Asymptotic Properties of an Approximate Maximum Likelihood Estimator for Stochastic PDEs

M. Huebner* S. Lototsky† B.L. Rozovskii‡

In: Yu. M. Kabanov, B. L. Rozovskii, and A. N. Shiryaev (editors), *Statistics and Control of Stochastic Processes (the Liptser festschrift)*, pp. 139–155. World Scientific, Singapore, 1997.

Abstract

A maximum likelihood estimate of a scalar parameter is constructed for a stochastic evolution system using the Galerkin approximation of the original equation. Conditions are established to guarantee the consistency and asymptotic normality of the estimator as the dimension of the approximation tends to infinity. Examples are given to illustrate the results.

1 Introduction

Many parameter estimation problems for stochastic partial differential equations can be reduced to the following model:

$$du(t,\omega) = (A_0(t)u(t,\omega) + \theta A_1(t)u(t,\omega) + f(t))\,dt + BdW(t), \tag{1.1}$$

where A_0, A_1, and B are linear operators, W is a cylindrical Brownian motion, all defined in some Hilbert space H, and θ is an unknown scalar parameter. A computable estimate of θ based on the observations of u can usually involve only a finite-dimensional approximation of the random field u. If the operators A_0, A_1, and B have a common system of eigenfunctions, then equation (1.1) is diagonalizable so that the maximum likelihood estimator (MLE) of θ can be easily constructed and its properties studied. In [1, 2] model (1.1) is considered under the assumption that A_0, A_1, and B are elliptic, selfadjoint operators with a common system of eigenfunctions in a bounded domain of $\mathbb{R}^d$. It was demonstrated in these works that under certain conditions the MLE is consistent, asymptotically normal, and asymptotically efficient as the dimension of

*Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824. This work was partially supported by NSF Grant #DMS 9509000

†Center for Applied Mathematical Sciences, University of Southern California, Los Angeles, CA 90089-1113. This work was partially supported by ONR Grant #N00014-95-1-0229.

‡Center for Applied Mathematical Sciences, University of Southern California, Los Angeles, CA 90089-1113. This work was partially supported by ONR Grant #N00014-95-1-0229 and ARO Grant DAAH 04-95-1-0164.

the approximation tends to infinity. It was mentioned in [2] that some of the results remain valid if the operators in (1.1) do not commute.

In this paper, equation (1.1) is studied in the general Hilbert space setting without any assumptions on the eigenfunctions of the operators A_0 and A_1, and the MLE of θ is based on the Galerkin approximation of (1.1). The objective is to establish conditions under which the MLE is consistent and asymptotically normal as the dimension of the approximation tends to infinity. These conditions are related to the separation of the corresponding sequences of the probability measures and can be verified in a number of examples (Section 4).

Unless the system is diagonalizable, the Galerkin approximation is usually not observable and therefore the results presented in this paper are mainly of theoretical interest. However the conditions for asymptotic optimality for maximum likelihood estimators obtained with this approach can be applied to approximate maximum likelihood estimators based on other finite-dimensional approximations of (1.1).

2 Stochastic Evolution Systems in Hilbert Scales

Let H be a real separable Hilbert space with the inner product $(\cdot,\cdot)_0$ and the corresponding norm $\|\cdot\|_0$. Consider a linear operator Λ on H such that $\|\Lambda f\|_0 \geq c\|f\|_0$ for all f from the domain of Λ. Then the powers Λ^s of Λ are defined for all $s \in I\!R$ and generate the spaces H^s as follows [3]:

— for $s \geq 0$, H^s is the domain of Λ^s;

— for $s < 0$, H^s is the completion of H with respect to the norm $\|\cdot\|_s := \|\Lambda^s \cdot\|_0$.

The collection of spaces $\{H^s\}_{s\in I\!R}$ is a *Hilbert scale* with the following properties [3]:

— for $s_2 > s_1$, the space H^{s_2} is continuously embedded in H^{s_2};

— for every $s \in I\!R$ and $r > 0$, the spaces (H^{s+r}, H^s, H^{s-r}) form a normal triple with the canonical bilinear functional
$\langle y_1, y_2\rangle_{s,r} = (\Lambda^{s-r}y_1, \Lambda^{s+r}y_2)_0$, where $y_1 \in H^{s-r},\ y_2 \in H^{s+r}$;

— $\Lambda^s(H^r) = H^{s-r}$.

