Consistent parameter estimation for partially observed diffusions with small noise

Mattew R. James, François Le Gland

UNITÉ DE RECHERCHE
INRIA-SOPHIA ANTIPOLIS

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P. 105
78153 Le Chesnay Cedex
France
Tél.:(1) 39 63 55 11

Rapports de Recherche

N° 1223

Programme 5
Automatique, Productique,
Traitement du Signal et des Données

CONSISTENT PARAMETER ESTIMATION FOR PARTIALLY OBSERVED DIFFUSIONS WITH SMALL NOISE

Matthew R. JAMES
François LE GLAND

Mai 1990

RR.1223

CONSISTENT PARAMETER ESTIMATION FOR PARTIALLY OBSERVED DIFFUSIONS WITH SMALL NOISE *

Estimation consistante de paramètres
pour les processus de diffusion partiellement observés

Matthew R. JAMES
Department of Mathematics
University of Kentucky
Lexington, KY 40506, USA

François LE GLAND
INRIA Sophia Antipolis
Route des Lucioles
F–06565 VALBONNE Cédex

*Research 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; by USACCE, under Contract DAJA45–90–C–0008, and by the CNRS-GR *"Automatique"*.

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.

1980 subject classifications: 62F12, 93E10, 93E11, 60F10

Résumé

On démontre la consistance du maximum de vraisemblance pour l'estimation de paramètres dans les processus de diffusion partiellement observés, dans le cas de petits bruits. Si le système déterministe sous-jacent est *identifiable*, alors tout estimateur du maximum de vraisemblance est proche de la vraie valeur du paramètre inconnu, pourvu que les bruits soient assez petits. La démonstration utilise des résultats de grandes déviations, qui sont obtenus par des techniques d'EDP (*vanishing viscosity*). On applique ce résultat à un problème de détection séquentielle, et on compare les notions déterministe et stochastique d'identifiabilité.

Mots–Clés: Estimation de paramètres, filtrage non–linéaire, grandes déviations.

Contents

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 [7] have investigated the EM algorithm, and Campillo and Le Gland [2] 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 Hijab [11], James and Baras [12], James [13]. The theory of large deviations for diffusions with small noise is presented in Freidlin and Wentzell [10]. 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 Baras and La Vigna [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 $(\Omega, \mathcal{F})$ we consider

- for each $\varepsilon > 0$, a family $\mathcal{M}^\varepsilon = \{P_{\theta,\varepsilon}\,,\, \theta \in \Theta\}$ 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 $P_{\theta,\varepsilon}$

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

$$dY_t = h_\theta(X_t)\,dt + dV_t^{\theta,\varepsilon}\ , \qquad Y_0 = 0\ ,$$

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

$$p_0^{\theta,\varepsilon}(x) \stackrel{\Delta}{=} C_{\theta,\varepsilon}\ \exp\{-\frac{1}{\varepsilon}S_0^\theta(x)\}\ . \tag{2.1}$$

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

(i) for all $\theta \in \Theta$, $b_\theta \in C_b^1(\mathbf{R}^m, \mathbf{R}^m)$, and $h_\theta \in C_b^2(\mathbf{R}^m, \mathbf{R}^d)$,

(ii) for all $\theta \in \Theta$, S_0^θ is convex, 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 $x \ne \bar{x}_0^\theta$. Assume also

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

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

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

(iii) for each $\delta > 0$, $R > 0$, there exists $\gamma > 0$ 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 $\mathcal{Y}_T$ denote the σ–algebra generated by the process Y on $C([0,T]\,;\,\mathbf{R}^d)$. The probability measures in $\mathcal{M}^\varepsilon$ are mutually absolutely continuous, and the *log-likelihood function* for estimating the parameter θ in the statistical model $\mathcal{M}^\varepsilon$ given $\mathcal{Y}_T$, can be expressed (note the minus sign) as

$$-\ell^\varepsilon(\theta) = \varepsilon \log \mathbf{E}^\dagger_{\theta,\varepsilon}(Z^{\theta,\varepsilon} \mid \mathcal{Y}_T) \; .$$

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

$$Z^{\theta,\varepsilon} \triangleq \frac{dP_{\theta,\varepsilon}}{dP^\dagger_{\theta,\varepsilon}} = \exp \frac{1}{\varepsilon} \left\{ \int_0^T h^*_\theta(X_s)\, dY_s - \tfrac{1}{2} \int_0^T |h_\theta(X_s)|^2\, ds \right\} \; ,$$

so that under $P^\dagger_{\theta,\varepsilon}$

$$dX_t = b_\theta(X_t)\, dt + dW^{\theta,\varepsilon}_t \; , \qquad X_0 \sim p^{\theta,\varepsilon}_0(x)\, dx \; ,$$

where $\{W^{\theta,\varepsilon}_t \, , \, t \geq 0\}$ 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 [2].

