\}.$$

Therefore, if $0 < \varepsilon \le \varepsilon_0$, then

$$P_{\alpha,\varepsilon}\left(\sup_{\theta \in \Theta} \rho(u^{\theta,\varepsilon}[Y], u^\theta[y^\alpha]) > \beta\right) \le P_{\alpha,\varepsilon}(\|Y - y^\alpha\| > \gamma) \overset{\varepsilon \downarrow 0}{\to} 0,$$

by (2.4). □

Proof of Lemma 4.2. Recall from (2.3) and (3.8) that

$$l^\varepsilon(\theta) = -\varepsilon \log \int_{\mathbf{R}^m} \exp\left\{-\frac{1}{\varepsilon} W^{\theta,\varepsilon}(x,T)\right\} dx \quad \text{a.s.,}$$

$$l_\alpha(\theta) = \inf_{x \in \mathbf{R}^m} W_\alpha^\theta(x,T).$$

From estimate (B.2) below, we see that

$$W^{\theta,\varepsilon}(x,T) \ge C_2|x| - C' \quad \text{a.s.}$$

for all $\varepsilon > 0$, $\theta \in \Theta$, where C' is random and satisfies the following estimate:

$$C' \leq C_0 \|Y - y^\alpha\|^2 + C_0.$$

From Lemma A.3, for all $\delta > 0$ there are $\varepsilon_0 > 0$, $\beta > 0$, and $c > 0$ such that if $0 < \varepsilon \leq \varepsilon_0$, then

$$\left\{ \sup_{\theta \in \Theta} \rho(W^{\theta, \varepsilon}, W_\alpha^\theta) < \beta, \|Y - y^\alpha\| < c \right\} \subset \left\{ \sup_{\theta \in \Theta} |l^\varepsilon(\theta) - l_\alpha(\theta)| < \delta \right\}.$$

Therefore, for $0 < \varepsilon \leq \varepsilon_0$,

$$P_{\alpha, \varepsilon}\left(\sup_{\theta \in \Theta} |l^\varepsilon(\theta) - l_\alpha(\theta)| > \delta \right)$$

$$\leq P_{\alpha, \varepsilon}\left(\sup_{\theta \in \Theta} \rho(W^{\theta, \varepsilon}, W_\alpha^\theta) > \beta \right) + P_{\alpha, \varepsilon}(\|Y - y^\alpha\| > c) \xrightarrow{\varepsilon \downarrow 0} 0,$$

by (2.4) and Lemma 4.7. □

5. Binary Sequential Detection

In this section we discuss some aspects of a binary detection problem studied by Baras and La Vigna [1], when the noise intensities are small.

Let $\Theta = \{0, 1\}$, and let X and Y be the signal and the observation processes described in Section 2. For $\varepsilon > 0$ fixed, we consider the two hypotheses H_0 and H_1. Under H_0 the law of (X, Y) is $P_{0, \varepsilon}$, whilst under H_1 the law of (X, Y) is $P_{1, \varepsilon}$. The problem is to determine which hypothesis is true, that is to detect the signal. In this section $\Omega = C([0, \infty), \mathbf{R}^{m+d})$.

A key technical assumption, essentially an identifiability condition, used in [1] is the following:

$$\int_0^\infty \left| \hat{h}_{0, \varepsilon}(t) - \hat{h}_{1, \varepsilon}(t) \right|^2 dt = \infty \quad P_{\alpha, \varepsilon}\text{-a.s.} \tag{5.1}$$

where

$$\hat{h}_{\theta, \varepsilon}(t) \triangleq E_{\theta, \varepsilon}(h_\theta(X_t) | \mathscr{Y}_t),$$

and

$$\mathscr{Y}_t \triangleq \sigma(Y_s, 0 \leq s \leq t).$$

The deterministic analogue of (5.1) is

$$\int_0^\infty |\dot{y}_t^0 - \dot{y}_t^1|^2 dt = \infty. \tag{5.2}$$

Clearly, (5.2) implies that the model $\mathscr{M}^0$ defined by (3.1) is identifiable. In fact, if

$$\sigma \triangleq \inf\left\{T \geq 0: \int_0^T |\dot{y}_t^0 - \dot{y}_t^1|^2\, dt > 0\right\},$$

then $\mathscr{M}^0$ is identifiable on each interval $[0, T]$ with $T > \sigma$.

The following result is a consequence of Theorem 4.1.

Theorem 5.1. *Assume* (5.2) *holds and* $T > \sigma$. *Let* $\hat{\theta}^\varepsilon$ *denote the MLE for the interval* $[0, T]$. *Then, for* $\alpha = 0, 1$,

$$P_{\alpha,\varepsilon}(\hat{\theta}^\varepsilon = \alpha) \xrightarrow{\varepsilon \downarrow 0} 1.$$

In [1] Baras and La Vigna use a threshold decision policy to decide which of the hypotheses is valid. Define the likelihood ratio

$$\Lambda_T^\varepsilon \triangleq \exp\left\{\frac{1}{\varepsilon}\int_0^T \left[\hat{h}_{1,\varepsilon}(t) - \hat{h}_{0,\varepsilon}(t)\right]^* dY_t - \frac{1}{2\varepsilon}\int_0^T \left[|\hat{h}_{1,\varepsilon}(t)|^2 - |\hat{h}_{0,\varepsilon}(t)|^2\right] dt\right\}.$$

Note that, as $\varepsilon \downarrow 0$,

$$\varepsilon \log \Lambda_T^\varepsilon \asymp \begin{cases} -\frac{1}{2}\int_0^T |\hat{h}_{1,0}(t) - \dot{y}_t^0|^2\, dt & \text{under } \mathrm{H}_0, \\ +\frac{1}{2}\int_0^T |\hat{h}_{0,1}(t) - \dot{y}_t^1|^2\, dt & \text{under } \mathrm{H}_1, \end{cases}$$

provided the limits $\hat{h}_{\theta,\alpha}(t) = \lim_{\varepsilon \downarrow 0} \hat{h}_{\theta,\varepsilon}(t)$ exist in $P_{\alpha,\varepsilon}$-probability ($\theta, \alpha = 0, 1$).