Let $(\Omega, \mathcal{F}, \mathbf{F} = (\mathcal{F}_t), P)$ be a filtered probability space with the usual assumtions, and let $W = (W(t))_{t\geq 0}$ be a *cylindrical Brownian motion* on this space. This means that a family of continuous martingales $W_t(f),\ f \in H$, is defined on $(\Omega, \mathcal{F}, \mathbf{F}, \mathrm{P})$ so that the quadratic covariation $\langle W.(f), W.(g)\rangle_t = t(f,g)_0$ for every $f, g \in H$. The stochastic integral $\int_0^T (\xi_t, dW(t))_0$ is defined [4, 5] for predictable processes $\xi = (\xi_t)_{0\leq t\leq T} \in L_2(\Omega\times[0,T]; H)$, that is $(\xi_t, y)_0$ is a predictable stochastic process for every $y \in H$.

Consider the following stochastic evolution system:

$$\begin{array}{rcl} du^\theta(t,\omega) & = & (A^\theta(t)u(t,\omega) + f(t))dt + B\,dW(t),\ 0 < t \leq T, \\ u^\theta|_{t=0} & = & u_0, \end{array} \tag{2.1}$$

where θ belongs to some parameter space $\Theta \subset I\!R$, and $A^\theta = A_0 + \theta A_1$. The linear operators $A_0,\ A_1$, and B are deterministic. Equation (2.1) will be studied under the following **assumptions**:

(A1) There exists $\alpha \geq 0$ such that $\Lambda^{-\alpha}B$ is a Hilbert-Schmidt operator in H;

(A2) There exists $\gamma \geq 0$ such that for every $t \in [0,T]$ the operators A_0 and A_1 are bounded and linear from $H^{\gamma-\alpha}$ to $H^{-\gamma-\alpha}$;

(A3) For every $\theta \in \Theta$ there exist positive numbers $\delta = \delta(\theta)$ and $K = K(\theta)$ such that for all $t \in [0,T]$ and $y \in H^{\gamma-\alpha}$,

$$\langle A^\theta y, y\rangle_{-\alpha,\gamma} \leq -\delta \|y\|^2_{\gamma-\alpha} + K\|y\|^2_{-\alpha}, \text{ and } \|A^\theta y\|_{-\gamma-\alpha} \leq K\|y\|_{\gamma-\alpha};$$

(A4) u_0 is an $\mathcal{F}_0$-measurable random element with values in $H^{-\alpha}$ and $\mathbf{E}\|u_0\|^2_{-\alpha} < \infty$;

(A5) $f = f(t)$ is an $H^{-\gamma-\alpha}$-valued deterministic function and $\int_0^T \|f(t)\|^2_{-\gamma-\alpha} < \infty$.

Assumption (A1) implies that $BW(t)$ is an $H^{-\alpha}$-valued Wiener process, therefore by Theorem 3.1.4 in [6] for every $\theta \in \Theta$ equation (2.1) has a unique solution u^θ in the normal triple $(H^{\gamma-\alpha}, H^{-\alpha}, H^{-\gamma-\alpha})$ and

$$u^\theta \in L_2(\Omega; \mathbf{C}([0,T]; H^{-\alpha})) \cap L_2(\Omega \times [0,T]; H^{\gamma-\alpha}). \tag{2.2}$$

The solution u satisfies

$$\mathbf{E} \sup_{0\leq t\leq T} \|u(t)\|^2_{-\alpha} + \mathbf{E}\int_0^T \|u(t)\|^2_{\gamma-\alpha} dt < \infty. \tag{2.3}$$

The approximate maximum likelihood estimator for the parameter θ will be based on a finite-dimensional approximation of (2.1). To study the properties of this estimator, it is further assumed that

(A6) Operators Λ and B have a common system of eigenfunctions $\{h_k\}_{k\geq 0}$ with the following properties:
 — $\{h_k\}_{k\geq 0}$ is a complete orthonormal system in H;
 — $h_k \in \cap_s H^s$ for all $k \geq 0$;

(A7) The eigenvalues of B are nonzero, and the operator A_1 is not identically zero.