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

$$\widehat{\theta}^\varepsilon \in \operatorname*{argmin}_{\theta \in \Theta} \ell^\varepsilon(\theta) \; .$$

The likelihood function can be computed through the solution of the Zakai equation

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

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

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

The filtering problem is discussed in detail in Liptser and Shiryayev [15]. The following lemma is proved in the Appendix.

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

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

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

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

$$(\Sigma^\theta) \qquad \begin{cases} \dot{x}^\theta_t = b_\theta(x^\theta_t)\ , & x^\theta_0 = \bar{x}^\theta_0 \\ \dot{y}^\theta_t = h_\theta(x^\theta_t)\ , & y_0 = 0. \end{cases}$$

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

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

see Freidlin and Wentzell [10].

Remark 2.2 As long as $\varepsilon > 0$, the probability measures in $\mathcal{M}^\varepsilon$ are mutually absolutely continuous, which allows us to define the log–likelihood function $-\ell^\varepsilon(\theta)$. 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 will be proved below.

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

3 Deterministic Parameter Estimation

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

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

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 systems in $\mathcal{M}^0$. Introduce the following definition:

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

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

$$\begin{aligned} J_\alpha^\theta(\xi, t) \triangleq\ & S_0^\theta(\xi_0) + \tfrac{1}{2}\int_0^t |\dot{\xi}_s - b_\theta(\xi_s)|^2\, ds \\ & + \tfrac{1}{2}\int_0^t |\dot{y}_s^\alpha - h_\theta(\xi_s)|^2\, ds - \tfrac{1}{2}\int_0^t |\dot{y}_s^\alpha|^2\, ds , \end{aligned} \tag{3.2}$$

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

$$W_\alpha^\theta(x,t) \triangleq \inf\left\{J_\alpha^\theta(\xi,t) \,:\, \xi_t = x\right\} . \tag{3.3}$$

The value function $W_\alpha^\theta(x,t)$ is continuous in (x,t) and is the unique *viscosity solution* of the Hamilton–Jacobi equation [12]

$$\frac{\partial}{\partial t} W_\alpha^\theta(x,t) + H_\alpha^\theta(x,t,DW_\alpha^\theta(x,t)) = 0 , \quad W_\alpha^\theta(x,0) = S_0^\theta(x) , \tag{3.4}$$

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

$$\begin{aligned} H_\alpha^\theta(x,t,\lambda) &\triangleq \max_{u \in \mathbf{R}^m} \left\{\lambda^*(b_\theta(x) + u) - \tfrac{1}{2}|u|^2\right\} - \tfrac{1}{2}|\dot{y}_t^\alpha - h_\theta(x)|^2 + \tfrac{1}{2}|\dot{y}_t^\alpha|^2 \\ &= b_\theta^*(x)\lambda + \tfrac{1}{2}|\lambda|^2 + h_\theta^*(x)\dot{y}_t^\alpha - \tfrac{1}{2}|h_\theta(x)|^2 . \end{aligned} \tag{3.5}$$

For definitions and an introduction to viscosity solutions of Hamilton–Jacobi equations, the reader is refered to Crandall and Lions [3], Crandall, Evans and Lions [5].

Consider the following functional, defined on Θ by

$$\ell_\alpha(\theta) \stackrel{\Delta}{=} \inf_{x\in\mathbf{R}^m} W_\alpha^\theta(x,T) = \inf\left\{J_\alpha^\theta(\xi,T) \,:\, \xi \in C([0,T]\,;\,\mathbf{R}^m)\right\} . \tag{3.6}$$

A *deterministic estimate* (DPE) of the unknown parameter θ in the model $\mathcal{M}^0$ on the basis of the observation record $\{y_t^\alpha\,,\,0 \le t \le T\}$ is defined by

$$\widehat{\theta}_\alpha \in M_\alpha \stackrel{\Delta}{=} \operatorname*{argmin}_{\theta\in\Theta} \ell_\alpha(\theta) . \tag{3.7}$$

The main result of this section is the following:

Theorem 3.2 *If the model $\mathcal{M}^0$ is identifiable, then for all $\alpha \in \Theta$*

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

Thus, under the identifiability hypothesis, the DPE is uniquely defined and the unknown parameter can, in principle, be computed exactly from (3.4), (3.6), (3.7). Before proving Theorem 3.2, we give a lemma which ensures that $\operatorname*{argmin}_{\theta\in\Theta} \ell_\alpha(\theta) \neq \emptyset$, and also provides useful estimates.