A threshold policy $u^\varepsilon = (\tau^\varepsilon, \delta^\varepsilon)$ consists of a $\{\mathscr{Y}_t, t \geq 0\}$-stopping time τ^ε and a $\mathscr{Y}_{\tau^\varepsilon}$-measureable $\{0, 1\}$-valued random variable δ^ε defined by

$$\tau^\varepsilon \triangleq \inf\{T \geq 0: \Lambda_T^\varepsilon \notin (e^{a/\varepsilon}, e^{b/\varepsilon})\},$$

$$\delta^\varepsilon \triangleq \begin{cases} 1 & \text{if } \Lambda_{\tau^\varepsilon}^\varepsilon = e^{b/\varepsilon}, \\ 0 & \text{if } \Lambda_{\tau^\varepsilon}^\varepsilon = e^{a/\varepsilon}, \end{cases}$$

for some constants $a < 0 < b$. If $\delta^\varepsilon = 1$ we decide that hypothesis H_1 is valid (i.e., that $\theta = 1$), whilst if $\delta^\varepsilon = 0$ we decide H_0 is (i.e., $\theta = 0$). Of course, our decision may be in error. Define an error probability for the policy u^ε:

$$e(u^\varepsilon) \triangleq P_{0,\varepsilon}(\delta^\varepsilon = 1) + P_{1,\varepsilon}(\delta^\varepsilon = 0).$$

Theorem 5.2. *If* (5.1) *holds, then*

$$e(u^\varepsilon) \xrightarrow{\varepsilon \downarrow 0} 0.$$

Proof. Under assumption (5.1), Baras and La Vigna [1] prove that

$$\tau^\varepsilon < \infty \qquad P_\alpha\text{-a.s.}$$

and

$$P_{0,\varepsilon}(\delta^\varepsilon = 1) = \frac{1 - e^{a/\varepsilon}}{e^{b/\varepsilon} - e^{a/\varepsilon}}, \qquad P_{1,\varepsilon}(\delta^\varepsilon = 0) = \frac{e^{a/\varepsilon}(e^{b/\varepsilon} - 1)}{e^{b/\varepsilon} - e^{a/\varepsilon}}.$$

Since $a < 0 < b$, the conclusion follows. □

Thus, assuming (5.1), the probability of making an incorrect decision converges to zero as $\varepsilon \downarrow 0$, and so (5.1) can be viewed as an identifiability criterion for the statistical model $\mathscr{M}^\varepsilon = \{P_{0,\varepsilon}, P_{1,\varepsilon}\}$.

We can define a deterministic threshold policy $u = (\tau, \delta)$ as follows. Define

$$F_T = \tfrac{1}{2}\int_0^T |\dot{y}_t^0 - \dot{y}_t|^2\, dt - \tfrac{1}{2}\int_0^T |\dot{y}_t^1 - \dot{y}_t|^2\, dt.$$

Let $a < 0 < b$ and set

$$\tau = \inf\{T \geq 0 \colon F_T \notin (a, b)\},$$

$$\delta = \begin{cases} 1 & \text{if} \quad F_\tau = b, \\ 0 & \text{if} \quad F_\tau = a. \end{cases}$$

Theorem 5.3. *Assume that* (5.2) *holds. Then, for any threshold policy* $u = (\tau, \delta)$ *with* $a < 0 < b$, *we have* $\tau < \infty$ *and*

$$\delta = 1 \qquad \textit{if and only if } \mathrm{H}_1 \textit{ is valid},$$

$$\delta = 0 \qquad \textit{if and only if } \mathrm{H}_0 \textit{ is valid}.$$

Proof. Under H_1, $y_t = y_t^1$ and, for $T > 0$,

$$F_T = \tfrac{1}{2}\int_0^T |\dot{y}_t^0 - \dot{y}_t^1|^2\, dt \geq 0.$$

By (5.2), $T_1 > 0$ exists such that $F_{T_1} = b$. Consequently $\tau \leq T_1$ and $\delta = 1$.

Similarly, under H_0, $y_t = y_t^0$ and, for $T > 0$,

$$F_T = -\tfrac{1}{2}\int_0^T |\dot{y}_t^1 - \dot{y}_t^0|^2\, dt \leq 0.$$

We conclude again $\tau < \infty$ and $\delta = 0$. □

Thus a deterministic threshold policy always makes the correct decision under the (stronger) identifiability condition (5.2).

To compute u^ε (approximately), Baras and La Vigna [1] use a numerical solution of the DMZ equation. The above suggests an approximation when $\varepsilon \downarrow 0$ is small. Now

$$F_T = F_T(y^0, y^1; y).$$

Compute approximations $\tilde{y}^0, \tilde{y}^1$ to y^0, y^1 by numerically integrating the differential system (3.1). Set

$$\tilde{F}_T^\varepsilon = F_T(\tilde{y}^0, \tilde{y}^1; Y),$$

where Y is the noisy observation record. Now define, for $a < 0 < b$,

$$\tilde{\tau}^\varepsilon = \inf\{T \geq 0: \tilde{F}_T^\varepsilon \notin (a,b)\},$$

$$\tilde{\delta}^\varepsilon = \begin{cases} 1 & \text{if } \tilde{F}_{\tilde{\tau}^\varepsilon}^\varepsilon = b, \\ 0 & \text{if } \tilde{F}_{\tilde{\tau}^\varepsilon}^\varepsilon = a. \end{cases}$$

If the integration is sufficiently accurate, then we expect, for $\alpha = 0, 1$,

$$P_{\alpha,\varepsilon}(\tilde{\delta}^\varepsilon = \delta^\varepsilon) \xrightarrow{\varepsilon \downarrow 0} 0.$$

Note that $|a|, b$ can be increased to increase the level of confidence.