3 Consistency and Asymptotic Normality of the Approximate Maximum Likelihood Estimator

For a positive integer N the operator Π^N acting from $\cup_s H^s$ to $\cap_s H^s$ is defined by $\Pi^N : u \mapsto \sum_{k=0}^N (u, h_k)_0\, h_k$, where $\{h_k\}_{k\geq 0}$ is the orthonormal basis of H defined in (A6). Then Π^N is an orthogonal projection in the space H, and $\|\Pi^n u\|_s \leq \|u\|_s$ for all $s \in I\!R$.

The finite-dimensional Galerkin approximation of (2.1) is defined by

$$u^{\theta,N}(t) = \Pi^n u_0 + \int_0^t \Pi^N \Big(A^\theta(s) u^{\theta,N}(s) + f(s)\Big) ds + \Pi^N B W^N(t), \tag{3.1}$$

where $W^N = W^N(t)$ is a standard Wiener process with the components $W_t(h_k),\ 0 \leq k \leq N$. It follows from assumption (A6) that $\Pi^N B$ is a diagonal matrix whose diagonal elements are the first N eigenvalues of the operator B.

Lemma 3.1. *Under the assumptions (A1)–(A6),*

$$\lim_{N\to\infty} \mathbf{E}\int_0^T \|u^{\theta,N}(t) - u^\theta(t)\|^2_{\gamma-\alpha} dt = 0. \tag{3.2}$$

Proof. Property (2.2) implies that

$$\lim_{N\to\infty} \mathbf{E}\int_0^T \|\Pi^N u^\theta(t) - u^\theta(t)\|^2_{\gamma-\alpha} dt = 0.$$

On the other hand, for $N \to \infty$, assumption (A3) and the Gronwall inequality imply that

$$\begin{aligned}\mathbf{E}\int_0^T \|u^{\theta,N}(t) - \Pi^N u^\theta(t)\|^2_{\gamma-\alpha} dt &\le C(\theta)\mathbf{E}\int_0^T \|\Pi^N A^\theta(u(t) - \Pi^N u(t))\|^2_{-\alpha} dt \\ &\le C(\theta)\mathbf{E}\int_0^T \|u^\theta(t) - \Pi^N u^\theta(t)\|^2_{\gamma-\alpha} dt \to 0,\end{aligned}$$

and (3.2) follows.

□

To simplify the future presentation, the solutions of (2.1) and (3.1) corresponding to the unknown (but fixed) value of the parameter $\theta_0 \in \Theta$ will be denoted by u and u^N, respectively. The objective is to estimate θ_0 given the trajectory of $u^N(t),\ 0 \le t \le T$.

Denote by $P_t^{\theta,N}$ the measure on $\mathbf{C}([0,t]; H^\alpha)$ generated by the solution $u^{\theta,N}(s)$ of (3.1) for $0 \le s \le t$. Since for each $\theta \in \Theta$ the process $u^{\theta,N}$ is finite dimensional, assumption (A7) and the results from [7] Chapter 7 imply that for each $\theta \in \Theta$ the measure $P_t^{\theta,N}$ is absolutely continuous with respect to the measure $P_t^{\theta_0,N}$ with the likelihood ratio (Radon-Nikodym derivative) given by

$$\begin{aligned}\frac{dP_t^{\theta,N}}{dP_t^{\theta_0,N}}(u^N) = \exp\Bigg\{&(\theta-\theta_0)\int_0^t \Big(\Pi^N B^{-2} A_1(s) u^N(s), du^N(s)\Big)_0 \\ &-\frac{\theta^2-\theta_0^2}{2}\int_0^t \|\Pi^N B^{-1} A_1(s) u^N(s)\|_0^2 ds \\ -(\theta-\theta_0)\int_0^t &\Bigg(\Pi^N B^{-1} A_1(s) u^N(s), \Pi^N B^{-1}(A_0(s) u^N(s) + f(s)\Bigg)_0 ds\Bigg\}.\end{aligned} \tag{3.3}$$