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

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

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

(ii) for all $R > 0$, $\delta > 0$ there exists $\gamma > 0$ 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 $\theta \longmapsto \ell_\alpha(\theta)$ is continuous.

PROOF. In the sequel, every constant independent of $\theta, \alpha \in \Theta$ and $(x,t) \in \mathbf{R}^m \times [0,T]$ will be 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,

$$|\xi_t|^2 \le \left(|\xi_s|^2 + \Delta \int_s^t |\dot{\xi}_\tau|^2 d\tau\right) \exp\{(t-s)/\Delta\} \,. \tag{3.8}$$

Since $\sup\limits_{\theta\in\Theta} \sup\limits_{x\in\mathbf{R}^m} |h_\theta(x)| \le C$ 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 Lagrangian in (3.2). It is easy to prove the following estimates

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

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

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 $4eC_1\Delta \le \frac{1}{2}$. For $n = 1,\ldots,N$ the Dynamic Programming principle implies

$$W_\alpha^\theta(z,n\Delta) = \inf_\xi \left\{ W(\xi_{(n-1)\Delta},(n-1)\Delta) + \int_{(n-1)\Delta}^{n\Delta} L_\alpha^\theta(\dot{\xi}_s,\xi_s,s) ds \;:\; \xi_{n\Delta} = z \right\} \,.$$

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

$$\begin{aligned} W_\alpha^\theta(\xi_{(n-1)\Delta}^n,(n-1)\Delta) + \int_{(n-1)\Delta}^{n\Delta} L_\alpha^\theta(\dot{\xi}_s^n,\xi_s^n,s) ds \;&\le\; W_\alpha^\theta(\xi_{n\Delta}^n,n\Delta) + \frac{\delta}{N} \\ &\le C_1|\xi_{n\Delta}^n|^2 + C + \frac{\delta}{N} \,. \end{aligned} \tag{3.9}$$

Then

$$\tfrac{1}{4}\int_{(n-1)\Delta}^{n\Delta} |\dot{\xi}_s^n|^2\, ds \le C_1|\xi_{n\Delta}^n|^2 + C + \frac{\delta}{N} \,,$$

and from (3.8)

$$|\xi^n_{n\Delta}|^2 \le \left(|\xi^n_{(n-1)\Delta}|^2 + \Delta\int_{(n-1)\Delta}^{n\Delta} |\dot\xi^n_\tau|^2 d\tau\right) e \le e\,|\xi^n_{(n-1)\Delta}|^2 + \tfrac12|\xi^n_{n\Delta}|^2 + \tfrac12(C+\frac{\delta}{N})/C_1\ ,$$

which implies

$$|\xi^n_{n\Delta}|^2 \le 2e\,|\xi^n_{(n-1)\Delta}|^2 + (C+\frac{\delta}{N})/C_1\ . \tag{3.10}$$

Define $\xi^\theta \in C([0,T]\,;\,\mathbf{R}^m)$ by $\xi^\theta_t = \xi^n_t$ for $t \in [(n-1)\Delta, n\Delta]$, $n = 1,\dots,N$. Then $\xi^\theta_T = x$ and by iterating (3.10) we obtain

$$|x|^2 \le C^N\,|\xi^\theta_0|^2 + C^N\ .$$

Now also, by iterating (3.9)

$$J^\theta_\alpha(\xi^\theta, T) \le W^\theta_\alpha(x,T) + \delta\ , \tag{3.11}$$

and consequently

$$W^\theta_\alpha(x,T) \ge J^\theta_\alpha(\xi^\theta,T) - \delta \ge S^\theta_0(\xi^\theta_0) - C \ge C|x| - C'\ ,$$

which proves (i).

Proof of (ii): Let $R > 0$, $\delta > 0$ and $x \in B(0,R)$. Choose ξ^θ as in (3.11). Then, from the above estimates,

$$\int_0^T |\dot\xi^\theta_s|^2\,ds \le C_R\ .$$

Using (3.8) we deduce that if $x \in B(0,R)$, then there exists $R' > 0$ such that $\xi^\theta_0 \in B(0,R')$ for all $\theta \in \Theta$. Therefore

$$\begin{aligned}
&W^{\theta'}_\alpha(x,T) - W^\theta_\alpha(x,T)\\
&\le\ J^{\theta'}_\alpha(\xi^\theta,T) - J^\theta_\alpha(\xi^\theta,T) + \tfrac14\delta\\
&=\ S^{\theta'}_0(\xi^\theta_0) - S^\theta_0(\xi^\theta_0) + \tfrac12\int_0^T |\dot\xi^\theta_s - b_{\theta'}(\xi^\theta_s)|^2\,ds - \tfrac12\int_0^T |\dot\xi^\theta_s - b_\theta(\xi^\theta_s)|^2\,ds\\
&\quad +\tfrac12\int_0^T |\dot y^\alpha_s - h_{\theta'}(\xi^\theta_s)|^2\,ds - \tfrac12\int_0^T |\dot y^\alpha_s - h_\theta(\xi^\theta_s)|^2\,ds + \tfrac14\delta\ .
\end{aligned}$$