Appendix A. Uniform Laplace's Asymptotic Formula

The purpose of this appendix is to prove Lemma A.3 below, which is an extension of the Laplace asymptotic formula to the situation where the integrand may depend on a parameter.

Lemma A.1. *Let $\Lambda \subset \mathbf{R}^p$ be compact. For any $\varphi \in C(\Lambda, \mathbf{R})$ define the set*

$$M(\varphi) \triangleq \operatorname*{argmin}_{\lambda \in \Lambda} \varphi(\lambda).$$

Let $f \in C(\Lambda, \mathbf{R})$. Then for all $\alpha > 0$ there is $\beta > 0$ depending on f, such that if

$$g \in C(\Lambda, \mathbf{R}) \quad \text{and} \quad \sup_{\lambda \in \Lambda} |f(\lambda) - g(\lambda)| < \beta,$$

then,

$$\forall \lambda \in M(g), \qquad d(\lambda, M(f)) < \alpha.$$

Proof. If not, $\alpha > 0$ and a sequence $\{g_i, i > 0\}$ exist such that

$$\sup_{\lambda \in \Lambda} |f(\lambda) - g_i(\lambda)| \to 0 \qquad \text{as} \quad i \to \infty,$$

and

$$d(\hat{\lambda}_i, M(f)) \geq \alpha \qquad \text{for some} \quad \hat{\lambda}_i \in M(g_i).$$

Since Λ is compact, we can assume that $\hat{\lambda}_i \to \lambda^* \in \Lambda$ as $i \to \infty$. Consequently,

$$d(\lambda^*, M(f)) \geq \alpha. \tag{A.1}$$

Let $\hat{\lambda}(f) \in M(f)$. Then

$$\begin{aligned} f(\hat{\lambda}_i) &= f(\hat{\lambda}(f)) + \left[g_i(\hat{\lambda}(f)) - f(\hat{\lambda}(f))\right] + \left[g_i(\hat{\lambda}_i) - g_i(\hat{\lambda}(f))\right] \\ &\quad + \left[f(\hat{\lambda}_i) - g_i(\hat{\lambda}_i)\right] \\ &\leq f(\hat{\lambda}(f)) + \left[g_i(\hat{\lambda}(f)) - f(\hat{\lambda}(f))\right] + \left[f(\hat{\lambda}_i) - g_i(\hat{\lambda}_i)\right] \\ &\leq f(\hat{\lambda}(f)) + 2 \sup_{\lambda \in \Lambda} |f(\lambda) - g_i(\lambda)|. \end{aligned}$$

Sending $i \to \infty$, we obtain $f(\lambda^*) \leq f(\hat{\lambda}(f))$. That is $\lambda^* \in M(f)$ which contradicts (A.1). □

Lemma A.2. *Let $\Lambda \subset \mathbf{R}^p$ be compact, and let $F^\lambda \in C(\mathbf{R}^m, \mathbf{R})$ be such that:*

(a) *There are constants $C > 0$, $C' > 0$ such that, for all $z \in \mathbf{R}^m$, $\lambda \in \Lambda$,*

$$F^\lambda(z) \geq C|z| - C'.$$

(b) *For all $R > 0$, $\delta > 0$ there is $\gamma > 0$ such that $|\lambda' - \lambda| < \gamma$ implies*

$$\sup_{z \in B(0,R)} |F^\lambda(z) - F^{\lambda'}(z)| < \delta.$$

Define $m^\lambda \triangleq \inf_{z \in \mathbf{R}^m} F^\lambda(z)$. Then:

(i) *A constant $R > 0$ exists such that, for all $\lambda \in \Lambda$,*

$$\operatorname*{argmin}_{z \in \mathbf{R}^m} F^\lambda(z) \subset B(0, R).$$

(ii) *The mapping $\lambda \mapsto m^\lambda$ is continuous.*

Proof. For any $\lambda \in \Lambda$ let $z^\lambda \in \operatorname{argmin}_{z \in \mathbf{R}^m} F^\lambda(z)$. The existence of z^λ follows from the continuity of F^λ and the coercivity hypothesis (a). Moreover,

$$m^\lambda = F^\lambda(z^\lambda) \geq C|z^\lambda| - C',$$

and thus, for all $\lambda \in \Lambda$,

$$|z^\lambda| \leq \frac{m^\lambda + C'}{C}.$$

Fix $\lambda_0 \in \Lambda$. By (b) for each $\delta > 0$ there is $\gamma > 0$ such that $|\lambda - \lambda_0| < \gamma$ implies

$$m^\lambda \leq F^\lambda(z^{\lambda_0}) = m^{\lambda_0} + [F^\lambda(z^{\lambda_0}) - F^{\lambda_0}(z^{\lambda_0})] \leq m^{\lambda_0} + \delta.$$

Then $|\lambda - \lambda_0| < \gamma$ implies

$$z^\lambda \in B(0, R) \qquad \text{with} \quad R \triangleq \frac{m^{\lambda_0} + \delta + C'}{C},$$

which proves (i).