The maximum likelihood estimate (MLE) $\hat\theta^N$ of θ_0 is obtained by maximizing the likelihood ratio (3.3), where t is replaced by T, with respect to $\theta \in \Theta$.

$$\hat\theta^N = \frac{\int_0^T \Big(\Pi^N B^{-2} A_1(t) u^N(t), du^N(t) - \Pi^N\Big(A_0(t)u^N(t) + f(t)\Big)dt\Big)_0}{\int_0^T \|\Pi^N B^{-1} A_1(s) u^N(t)\|_0^2 dt}. \tag{3.4}$$

By convention, $\hat\theta^N = 0$ if $\int_0^T \|\Pi^N B^{-1} A_1(s) u^N(t)\|_0^2 dt = 0$.
If $\int_0^T \|\Pi^N B^{-1} A_1(s) u^N(t)\|_0^2 dt > 0$, then (3.4) implies that $\hat\theta^N$ satisfies

$$\hat\theta^N = \theta_0 + \frac{\int_0^T \Big(\Pi^N B^{-1} A_1(t) u^N(t), dW^N(t)\Big)_0}{\int_0^T \|\Pi^N B^{-1} A_1(t) u^N(t)\|_0^2 dt}. \tag{3.5}$$

It follows from assumption (A7) that the operator $\Pi^N B^{-1}A_1(t)$ is not identical zero for all N that are greater or equal to some N_0. Thus, by Lemma 7.1 in [7],

$$\mathrm{P}\Big(\int_0^T \|\Pi^N B^{-1}A_1(t)u^N(t)\|_0^2 dt > 0\Big) = 1,\ N \geq N_0,$$

so that (3.5) holds with probability 1 for all sufficiently large N.

Direct computations show that Fisher's information $I_N(\theta_0)$ corresponding to the likelihood ratio (3.3) with $t = T$ is given by

$$I_N(\theta_0) = \mathbf{E}\int_0^T \|\Pi^N B^{-1}A_1(t)u^{\theta_0,N}(t)\|_0^2 dt. \tag{3.6}$$

Theorem 3.1.

(i) If

$$\mathrm{P} - \lim_{N\to\infty}\int_0^T \|\Pi^N B^{-1}A_1(t)u^N(t)\|_0^2 dt = \infty, \tag{3.7}$$

then

$$\mathrm{P} - \lim_{N\to\infty} \hat{\theta}^N = \theta_0.$$

If, in addition,

$$\mathrm{P} - \lim_{N\to\infty} (I_N(\theta_0))^{-1}\int_0^T \|\Pi^N B^{-1}A_1(t)u^N(t)\|_0^2 dt = 1, \tag{3.8}$$

then $\hat{\theta}^N$ is asymptotically normal with normalizing factor $\sqrt{I_N(\theta_0)}$, i.e., $\sqrt{I_N(\theta_0)}(\theta-\theta_0)$ converges in distribution, as $N \to \infty$, to a Gaussian random variable with zero mean and unit variance.

(ii) If

$$\int_0^T \|B^{-1}A_1(t)g(t)\|_0^2 dt \leq C\int_0^T \|g(t)\|_{\gamma-\alpha}^2 dt \tag{3.9}$$

for all $g \in L_2([0,T]; H^{\gamma-\alpha})$, then

$$\mathrm{P} - \lim_{N\to\infty} \hat{\theta}^N = \theta_0 + \frac{\displaystyle\int_0^T (B^{-1}A_1(t)u(t), dW(t))_0}{\displaystyle\int_0^T \|B^{-1}A_1(t)u(t)\|_0^2 dt}. \tag{3.10}$$

Proof. *Part (i).* Consistency of $\hat{\theta}^N$ follows from (3.5) and the following result:

Lemma 3.2. *If $\{f_n(t)\}_{n\geq 1}$ is a sequence of predictable random functions in $L_2(\Omega \times [0,T]; H)$ such that*

$$P - \lim_{N\to\infty}\int_0^T \|f_n(t)\|_0^2 dt = \infty\ ,$$

then

$$P - \lim_{N\to\infty} \frac{\displaystyle\int_0^T \Big(f_n(t), dW(t)\Big)_0}{\displaystyle\int_0^T \|f_n(t)\|_0^2\, dt} = 0.$$

This result is not new, but, for the sake of completeness, we give a short proof in the Appendix.