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

$$|S^{\theta'}_0(\xi^\theta_0) - S^\theta_0(\xi^\theta_0)| < \tfrac14\delta\ ,$$

$$\tfrac12\left|\int_0^T |\dot y^\alpha_s - h_{\theta'}(\xi^\theta_s)|^2\,ds - \int_0^T |\dot y^\alpha_s - h_\theta(\xi^\theta_s)|^2\,ds\right| < \tfrac14\delta\ .$$

Also

$$\tfrac{1}{2}\left|\int_0^T |\dot{\xi}_s^\theta - b_{\theta'}(\xi_s^\theta)|^2\,ds - \int_0^T |\dot{\xi}_s^\theta - b_\theta(\xi_s^\theta)|^2\,ds\right|$$

$$\leq \left\{\int_0^T |\dot{\xi}_s^\theta|^2\,ds\right\}^{1/2} \left\{\int_0^T |b_{\theta'}(\xi_s^\theta) - b_\theta(\xi_s^\theta)|^2\,ds\right\}^{1/2}$$

$$+\tfrac{1}{2}\int_0^T |b_{\theta'}(\xi_s^\theta) - b_\theta(\xi_s^\theta)|\,|b_{\theta'}(\xi_s^\theta) + b_\theta(\xi_s^\theta)|\,ds < \tfrac{1}{4}\delta\ ,$$

if $|\theta' - \theta|$ is small enough. Hence, there exists $\gamma > 0$ 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. □

PROOF OF THEOREM 3.2. From (3.2), (3.3) and (3.6) we have

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

for all $\theta \in \Theta$ and $\xi \in C([0,T]\,;\mathbf{R}^m)$, so that $\ell_\alpha(\theta) \geq c_\alpha$. From (3.1) we have

$$J_\alpha^\alpha(x^\alpha,T) = c_\alpha\ ,$$

so that for all $\theta \in \Theta$, $\ell_\alpha(\alpha) = c_\alpha \leq \ell_\alpha(\theta)$. Therefore $\alpha \in M_\alpha$.

Assume that $\widehat{\theta} \in M_\alpha$. Then $\ell_\alpha(\widehat{\theta}) = \ell_\alpha(\alpha)$ and

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

for some $\widehat{\xi} \in C([0,T]\,;\mathbf{R}^m)$, since $J_\alpha^{\widehat{\theta}}(\cdot,T)$ is lower semi-continuous. Then from (3.2)

$$(i)\qquad S_0^{\widehat{\theta}}(\widehat{\xi}_0) = 0\ ,$$

$$(ii)\qquad \dot{\widehat{\xi}}_s = b_{\widehat{\theta}}(\widehat{\xi}_s)\ ,\qquad 0 \leq s \leq T\ ,$$

$$(iii)\qquad \dot{y}_s^\alpha = h_{\widehat{\theta}}(\widehat{\xi}_s)\ ,\qquad 0 \leq s \leq T\ .$$

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

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

Now since the model $\mathcal{M}^0$ is identifiable, this equality forces $\widehat{\theta} = \alpha$, which proves the theorem. □

Remark 3.4 The notion of identifiability is reminiscent of a notion of observability for nonlinear systems, which also has a variational characterization, see James [13] [14].

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 $\{\widehat{\theta}^\varepsilon\,,\, \varepsilon > 0\}$ converges in $P_{\alpha,\varepsilon}$-probability to the deterministic set M_α: for all $\delta > 0$

$$P_{\alpha,\varepsilon}(d(\widehat{\theta}^\varepsilon, M_\alpha) > \delta) \xrightarrow{\varepsilon\downarrow 0} 0\ ,$$

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

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

The proof of this theorem depends on a technical extension of large deviations limit results for nonlinear filtering contained in James and Baras [12], James [13]. 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 $\{\ell^\varepsilon(\theta)\,,\, \varepsilon > 0\}$ converges in $P_{\alpha,\varepsilon}$-probability uniformly in $\theta \in \Theta$ to $\ell_\alpha(\theta)$: for all $\delta > 0$*

$$P_{\alpha,\varepsilon}(\sup_{\theta\in\Theta} |\ell^\varepsilon(\theta) - \ell_\alpha(\theta)| > \delta) \xrightarrow{\varepsilon\downarrow 0} 0\ .$$

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 exists $\gamma > 0$ such that

$$\{\sup_{\theta\in\Theta} |\ell^\varepsilon(\theta) - \ell_\alpha(\theta)| < \gamma\} \subset \{d(\widehat{\theta}^\varepsilon, M_\alpha) < \delta\}\ .$$

Therefore, by Lemma 4.2

$$P_{\alpha,\varepsilon}(d(\widehat{\theta}^\varepsilon, M_\alpha) > \delta) \leq P_{\alpha,\varepsilon}(\sup_{\theta\in\Theta} |\ell^\varepsilon(\theta) - \ell_\alpha(\theta)| > \gamma) \xrightarrow{\varepsilon\downarrow 0} 0\ ,$$

which proves (i).