By (b) again, this implies

$$m^{\lambda_0} \le F^{\lambda_0}(z^\lambda) = m^\lambda + [F^{\lambda_0}(z^\lambda) - F^\lambda(z^\lambda)]$$
$$\le m^\lambda + \sup_{z \in B(0,R)} |F^{\lambda_0}(z) - F^\lambda(z)| \le m^\lambda + \delta,$$

and the proof of the lemma is now complete. □

The next lemma is a variant of Laplace's asymptotic method.

Lemma A.3. *Let $\Lambda \subset \mathbf{R}^p$ be compact, and let $F^\lambda, G^\lambda \in C(\mathbf{R}^m, \mathbf{R})$ be such that:*

(a) *There are constants $C > 0$, $C' > 0$ such that, for all $z \in \mathbf{R}^m$, $\lambda \in \Lambda$,*

$$F^\lambda(z) \ge C|z| - C', \qquad G^\lambda(z) \ge C|z| - C'.$$

(b) *For all $R > 0$, $\delta > 0$ there is $\gamma > 0$ such that $|\lambda' - \lambda| < \gamma$ implies*

$$\sup_{z \in B(0,R)} |F^\lambda(z) - F^{\lambda'}(z)| < \delta, \qquad \sup_{z \in B(0,R)} |G^\lambda(z) - G^{\lambda'}(z)| < \delta.$$

Let ρ denote a metric on $C(\mathbf{R}^m, \mathbf{R})$ corresponding to uniform convergence on compact sets. Then, for all $\delta > 0$ there are $\beta > 0$, $\varepsilon_0 > 0$ (depending on G) such that $0 < \varepsilon \le \varepsilon_0$ and

$$\sup_{\lambda \in \Lambda} \rho(F^\lambda, G^\lambda) < \beta$$

implies

$$\sup_{\lambda \in \Lambda} \left| \varepsilon \log \int_{\mathbf{R}^m} \exp\left\{ -\frac{1}{\varepsilon} F^\lambda(z) \right\} dz + \inf_{z \in \mathbf{R}^m} G^\lambda(z) \right| < \delta.$$

Proof. Define

$$m^\lambda(F) \triangleq \inf_{z \in \mathbf{R}^m} F^\lambda(z), \qquad m^\lambda(G) \triangleq \inf_{z \in \mathbf{R}^m} G^\lambda(z).$$

Lower Bound. It follows from Lemma A.2 that the mappings $\lambda \mapsto m^\lambda(F)$ and $\lambda \mapsto m^\lambda(G)$ are continuous. Further, there is a constant $R > 0$ such that

$$\operatorname*{argmin}_{z \in \mathbf{R}^m} G^\lambda(z) \subset B(0, \tfrac{1}{2}R),$$

for all $\lambda \in \Lambda$. Thus we can choose $0 < \beta < \frac{1}{12}\delta$ such that $\sup_{\lambda \in \Lambda} \rho(F^\lambda, G^\lambda) < \beta$ implies

$$\sup_{\lambda \in \Lambda} |m^\lambda(F) - m^\lambda(G)| < \tfrac{1}{3}\delta$$

and

$$\operatorname*{argmin}_{z \in \mathbf{R}^m} F^\lambda(z) \subset B(0, R)$$

for all $\lambda \in \Lambda$. Set

$$B_\delta^\lambda \triangleq \{z \in \mathbf{R}^m : F^\lambda(z) - m^\lambda(F) < \tfrac{1}{3}\delta\}.$$

Increasing R if necessary, $B_\delta^\lambda \subset B(0, R)$ for all $\lambda \in \Lambda$ by the uniform coercivity hypothesis (a).

Now $(z, \lambda) \mapsto G^\lambda(z)$ is uniformly continuous on $B(0, R) \times \Lambda$, so $r > 0$ exists such that

$$|z - z'| + |\lambda - \lambda'| < r \quad \text{implies} \quad |G^\lambda(z) - G^{\lambda'}(z')| < \tfrac{1}{6}\delta,$$

and also, since $0 < \beta < \frac{1}{12}\delta$,

$$|F^\lambda(z) - F^{\lambda'}(z')| \leq 2\tfrac{1}{12}\delta + \tfrac{1}{6}\delta = \tfrac{1}{3}\delta,$$

for any $z, z' \in B(0, R)$ and any $\lambda, \lambda' \in \Lambda$. Let $z^\lambda \in \operatorname{argmin}_{z \in \mathbf{R}^m} F^\lambda(z)$. Then $z^\lambda \in B(0, R)$ and

$$|z - z^\lambda| < r \quad \text{implies} \quad |F^\lambda(z) - m^\lambda(F)| < \tfrac{1}{3}\delta,$$

for all $\lambda \in \Lambda$. That is $B(z^\lambda, r) \subset B_\delta^\lambda$ for all $\lambda \in \Lambda$. Therefore

$$\infty > v_R \geq \mu(B_\delta^\lambda) \geq v_r > 0,$$

where μ denotes the Lebesgue measure in $\mathbf{R}^m$, and v_r (resp. v_R) denotes the Lebesgue measure of a ball of radius r (resp. R) in $\mathbf{R}^m$.