To prove the asymptotic normality, consider

$$M_t^N := \frac{\int_0^t (\Pi^N B^{-1} A_1(s) u^N(s), dW^N(s))_0 ds}{\sqrt{\mathbf{E}\int_0^T \|\Pi^N B^{-1} A_1(s) u^N(s)\|_0^2\, ds}}.$$

Then $(M_t^N, \mathcal{F}_t)_{0\le t\le T}$ is a continuous square integrable martingale with quadratic characteristic

$$\langle M^N\rangle_t = \frac{\int_0^t \|\Pi^N B^{-1} A_1(s) u^N(s)\|_0^2 ds}{\mathbf{E}\int_0^T \|\Pi^N B^{-1} A_1(s) u^N(s)\|_0^2\, ds}.$$

By assumption (3.8), $\mathrm{P}-\lim_{N\to\infty}\langle M^N\rangle_T = 1$.

On the other hand, if $(w_1(t), \mathcal{F}_t)_{0\le t\le T}$ is a one dimensional Wiener process and $M_t := w_1(t)/\sqrt{T}$, then $(M_t, \mathcal{F}_t)_{0\le t\le T}$ is a continuous square integrable martingale and $\langle M\rangle_T = 1$. Therefore, by Theorem 5.5.4(II) in [8]

$$\lim_{N\to\infty} M_T^N = M_T$$

in distribution. Since M_T is a Gaussian random variable with zero mean and unit variance, the asymptotic normality follows.

Part (ii). The right-hand side of (3.10) is well defined because the process $B^{-1}A_1(t)u(t)$ is in the space $L_2(\Omega\times[0,T];H)$, by (3.9) and (2.2), and, for every $y\in H$, the process $(B^{-1}A_1(t)u(t), y)_0$ is continuous and thus predictable. It suffices to show that

$$\lim_{N\to\infty} \mathbf{E}\int_0^T \|\Pi^N B^{-1} A_1(t) u^N(t) - B^{-1}A_1(t)u\|_0^2 dt = 0. \tag{3.11}$$

Then the assertion (3.10) follows from (3.5) and the properties of the stochastic integral [4, 6]. By the triangle inequality,

$$\begin{aligned}\mathbf{E}&\int_0^T \|\Pi^N B^{-1} A_1(t) u^N(t) - B^{-1}A_1(t)u(t)\|_0^2\, dt\\ &\le 2\mathbf{E}\int_0^T \|\Pi^N B^{-1} A_1(t) u^N(t) - \Pi^N B^{-1}A_1(t)u\|_0^2\, dt\\ &\quad + 2\mathbf{E}\int_0^T \|\Pi^N B^{-1} A_1(t) u(t) - B^{-1}A_1(t)u\|_0^2\, dt.\end{aligned} \tag{3.12}$$

The first term on the right of (3.12) tends to zero as $N\to\infty$ since $B^{-1}A_1 u \in L_2(\Omega\times[0,T];H)$, and the second term tends to zero by (3.9) and (3.2). This proves (3.11).

□

The conditions (3.7) and (3.9) of Theorem 2.1 hold for a large class of SPDE's. In [2] they were verified for elliptic operators A_0 and A_1 which have the same eigenfunctions as the operators Λ and B. Other examples to which Theorem 2.1 applies are given in Section 4 below.

Despite their technical nature, conditions (3.7) and (3.9) can be related to some standard statistical notions.

Definition. Let $\{P_n\}_{n\geq 1}$ and $\{\tilde{P}_n\}_{n\geq 1}$ be sequences of probability measures on measurable spaces $(\Omega_n, \mathcal{F}_n)$. The sequences are called *completely separated* if there exist sets $A_n \in \mathcal{F}_n$ such that $\lim_{n\to\infty} P_n(A_n) = 1$ and $\lim_{n\to\infty} \tilde{P}_n(A_n) = 0$. The sequences are called *entirely separated* if there is a sequence $n_k \uparrow \infty,\ k \to \infty$ such that the sequences $\{P_{n_k}\}_{k\geq 1}$ and $\{\tilde{P}_{n_k}\}_{k\geq 1}$ are completely separated.