The proof of (ii) follows at once from (i) and Theorem 3.2. □

As in James and Baras [12], James [13], we employ the vanishing viscosity method of Evans and Ishii [8]. We proceed by a logarithmic change of variables used by Fleming and Mitter [9]. Define

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

The $\mathcal{Y}_t$–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 $\Omega_0 \equiv \{\eta \in C([0,T]\,;\,\mathbf{R}^d) \,:\, \eta_0 = 0\}$, which we denote by $u^{\theta,\varepsilon}[\eta](x,t)$, see [9] and [12]. 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

$$\begin{gathered} \frac{\partial}{\partial t}u^{\theta,\varepsilon}[\eta](x,t) - \tfrac{1}{2}\varepsilon\Delta u^{\theta,\varepsilon}[\eta](x,t) + H^{\theta,\varepsilon}[\eta](x,t,Du^{\theta,\varepsilon}[\eta](x,t)) = 0 \\ u^{\theta,\varepsilon}[\eta](x,0) = S_0^\theta(x) - \varepsilon \log C_{\theta,\varepsilon} \end{gathered} \tag{4.2}$$

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

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

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

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

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

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

Lemma 4.3 *We have*

$$\lim_{\varepsilon \downarrow 0} u^{\theta,\varepsilon}[\eta](x,t) = u^\theta[\eta](x,t) \ ,$$

uniformly in $\theta \in \Theta$ and $t \in [0,T]$ and uniformly on compact subsets of $\eta \in \Omega_0$ and $x \in \mathbf{R}^m$.

PROOF. The following estimates are obtained as in James and Baras [12], James [13], using methods introduced in Evans and Ishii [8], Crandall and Lions [4]. Let $R > 0$ and $K \subset \Omega_0$ be compact. Then if $\varepsilon > 0$ is sufficiently small, we have

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

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

$$|u^{\theta,\varepsilon}[\eta](x,t) - u^{\theta,\varepsilon}[\eta](x,s)| \leq C(\sqrt{\varepsilon}|t-s|^{\frac{1}{2}} + |t-s|)$$

for some constant $C > 0$ and for all $\theta \in \Theta$, $t \in [0,T]$, $\eta \in K$ and $x \in B(0,R)$. By the Arzela–Ascoli theorem, there is a subsequence $\varepsilon_k \downarrow 0$ such that $u^{\theta,\varepsilon_k}[\eta]$ converges uniformly on $B(0,R) \times [0,T]$ to a continuous function w. This function satisfies the Hamilton–Jacobi equation (4.4), and by uniqueness, $w = u^\theta[\eta]$ (Crandall and Lions [3]). Hence $u^{\theta,\varepsilon}[\eta] \to u^\theta[\eta]$ as $\varepsilon \downarrow 0$.

Now $u^\theta[\eta]$ is a continuous function of $\eta \in K$, $\theta \in \Theta$ (see the proof of Lemma 3.3 (ii)). Using this fact and the uniform estimate above we conclude that the convergence is uniform. □

Now

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

and

$$W^\theta_\alpha(x,t) = u^\theta[y^\alpha](x,t) - h^*_\theta(x)y^\alpha_t \ .$$

Lemma 4.4 *We have*

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

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 $\delta > 0$

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

Choose $\beta > 0$ such that $\|\eta - y^\alpha\| < \beta$ implies

$$\sup_{\theta\in\Theta} \rho(u^\theta[\eta], u^\theta[y^\alpha]) < \tfrac{1}{2}\delta \ .$$

From Lemma 4.3, if $\|\eta - y^\alpha\| < \beta$ and $0 < \varepsilon \leq \varepsilon_0$ then

$$\sup_{\theta\in\Theta} \rho(u^{\theta,\varepsilon}[\eta], u^\theta[\eta]) < \tfrac{1}{2}\delta \ .$$