Now

$$\begin{aligned} a^\lambda(\varepsilon) &\triangleq \int_{\mathbf{R}^m} \exp\left\{-\frac{1}{\varepsilon}F^\lambda(z)\right\} dz \\ &\geq \int_{B_\delta^\lambda} \exp\left\{-\frac{1}{\varepsilon}F^\lambda(z)\right\} dz \geq \mu(B_\delta^\lambda) \exp\left\{-\frac{1}{\varepsilon}(m^\lambda(F) + \tfrac{1}{3}\delta)\right\} \end{aligned}$$

and

$$\begin{aligned} \varepsilon \log a^\lambda(\varepsilon) &\geq \varepsilon \log v_r - m^\lambda(F) - \tfrac{1}{3}\delta \\ &\geq \varepsilon \log v_r - m^\lambda(G) - \tfrac{2}{3}\delta \geq -m^\lambda(G) - \delta, \end{aligned}$$

provided $0 < \varepsilon \leq \varepsilon_1$ for some ε_1 independent of $\lambda \in \Lambda$.

Upper Bound. Let $0 < \nu < 1$. The uniform coercivity hypothesis (a) implies

$$\begin{aligned} a^\lambda(\varepsilon) &\leq \int_{\mathbf{R}^m} \exp\left\{-\frac{1-\nu}{\varepsilon}F^\lambda(z)\right\} \exp\left\{-\frac{\nu}{\varepsilon}F^\lambda(z)\right\} dz \\ &\leq \exp\left\{-\frac{1-\nu}{\varepsilon}m^\lambda(F)\right\} \int_{\mathbf{R}^m} \exp\left\{-\frac{\nu}{\varepsilon}F^\lambda(z)\right\} dz \\ &\leq \exp\left\{-\frac{1-\nu}{\varepsilon}m^\lambda(F)\right\} \exp\left\{\frac{\nu C'}{\varepsilon}\right\} \int_{\mathbf{R}^m} \exp\left\{-\frac{\nu C}{\varepsilon}|z|\right\} dz \\ &\leq \exp\left\{-\frac{1-\nu}{\varepsilon}m^\lambda(F)\right\} \exp\left\{\frac{\nu C'}{\varepsilon}\right\} \left(\frac{\varepsilon}{\nu C}\right)^m, \end{aligned}$$

for all $\varepsilon > 0$. Therefore

$$\begin{aligned}\varepsilon \log a^{\lambda}(\varepsilon) &\leq -(1-\nu)m^{\lambda}(F) + \nu C' + m\varepsilon(\log\varepsilon - \log \nu C)\\ &\leq -m^{\lambda}(G) + \tfrac{1}{3}(1-\nu)\delta + \nu m^{\lambda}(G) + \nu C' + m\varepsilon(\log\varepsilon - \log\nu C).\end{aligned}$$

Choose ν so small that $\nu m^{\lambda}(G) + \nu C' < \frac{1}{3}\delta$. Next, choose $0 < \varepsilon_0 < \varepsilon_1$ such that $m\varepsilon(\log\varepsilon - \log\nu C) < \frac{1}{3}\delta$ for $0 < \varepsilon < \varepsilon_0$. Then we have

$$\varepsilon \log a^{\lambda}(\varepsilon) \leq -m^{\lambda}(G) + \delta$$

provided $0 < \varepsilon \leq \varepsilon_0$. □

Appendix B. Proof of Lemma 2.1

From Sections 2 and 4 we have

$$l^{\varepsilon}(\theta) = -\varepsilon \log \int_{\mathbf{R}^m} q^{\theta,\varepsilon}(x,T)\exp\left\{\frac{1}{\varepsilon}h_{\theta}^{*}(x)Y_T\right\}dx,$$

where, for a.e. $\omega \in \Omega_0$,

$$q^{\theta,\varepsilon} \in C_b^{1,2}(\mathbf{R}^m \times [0,T])$$

solves the *robust* DMZ equation, see [4] and [7],

$$\begin{aligned}&\frac{\partial}{\partial t}q^{\theta,\varepsilon}(x,t) - \tfrac{1}{2}\varepsilon\Delta q^{\theta,\varepsilon}(x,t) + \bar{g}_{\theta}^{*}(x,t)Dq^{\theta,\varepsilon}(x,t)\\ &\qquad + \frac{1}{\varepsilon}\bar{V}^{\theta,\varepsilon}(x,t)q^{\theta,\varepsilon}(x,t) = 0, \qquad \text{(B.1)}\\ &q^{\theta,\varepsilon}(x,0) = p_0^{\theta,\varepsilon}(x),\end{aligned}$$

with $\bar{g}_{\theta}(x,t) = g_{\theta}(x,Y_t)$ and $\bar{V}^{\theta,\varepsilon}(x,t) = V^{\theta,\varepsilon}(x,Y_t)$ as defined in (4.3).

Fix $\varepsilon > 0$ and $\omega \in \Omega_0$ such that the above holds. Now $|\bar{g}_{\theta}(x,t)| \leq C$ and $|\bar{V}^{\theta,\varepsilon}(x,t)| \leq \bar{C}$ in $\mathbf{R}^m \times [0,T]$. Then

$$\frac{\partial}{\partial t}q^{\theta,\varepsilon}(x,t) - \tfrac{1}{2}\varepsilon\Delta q^{\theta,\varepsilon}(x,t) + \bar{g}_{\theta}^{*}(x,t)Dq^{\theta,\varepsilon}(x,t) - \frac{1}{\varepsilon}\bar{C}q^{\theta,\varepsilon}(x,t) \leq 0,$$

and the fundamental solution $\Gamma^{\theta,\varepsilon}(x,t;x',0)$ of (B.1) satisfies the following estimate, see [13]:

$$\Gamma^{\theta,\varepsilon}(x,t;x',0) \leq C(\varepsilon t)^{-m/2}\exp\left\{-c\frac{|x-x'|^2}{\varepsilon t}\right\},$$

where $C > 0$, and $c > 0$ are independent of $\theta \in \Theta$. By the maximum principle, for all $(x,t) \in \mathbf{R}^m \times [0,T]$,