Theorem 3.2. *(1) If the sequences of measures $\{P_T^{\theta,N}\}_{N\geq 1}$ and $\{P_T^{\theta_0,N}\}_{N\geq 1}$ are completely separated, then (3.7) holds.*

(2) If (3.7) holds, then $\{P_T^{\theta,N}\}_{N\geq 1}$ and $\{P_T^{\theta_0,N}\}_{N\geq 1}$ are entirely separated.

Proof. It follows from Corollary IV.1.37 in [9] that the Hellinger process $h_t^N(1/2)$ corresponding to the likelihood ratio (3.3) and the reference measure $P_T^{\theta_0,N}$ is given by

$$h_t^N(1/2) = \frac{(\theta-\theta_0)^2}{8}\int_0^T \|\Pi^N B^{-1}A_1(t)u^N(t)\|_0^2 dt.$$

The first statement of the theorem is proved in the same way as property (i) from Theorem V.2.4a in [9] and the second statement follows from [9] Theorem V.2.4b.

□

Denote by P^θ the measure on $\mathbf{C}([0,T]; H^{-\alpha})$ generated by the solution of (2.1).

Theorem 3.3. *The measures P^θ are equivalent for all $\theta \in \Theta$ if (3.9) holds.*

Proof. It follows from Corollary 1 in [4] that the measures P^{θ_1} and P^{θ_2} are equivalent if and only if

$$\mathbf{E}\int_0^T \|B^{-1}A_1(t)u^\theta(t)\|_0^2 dt < \infty \tag{3.13}$$

for $\theta = \theta_1$ and $\theta = \theta_2$. This follows from the conditions (3.9) and (2.2).

□

For the model with commuting operators discussed in [2] the conditions (3.7) and (3.9) of Theorem 3.1 are equivalent to the absolute continuity or singularity of the measures P^θ for different $\theta \in \Theta$, respectively.

4 Examples

In this section, Theorem 3.1 is applied to three parameter estimation models of the type (2.1). To simplify the notation, the superscript θ is omitted:

1. $du = (\theta\Delta + a(x))udt + (1-\Delta)^{-1/2}dW,\ x \in [0,1]$, with zero boundary conditions;
2. $du = (\Delta + \theta a(x)\partial/\partial x)udt + dW,\ x \in [0,1]$, with periodic boundary conditions;
3. $du = (\Delta + \theta a(x))udt + dW,\ x \in [0,1]$, with either zero or periodic boundary conditions.

For the first two examples, conditions (3.7) and (3.8) of Theorem 3.1 are fulfilled so that the maximum likelihood estimator (3.4) for θ is consistent and asymptotically normal, whereas for the third example, condition (3.9) holds so that the MLE (3.4) converges to (3.10).

In the following, C denotes a real number, independent of N and independent of the solutions of (2.1) and (3.1). The constant C may depend on other parameters, as indicated, and the value of C may be different at different places. For sequences $\{a_n\}_{n\geq 1}$ and $\{b_n\}_{n\geq 1}$ of positive real numbers the notation $a_n \asymp b_n$ means that

$$0 < \liminf_n a_n/b_n \leq \limsup_n a_n/b_n < \infty.$$

Example 1

This system is governed by the following equation with zero boundary conditions.

$$\begin{aligned} &du(t,x) = (\theta\Delta + a(x))u(t,x)dt + (1-\Delta)^{-1/2}\,dW(t,x),\ 0<t\leq T,\ x\in(0,1), \\ &u(0,x) = 0,\ \ u(t,0) = u(t,1) = 0. \end{aligned} \tag{4.1}$$