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

$$
\begin{aligned}
& P_{\alpha,\varepsilon}(\sup_{\theta\in\Theta} \rho(u^{\theta,\varepsilon}[Y], u^\theta[y^\alpha]) > \delta) \\
\le\ & P_{\alpha,\varepsilon}(\sup_{\theta\in\Theta} \rho(u^{\theta,\varepsilon}[Y], u^\theta[Y]) > \tfrac{1}{2}\delta\,;\ \|Y - y^\alpha\| < \beta) \\
& + P_{\alpha,\varepsilon}(\sup_{\theta\in\Theta} \rho(u^{\theta}[Y], u^\theta[y^\alpha]) > \tfrac{1}{2}\delta\,;\ \|Y - y^\alpha\| < \beta) \\
& + P_{\alpha,\varepsilon}(\|Y - y^\alpha\| > \beta) \le P_{\alpha,\varepsilon}(\|Y - y^\alpha\| > \beta) \xrightarrow{\varepsilon\downarrow 0} 0\ ,
\end{aligned}
$$

by (2.4). □

PROOF OF LEMMA 4.2. Recall from (2.3) and (3.6) that

$$
\begin{aligned}
\ell^\varepsilon(\theta) &= -\varepsilon \log \int_{\mathbf{R}^m} \exp\left\{-\frac{1}{\varepsilon} W^{\theta,\varepsilon}(x,T)\right\} dx \quad \text{a.s.} \\
\ell_\alpha(\theta) &= \inf_{x\in\mathbf{R}^m} W_\alpha^\theta(x,T)\ .
\end{aligned}
$$

From the proof of Lemma 2.1 we see that

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

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

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

From Lemma A.3, there exists $\varepsilon_0 > 0$, $\beta > 0$ and $c > 0$ such that $0 < \varepsilon \le \varepsilon_0$

$$
\sup_{\theta\in\Theta} \rho(W^{\theta,\varepsilon}, W_\alpha^\theta) < \beta \quad \text{and} \quad \|Y - y^\alpha\| < c
$$

implies

$$
\sup_{\theta\in\Theta} |\ell^\varepsilon(\theta) - \ell_\alpha(\theta)| < \delta\ .
$$

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

$$
\begin{aligned}
& P_{\alpha,\varepsilon}(\sup_{\theta\in\Theta} |\ell^\varepsilon(\theta) - \ell_\alpha(\theta)| > \delta) \\
\le\ & P_{\alpha,\varepsilon}(\sup_{\theta\in\Theta} \rho(W^{\theta,\varepsilon}, W_\alpha^\theta) > \beta) + P_{\alpha,\varepsilon}(\|Y - y^\alpha\| > c) \xrightarrow{\varepsilon\downarrow 0} 0\ ,
\end{aligned}
$$

by (2.4) and Lemma 4.4. □

5 Binary Sequential Detection

In this section we discuss some aspects of a binary detection problem studied by Baras and LaVigna [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 Baras and LaVigna [1] is the following

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

where

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

and

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

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 $\mathcal{M}^0$ defined by (3.1) is identifiable. In fact, if

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

then $\mathcal{M}^0$ 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 $T > \sigma$. Define the MLE $\hat{\theta}^\varepsilon$ 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 LaVigna use a threshold decision policy to decide which of the hypotheses is valid. Define the likelihood ratio

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

Note that as $\varepsilon \downarrow 0$,

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

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

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

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

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 (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 LaVigna [1] prove that

$$\tau^\varepsilon < \infty \quad \text{a.s.}$$

and

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

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 $\mathcal{M}^\varepsilon = \{P_{0,\varepsilon}\,,\, P_{1,\varepsilon}\}$.

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

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

Let $a < 0 < b$ and set

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

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

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), there exists $T_1 > 0$ 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 LaVigna [1] use a numerical solution of the Zakai 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.

Remark 5.4 In practice, the initial condition x_0 is not known, so that one would have to estimate x_0 also, for instance using an observer.

References

[1] J.S. BARAS and A. LA VIGNA, Real time sequential detection for diffusion signals, *preprint.*

[2] F. CAMPILLO and F. LE GLAND, MLE for partially observed diffusions: direct maximization vs. the EM algorithm, *Stochastic Processes and Applications* **33** (2) 245–274 (1989).

[3] M.G. CRANDALL and P.L. LIONS, Viscosity solutions of Hamilton–Jacobi equations, *Transactions of the AMS* **277** (1) 1–42 (1983).

[4] M.G. CRANDALL and P.L. LIONS, Two approximations of solutions of Hamilton–Jacobi equations, *Mathematics of Computation* **43** (167) 1–19 (1984).

[5] M.G. CRANDALL, L.C. EVANS and P.L. LIONS, Some properties of viscosity solutions of Hamilton–Jacobi equations, *Transactions of the AMS* **282** (2) 487–502 (1984).