$$0 \leq q^{\theta,\varepsilon}(x,t) \leq \left\{\int_{\mathbf{R}^m} \Gamma^{\theta,\varepsilon}(x,t;x',0) p_0^{\theta,\varepsilon}(x')\, dx'\right\} \exp\left\{\frac{\bar{C}t}{\varepsilon}\right\},$$

and

$$\begin{aligned}
&\int_{\mathbf{R}^m} \Gamma^{\theta,\varepsilon}(x,t;x',0) p_0^{\theta,\varepsilon}(x')\, dx' \\
&\quad \leq C(\varepsilon t)^{-m/2} \int_{\mathbf{R}^m} \exp\left\{-c\frac{|x-x'|^2}{\varepsilon t}\right\} \exp\left\{-\frac{1}{\varepsilon}(C_2|x'| - C_2')\right\} dx' \\
&\quad \leq C \exp\left\{-\frac{1}{\varepsilon}(C_2|x| - C'')\right\}.
\end{aligned}$$

Therefore

$$W^{\theta,\varepsilon}(x,t) \leq C_2|x| - C', \tag{B.2}$$

where C' is random and satisfies the following estimate:

$$C' \leq C_0\|Y - y^\alpha\|^2 + C_0.$$

Therefore, by the Lebesgue dominated convergence theorem, it is enough to show that if $\theta_k \to \theta_0$ in Θ as $k \to \infty$, then $q^{\theta_k,\varepsilon}(x,T) \to q^{\theta_0,\varepsilon}(x,T)$ for each $x \in \mathbf{R}^m$. The difference $z_k \triangleq q^{\theta_k,\varepsilon} - q^{\theta_0,\varepsilon}$ satisfies

$$\begin{aligned}
&\frac{\partial}{\partial t} z_k(x,t) - \tfrac{1}{2}\varepsilon\Delta z_k(x,t) + \bar{g}_{\theta_k}^*(x,t) D z_k(x,t) + \frac{1}{\varepsilon}\bar{V}^{\theta_k,\varepsilon}(x,t) z_k(x,t) \\
&\quad = -\left[\bar{g}_{\theta_k}(x,t) - \bar{g}_{\theta_0}(x,t)\right]^* D q^{\theta_0,\varepsilon}(x,t) \\
&\qquad - \frac{1}{\varepsilon}[\bar{V}^{\theta_k,\varepsilon}(x,t) - \bar{V}^{\theta_0,\varepsilon}(x,t)] q^{\theta_0,\varepsilon}(x,t),
\end{aligned}$$

and hence

$$\frac{\partial}{\partial t} z_k(x,t) - \tfrac{1}{2}\varepsilon\Delta z_k(x,t) + \bar{g}_{\theta_k}^*(x,t) D z_k(x,t) - \frac{1}{\varepsilon}\bar{C} z_k(x,t) \leq \frac{1}{\varepsilon} C\rho_k,$$

where $\rho_k \to 0$ as $k \to \infty$. Then by the maximum principle

$$z_k(x,t) \leq \left\{\int_{\mathbf{R}^m} \Gamma^{\theta_k,\varepsilon}(x,t;x',0) z_k(x',0)\, dx' + \frac{1}{\varepsilon} Ct\rho_k\right\} \exp\left\{\frac{\bar{C}t}{\varepsilon}\right\}.$$

Now

$$|z_k(x,0)| \leq \left\{|C_{\theta_k,\varepsilon} - C_{\theta_0,\varepsilon}| + \frac{1}{\varepsilon}|S_0^{\theta_k}(x) - S_0^{\theta_0}(x)|\right\} \exp\left\{-\frac{1}{\varepsilon}(C_2|x| - C_2')\right\}.$$

By the Lebesgue dominated convergence theorem again, we obtain

$$\limsup_{k\to\infty}\{q^{\theta_k,\varepsilon}(x,T) - q^{\theta_0,\varepsilon}(x,T)\} \le \limsup_{k\to\infty} z_k(x,T) \le 0.$$

Similarly, we obtain the reverse inequality and conclude □

Appendix C. Proof of Lemma 3.3

In what follows every constant independent of $\theta, \alpha \in \Theta$ and $(x,t) \in \mathbf{R}^m \times [0,T]$ is denoted by C or C'. For any absolutely continuous function $\xi \in C([0,T];\mathbf{R}^m)$ and any $\Delta > 0$, we have

$$|\xi_t|^2 \le |\xi_s|^2 + \frac{1}{\Delta}\int_s^t |\xi_\tau|^2\,d\tau + \Delta\int_s^t |\dot{\xi}_\tau|^2\,d\tau,$$

and, by Gronwall's lemma,

$$|\xi_t|^2 \le \left\{|\xi_s|^2 + \Delta\int_s^t |\dot{\xi}_\tau|^2\,d\tau\right\} e^{(t-s)/\Delta}. \tag{C.1}$$