It is assumed that the parameter θ is positive, the function $a = a(x)$ belongs to $\mathbf{C}^\infty([0,1])$, and $a^{(n)}(0+) = a^{(n)}(1-) = 0$ for $n \geq 1$, where $a^{(n)}(x)$ is the n-th derivative of a. Let $H = L_2([0,1])$, with orthonormal basis $h_k(x) = \sqrt{2}\sin(\pi k x)$ for $k \geq 1$. The operator $\Lambda : L_2([0,1]) \to L_2([0,1])$ defined by $\Lambda h_k = \pi k\, h_k$ satisfies the assumptions from Section 2. Therefore Λ generates a Hilbert scale $\{H^s\}_{s\in I\!R}$. The operators corresponding to the evolution system (2.1) in Section 1 are as follows:

$A_1 = \Delta$ satisfies $A_1 h_k = -(k\pi)^2 h_k$.

$A_0 = a(x)$ satisfies $A_0 h_k = \sum_{l=1}^{\infty}(a_{l-k} - a_{l+k})h_l$.

$B = (1-\Delta)^{-1/2}$ satisfies $Bh_k = (1+(\pi k)^2)^{-1/2}h_k$, and B is a Hilbert-Schmidt operator in $H = L_2([0,1])$.

The assumptions on $a(x)$ imply that $|a_k| \leq C(r)/k^r$ for every $r > 0$ so that

$$\|A_0 f\|_s \leq C(s)\|f\|_s \tag{4.2}$$

for all $s \in I\!R$. Therefore conditions (A1)–(A7) are fulfilled for all $\theta > 0$ with $\alpha = 0,\ \gamma = 1$. Consequently, for every $\theta > 0$, there is a unique solution u of (4.1) satisfying

$$\mathbf{E}\sup_{0\leq t\leq T}\|u(t)\|_0^2 + \mathbf{E}\int_0^T \|u(t)\|_1^2 dt < \infty.$$

The objective now is to show that (3.7) and (3.8) hold for all $\theta > 0$. To simplify the presentation it is assumed that $\theta = 1$. The variation of parameters formula yields:

$$\begin{aligned} \left(u^N(t), h_k\right)_0 &= \int_0^t e^{-(\pi k)^2(t-s)}\left(A_0 u^N(s), h_k\right)_0 ds \\ &+ \left(1+(\pi k)^2\right)^{-1/2}\int_0^t e^{-(\pi k)^2(t-s)}\,dW_k(s). \end{aligned}$$

The first step is to show that Fisher's information (3.6) diverges as N tends to infinity:

$$I_N(\theta_0) = E\int_0^T \|\Pi^N B^{-1} A_1 u^N(t)\|_0^2\,dt \asymp N^3. \tag{4.3}$$

The following notation is used for the sake of brevity:

$$\xi_k(t) = \int_0^t (\pi k)^2 e^{-(\pi k)^2(t-s)} dW_k(s), \;\; X^N(t) = \sum_{k=1}^N \xi_k^2(t),$$
$$\eta_k^N(t) = (\pi k)^2 \sqrt{1+(\pi k)^2} \int_0^t e^{-(\pi k)^2(t-s)} \Big(A_0 u^N(s), h_k \Big)_0 \, ds.$$

With this notation,

$$\|\Pi^N B^{-1} A_1 u^N(t)\|_0^2 = X^N(t) + 2\sum_{k=1}^N \xi_k(t)\eta_k^N(t) + \sum_{k=1}^N (\eta_k^N(t))^2. \tag{4.4}$$

By the inequality $(1/2)x^2 - y^2 \le (x+y)^2 \le 2(x^2+y^2)$, the norm (4.4) can be bracketed as follows.

$$\frac{1}{2}X^N(t) - \sum_{k=1}^N (\eta_k^N(t))^2 \le \|\Pi^N B^{-1} A_1 u^N(t)\|_0^2 \le 2X^N(t) + 2\sum_{k=1}^N (\eta_k^N(t))^2. \tag{4.5}$$

Since for each t the random variables $\{\xi_k(t)\}_{k\ge 1}$ are independent and Gaussian, it follows that

$$\mathbf{E}\int_0^T X^N(t)dt \asymp \sum_{k=1}^N k^2 \asymp N^3. \tag{4.6}$$

The asymptotics in (4.3) holds if the expectation $\mathbf{E}\int_0^T (\eta_k^N(t))^2 dt$ is bounded. For this the following lemma is necessary.