[6] M.H.A. DAVIS, On a multiplicative functional transformation arising in nonlinear filtering theory, *Z. Wahrsch. Verw. Gebiete* **54** (2) 125–139 (1980).

[7] A. DEMBO and O. ZEITOUNI, Parameter estimation of partially observed continuous-time stochastic processes via the EM algorithm, *Stochastic Processes and Applications* **23** (1) 91–113 (1986).

[8] L.C. EVANS and H. ISHII, A PDE approach to some asymptotic problems concerning random differential equations with small noise intensities, *Annales de l'Institut Henri Poincaré–Analyse Non Linéaire* **2** (1) 1–20 (1985).

[9] W.H. FLEMING and S.K. MITTER, Optimal control and nonlinear filtering for nondegenerate diffusion processes, *Stochastics* **8** (1) 63–77 (1982).

[10] M.I. FREIDLIN and A.D. WENTZELL, *Random Perturbations of Dynamical Systems*, Springer–Verlag (1984).

[11] O. HIJAB, Asymptotic Bayesian estimation of a first order equation with small diffusion, *Annals of Probability* **12** (3) 890–902 (1984).

[12] M.R. JAMES and J.S. BARAS, Nonlinear filtering and large deviations: A PDE-control theoretic approach, *Stochastics* **23** (3) 391–412 (1988).

[13] M.R. JAMES, *Asymptotic Nonlinear Filtering and Large Deviations with Application to Observer Design*, Ph.D. Dissertation, University of Maryland, May 1988. *(Technical Report SRC–TR–88–28, Systems Research Center).*

[14] M.R. JAMES, Finite time observer design by probabilistic–variational methods, *Report # 89–12, Lefschetz Center for Dynamical Systems.*

[15] R.S. LIPTSER and A.N. SHIRYAYEV, *Statistics of Random Processes*, Springer–Verlag (1977).

A Appendix

This Appendix contains some technical results used in the paper, and a proof of Lemma 2.1.

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

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

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

$$\sup_{\lambda \in \Lambda} |f(\lambda) - g(\lambda)| < \beta \quad \textit{implies} \quad \forall \lambda \in M(g) \ , \ d(\lambda, M(f)) < \alpha \ .$$

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

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

and

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

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

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

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

$$\begin{aligned} f(\widehat{\lambda}_i) &= f(\widehat{\lambda}(f)) + [g_i(\widehat{\lambda}(f)) - f(\widehat{\lambda}(f))] + [g_i(\widehat{\lambda}_i) - g_i(\widehat{\lambda}(f))] + [f(\widehat{\lambda}_i) - g_i(\widehat{\lambda}_i)] \\ &\leq f(\widehat{\lambda}(f)) + [g_i(\widehat{\lambda}(f)) - f(\widehat{\lambda}(f))] + [f(\widehat{\lambda}_i) - g_i(\widehat{\lambda}_i)] \\ &\leq f(\widehat{\lambda}(f)) + 2 \sup_{\lambda \in \Lambda} |f(\lambda) - g_i(\lambda)| \ , \end{aligned}$$

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

Lemma A.2 *Let $\Lambda \subset \mathbf{R}^p$ be compact, and $F^\lambda \in C(\mathbf{R}^m, \mathbf{R})$ 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 exists $\gamma > 0$ 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) there exists a constant $R > 0$ 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 exists $\gamma > 0$ 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

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

which proves (i).

By (b) again, this implies

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

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 $F^\lambda, G^\lambda \in C(\mathbf{R}^m, \mathbf{R})$ be such that*

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

$$F^\lambda(z) \geq C|z| - C' , \quad G^\lambda(z) \geq C|z| - C' ,$$

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

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

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 exists $\beta > 0$, $\varepsilon_0 > 0$ (depending on G) such that $0 < \varepsilon \leq \varepsilon_0$ and

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

implies

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

PROOF. Define

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

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

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

for all $\lambda \in \Lambda$. Thus we can choose $0 < \beta < \delta/12$ 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 there exists $r > 0$ 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 $B(z^\lambda, r) \subset B^\lambda_\delta$ for all $\lambda \in \Lambda$. Therefore

$$\infty > v_R \geq \mu(B^\lambda_\delta) \geq v_r > 0 \ ,$$

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

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

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

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

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

We turn now to the

PROOF OF LEMMA 2.1. From Sections 2 and 4 we have

$$\ell^\varepsilon(\theta) = -\varepsilon \log \int_{\mathbf{R}^m} q^{\theta,\varepsilon}(x,T) \exp\{\frac{1}{\varepsilon} h_\theta^*(x) Y_T\}\, dx \quad \text{a.s.},$$