Since $\sup_{\theta\in\Theta}\sup_{x\in\mathbf{R}^m}|h_\theta(x)| \le C$ it follows that, for all $\alpha \in \Theta$,

$$\tfrac{1}{2}\int_0^T |\dot{y}_s^\alpha|^2\,ds \le C,$$

and hence, for all $\alpha, \theta \in \Theta$,

$$W_\alpha^\theta(x,t) \ge -C, \qquad (x,t) \in \mathbf{R}^m \times [0,T].$$

Let $L_\alpha^\theta(\dot{\xi}, \xi, t)$ denote the Lagrangean in (3.4). It is easy to prove the following estimates:

$$\begin{aligned}\tfrac{1}{4}\int_s^t |\dot{\xi}_\tau|^2\,d\tau &\le \tfrac{1}{2}\int_s^t |\dot{\xi}_\tau - b_\theta(\xi_\tau)|^2\,d\tau + \tfrac{1}{2}\int_s^t |b_\theta(\xi_\tau)|^2\,d\tau \\ &\le \int_s^t L_\alpha^\theta\big(\dot{\xi}_\tau, \xi_\tau, \tau\big)\,d\tau + C,\end{aligned}$$

$$\tfrac{1}{2}\int_s^t |\dot{\xi}_\tau - b_\theta(\xi_\tau)|^2\,d\tau \le \int_s^t |\dot{\xi}_\tau|^2\,d\tau + \int_s^t |b_\theta(\xi_\tau)|^2\,d\tau \le \int_s^t |\dot{\xi}_\tau|^2\,d\tau + C.$$

In particular

$$\tfrac{1}{4}\int_0^t |\dot{\xi}_s|^2\,ds - C \le J_\alpha^\theta(\xi,t) \le S_0^\theta(\xi_0) + \int_0^t |\dot{\xi}_s|^2\,ds + C.$$

Proof of (i). Setting $\xi \equiv x$ on $[0,T]$, gives, for $0 \le t \le T$,

$$W_\alpha^\theta(x,t) \le J_\alpha^\theta(\xi,t) \le S_0^\theta(x) + C \le C_1|x|^2 + C.$$

Choose $\Delta > 0$ such that $N = T/\Delta$ is an integer and $4eC_1\Delta \leq \frac{1}{2}$. For $n = 1,\ldots,N$ the Dynamic Programming principle implies

$$W_\alpha^\theta(z,t_n) = \inf\left\{W_\alpha^\theta(\xi_{t_{n-1}},t_{n-1}) + \int_{t_{n-1}}^{t_n} L_\alpha^\theta\left(\dot{\xi}_s,\xi_s,s\right)ds\colon \xi_{t_n} = z\right\},$$

with $t_n = n\Delta$. Given $\delta > 0$, recursively select $\xi^n \in C([0,T];\mathbf{R}^m)$ for $n = N,\ldots,1$ as follows: $\xi_{t_N}^N = x$, $\xi_{t_n}^{n-1} = \xi_{t_n}^n$, and

$$W_\alpha^\theta(\xi_{t_{n-1}}^n,t_{n-1}) + \int_{t_{n-1}}^{t_n} L_\alpha^\theta\left(\dot{\xi}_s^n,\xi_s^n,s\right)ds \leq W_\alpha^\theta(\xi_{t_n}^n,t_n) + \frac{\delta}{N} \tag{C.2}$$

$$\leq C_1|\xi_{t_n}^n|^2 + C + \frac{\delta}{N}. \tag{C.3}$$

Then

$$\tfrac{1}{4}\int_{t_{n-1}}^{t_n} |\dot{\xi}_s^n|^2\,ds \leq C_1|\xi_{t_n}^n|^2 + C + \frac{\delta}{N},$$

and from (C.1)

$$\begin{aligned}|\xi_{t_n}^n|^2 &\leq \left\{|\xi_{t_{n-1}}^n|^2 + \Delta\int_{t_{n-1}}^{t_n}|\dot{\xi}_\tau^n|^2\,d\tau\right\}e\\ &\leq e|\xi_{t_{n-1}}^n|^2 + \tfrac{1}{2}|\xi_{t_n}^n|^2 + \frac{\frac{1}{2}(C+\delta/N)}{C_1},\end{aligned}$$

which implies

$$|\xi_{t_n}^n|^2 \leq 2e|\xi_{t_{n-1}}^n|^2 + \frac{C+\delta/N}{C_1}. \tag{C.4}$$

Define $\xi^\theta \in C([0,T];\mathbf{R}^m)$ by $\xi_t^\theta = \xi_t^n$ for $t \in [t_{n-1},t_n]$, $n = 1,\ldots,N$. Then $\xi_T^\theta = x$ and by iterating (C.4) we obtain

$$|x|^2 \leq C^N|\xi_0^\theta|^2 + C^N.$$

Now also, by iterating (C.3),

$$J_\alpha^\theta(\xi^\theta,T) \leq W_\alpha^\theta(x,T) + \delta, \tag{C.5}$$

and consequently

$$W_\alpha^\theta(x,T) \geq J_\alpha^\theta(\xi^\theta,T) - \delta \geq S_0^\theta(\xi_0^\theta) - C \geq C|x| - C',$$

which proves (i).

Proof of (ii). Let $R > 0$, $\delta > 0$, and $x \in B(0, R)$. Choose ξ^θ as in (C.6). Then, from the above estimates,

$$\int_0^T |\dot{\xi}_s^\theta|^2 \, ds \le C_R.$$

Using (C.1) we deduce that if $x \in B(0, R)$, then $R' > 0$ exists such that $\xi_0^\theta \in B(0, R')$ for all $\theta \in \Theta$. Therefore

$$\begin{aligned} W_\alpha^{\theta'}(x,T) - W_\alpha^\theta(x,T) &\le J_\alpha^{\theta'}(\xi^\theta, T) - J_\alpha^\theta(\xi^\theta, T) + \tfrac{1}{4}\delta \\ &= S_0^{\theta'}(\xi_0^\theta) - S_0^\theta(\xi_0^\theta) + \tfrac{1}{2}\int_0^T |\dot{\xi}_s^\theta - b_{\theta'}(\xi_s^\theta)|^2 \, ds \\ &\quad -\tfrac{1}{2}\int_0^T |\dot{\xi}_s^\theta - b_\theta(\xi_s^\theta)|^2 \, ds \\ &\quad +\tfrac{1}{2}\int_0^T |\dot{y}_s^\alpha - h_{\theta'}(\xi_s^\theta)|^2 \, ds \\ &\quad -\tfrac{1}{2}\int_0^T |\dot{y}_s^\alpha - h_\theta(\xi_s^\theta)|^2 \, ds + \tfrac{1}{4}\delta. \end{aligned}$$