where for a.e. $\omega \in \Omega$, $q^{\theta,\varepsilon} \in C_b^{1,2}(\mathbf{R}^m \times [0,T])$ and solves the "robust" Zakai equation

$$\frac{\partial}{\partial t} q^{\theta,\varepsilon}(x,t) - \tfrac{1}{2}\varepsilon \Delta q^{\theta,\varepsilon}(x,t) + \tilde{g}_\theta^*(x,t) D q^{\theta,\varepsilon}(x,t) + \frac{1}{\varepsilon}\tilde{V}^{\theta,\varepsilon}(x,t) q^{\theta,\varepsilon}(x,t) = 0 \ ,$$

$$q^{\theta,\varepsilon}(x,0) = p_0^{\theta,\varepsilon}(x) \ ,$$

with

$$\tilde{V}^{\theta,\varepsilon}(x,t) \triangleq \tilde{V}^\theta(x,t) + \tfrac{1}{2}\varepsilon Y_t^* \Delta h_\theta(x) + \varepsilon \mathrm{div} \tilde{g}_\theta(x,t) \ ,$$

$$\tilde{V}^\theta(x,t) \triangleq \tfrac{1}{2}|h_\theta(x)|^2 + b_\theta^* Y_t^* D h_\theta(x) - \tfrac{1}{2}(Dh_\theta(x))^* Y_t Y_t^* Dh_\theta(x) \ ,$$

$$\tilde{g}_\theta(x,t) \triangleq b_\theta(x) - Y_t^* D h_\theta(x) \ ;$$

see Davis [6]. Fix $\varepsilon > 0$ and $\omega \in \Omega$ such that the above holds. Now $|\tilde{g}_\theta(x,t)| \leq C$ and $|\tilde{V}^{\theta,\varepsilon}(x,t)| \leq C$ in $\mathbf{R}^m \times [0,T]$. Then

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

and by the maximum principle, for all $(x,t) \in \mathbf{R}^m \times [0,T]$

$$0 \leq q^{\theta,\varepsilon}(x,t) \leq \exp\{\frac{CT}{\varepsilon}\} p_0^{\theta,\varepsilon}(x) \leq \exp\{-\frac{1}{\varepsilon}(C_2|x| - C_2' - CT)\} \ ,$$

i.e.

$$W^{\theta,\varepsilon}(x,t) \geq C_2|x| - C_2' - CT \ ,$$

where C is random and satisfies the following estimate

$$C \le C_0 \sup_{0\le t\le T} |Y_t - y_t^\alpha|^2 + C_\alpha \ .$$

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 $z \triangleq q^{\theta_k,\varepsilon} - q^{\theta_0,\varepsilon}$ satisfies

$$\frac{\partial}{\partial t} z(x,t) - \tfrac{1}{2}\varepsilon \Delta z(x,t) + \tilde{g}^*_{\theta_k}(x,t) Dz(x,t) + \frac{1}{\varepsilon}\tilde{V}^{\theta_k,\varepsilon}(x,t) z(x,t)$$

$$= -[\tilde{g}_{\theta_k}(x,t) - \tilde{g}_{\theta_0}(x,t)]^* Dq^{\theta_0,\varepsilon}(x,t) - \frac{1}{\varepsilon}[\tilde{V}^{\theta_k,\varepsilon}(x,t) - \tilde{V}^{\theta_0,\varepsilon}(x,t)]\, q^{\theta_0,\varepsilon}(x,t) \ ,$$

and hence

$$\frac{\partial}{\partial t} z(x,t) - \tfrac{1}{2}\varepsilon \Delta z(x,t) + \tilde{g}^*_{\theta_k}(x,t) Dz(x,t) - \frac{1}{\varepsilon} C z(x,t) \le C_{\theta_0} \rho(\theta_k,\theta_0)(1 + \frac{1}{\varepsilon}) \ ,$$

where $\rho(\theta_k,\theta_0) \to 0$ as $k \to \infty$. Then by the maximum principle

$$z(x,t) \le \exp\{\frac{CT}{\varepsilon}\}\, z(x,0) + T\, \exp\{\frac{CT}{\varepsilon}\}\, C_{\theta_0} \rho(\theta_k,\theta_0)(1 + \frac{1}{\varepsilon}) \ .$$

Now

$$z(x,0) \le \frac{C}{\varepsilon}\, \exp\{-\frac{1}{\varepsilon}(C_2|x| - C_2')\}\, |S_0^{\theta_0}(x) - S_0^{\theta_k}(x)| \ .$$

Consequently, sending $k \to \infty$ we obtain

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

Similarly, we obtain the reverse inequality and conclude. □

Imprimé en France
par
l'Institut National de Recherche en Informatique et en Automatique

ISSN 0249-6399