Now, if $|\theta' - \theta|$ is small enough,

$$|S_0^{\theta'}(\xi_0^\theta) - S_0^\theta(\xi_0^\theta)| < \tfrac{1}{2}\delta,$$

$$\tfrac{1}{2}\left| \int_0^T |\dot{y}_s^\alpha - h_{\theta'}(\xi_s^\theta)|^2 \, ds - \int_0^T |\dot{y}_s^\alpha - h_\theta(\xi_s^\theta)|^2 \, ds \right| < \tfrac{1}{4}\delta.$$

Also

$$\begin{aligned} &\tfrac{1}{2}\left| \int_0^T |\dot{\xi}_s^\theta - b_{\theta'}(\xi_s^\theta)|^2 \, ds - \int_0^T |\dot{\xi}_s^\theta - b_\theta(\xi_s^\theta)|^2 \, ds \right| \\ &\quad \le \left\{ \int_0^T |\dot{\xi}_s^\theta|^2 \, ds \right\}^{1/2} \left\{ \int_0^T |b_{\theta'}(\xi_s^\theta) - b_\theta(\xi_s^\theta)|^2 \, ds \right\}^{1/2} \\ &\quad\quad + \tfrac{1}{2}\int_0^T |b_{\theta'}(\xi_s^\theta) - b_\theta(\xi_s^\theta)| \, |b_{\theta'}(\xi_s^\theta) + b_\theta(\xi_s^\theta)| \, ds < \tfrac{1}{4}\delta, \end{aligned}$$

if $|\theta' - \theta|$ is small enough. Hence, $\gamma > 0$ exists such that $|\theta' - \theta| < \gamma$ implies

$$W_\alpha^{\theta'}(x,T) - W_\alpha^\theta(x,T) < \delta.$$

Reversing the role of θ' and θ proves (ii).

Finally, (iii) follows from (i)–(ii) and Lemma A.2. □

References

1. Baras, J. S., and La Vigna, A. (1987), Real-time sequential hypothesis testing for diffusion signals, Proc. 26th IEEE CDC, Los Angeles, 1987, pp. 1153–1157, IEEE Press, New York.

2. Barles, G., and Perthame, B. (1987), Discontinuous solutions of deterministic optimal stopping time problems, Modél. Math. Anal. Numér. 21:557–579.
3. Campillo, F., and Le Gland, F. (1989), MLE for partially observed diffusions: direct maximization vs. the EM algorithm, Stochastic Process. Appl. 33:245–274.
4. Clark, J. M. C. (1978), The design of robust approximations to the stochastic differential equations of non-linear filtering, in: Communication Systems and Random Processes Theory (ed. J. K. Skwirzynski), pp. 721–734, Sijthoff and Nordhoff, Alphen aan den Rijn.
5. Crandall, M. G., Evans, L. C., and Lions, P. L. (1984), Some properties of viscosity solutions of Hamilton–Jacobi equations, Trans. Amer. Math. Soc. 282:487–502.
6. Crandall, M. G., and Lions, P. L. (1983), Viscosity solutions of Hamilton–Jacobi equations, Trans. Amer. Math. Soc. 277:1–42.
7. Davis, M. H. A. (1980), On a multiplicative functional transformation arising in non-linear filtering theory, Z. Wahrsch. Verw. Gebiete 54:125–139.
8. Dembo, A., and Zeitouni, O. (1986), Parameter estimation of partially observed continuous-time stochastic processes via the EM algorithm, Stochastic Process. Appl. 23:91–113.
9. Elliott, R. J. (1982), Stochastic Calculus and Applications, Springer-Verlag, New York.
10. Evans, L. C., and Ishii, H. (1985), A PDE approach to some asymptotic problems concerning random differential equations with small noise intensities, Ann. Inst. H. Poincaré Anal. Non Linéaire 2:1–20.
11. Fleming, W. H., and Mitter, S. K. (1982), Optimal control and nonlinear filtering for nondegenerate diffusion processes, Stochastics 8:63–77.
12. Freidlin, M. I., and Wentzell, A. D. (1984), Random Perturbations of Dynamical Systems, Springer-Verlag, New York.
13. Friedman, A. V. (1964), Partial Differential Equations of Parabolic Type, Prentice-Hall, Englewood Cliffs, NJ.
14. Hijab, O. (1984), Asymptotic Bayesian estimation of a first order equation with small diffusion, Ann. Probab. 12:890–902.
15. James, M. R. (1988), Asymptotic Nonlinear Filtering and Large Deviations with Application to Observer Design, Ph.D. Dissertation, University of Maryland (SRC Technical Report PhD-88-1, Systems Research Center).
16. James, M. R. (1991), Finite time observer design by probabilistic-variational methods, SIAM J. Control Optim. 29:954–967.
17. James, M. R., and Baras, J. S. (1988), Nonlinear filtering and large deviations: a PDE-control theoretic approach, Stochastics 23:391–412.
18. Pardoux, E. (1991), Filtrage non-linéaire et équations aux dérivées partielles stochastiques associées, in: Ecole d'Eté de Probabilités de Saint-Flour XIX, 1989 (ed. P. L. Hennequin), pp. 69–163 (LNM, 1464), Springer-Verlag, New York.

Accepted 6 October 1993