Lemma 4.1. *If $a>0$ and $f(t)\ge 0$ then*

$$\int_0^T \Big(\int_0^t e^{-a(t-s)} f(s)ds \Big)^2 dt \le \frac{\int_0^T f^2(t)dt}{a^2}.$$

The above inequality can be easily verified by direct computation; for the sake of completeness, the main steps are given in the Appendix.

By Lemma 4.1,

$$\mathbf{E}\int_0^T (\eta_k^N(t))^2 dt \le \mathbf{E}\int_0^T (1+(\pi k)^2) \Big(A_0 u^N(t), h_k \Big)_0^2 \, dt,$$

so that by (3.2), (4.2), and (2.3),

$$\lim_{N\to\infty} \mathbf{E}\int_0^T \sum_{k=1}^N (\eta_k^N(t))^2 dt \le C\mathbf{E}\int_0^T \|u(t)\|_1^2 dt < \infty. \tag{4.7}$$

The last inequality together with (4.5) and (4.6) implies the asymptotics for Fisher's information in (4.3).

Now, (3.7) can be rewritten as follows.

$$\begin{aligned}\frac{\int_0^T \|\Pi^N B^{-1} A_1 u^N(t)\|_0^2 \, dt}{\mathbf{E}\int_0^T \|\Pi^N B^{-1} A_1 u^N(t)\|_0^2 \, dt} &= \frac{\int_0^T X^N(t)\, dt}{I_N} + 2\frac{\int_0^T \sum_{k=1}^N \xi_k(t)\eta_k^N(t)\, dt}{I_N} \\ &+ \frac{\int_0^T \sum_{k=1}^N (\eta_k^N(t))^2 dt}{I_N}.\end{aligned}$$

Both the second and third terms on the right-hand side converge to zero by (4.3) and (4.7). For the second term it is necessary to first apply the inequality

$$|2xy| \leq \delta x^2 + \delta^{-1} y^2 \tag{4.8}$$

which holds for every $\delta > 0$ and every real x, y. The precise arguments are given below. It will be shown next that the first term converges to 1.By direct computation,

$$var(X^N(t)) = \sum_{k=1}^{N} var(\xi_k^2(t)) \leq CN^5, \ \forall t \in [0,T].$$

Define

$$Y^N := \frac{\int_0^T (X^N(t) - \mathbf{E}X^N(t))dt}{\mathbf{E}\int_0^T X^N(t)dt}.$$

Then

$$\frac{\int_0^T X^N(t)dt}{\mathbf{E}\int_0^T X^N(t)dt} = 1 + Y^N$$

and

$$\mathbf{E}Y_N^2 \leq \frac{T\int_0^T (var(X^N(t))dt}{\left(\mathbf{E}\int_0^T X^N(t)dt\right)^2} \leq \frac{C}{N} \to 0 \quad \text{as } N \to \infty,$$

so that by Chebychev's inequality, $\mathrm{P} - \lim_{N\to\infty} Y_N = 0$ and

$$\mathrm{P} - \lim_{N\to\infty} \frac{\int_0^T X^N(t)dt}{\mathbf{E}\int_0^T X^N(t)dt} = 1. \tag{4.9}$$

The δ-inequality (4.8) can be written in the following form: $(1-\delta)x^2 + (1-\delta^{-1})y^2 \leq (x+y)^2 \leq (1+\delta)x^2 + (1+\delta^{-1})y^2$. This and (4.5) imply

$$\begin{aligned}
(1-\delta)\mathbf{E}\int_0^T X^N(t)dt &+ (1-\frac{1}{\delta})\mathbf{E}\int_0^T \sum_{k=1}^{N}(\eta_k^N(t))^2\, dt \\
&\leq \mathbf{E}\int_0^T \|\Pi^N B^{-1} A_1 u(t)\|_0^2 dt \\
&\leq (1+\delta)\mathbf{E}\int_0^T X^N(t)dt + (1+\frac{1}{\delta})\mathbf{E}\int_0^T \sum_{k=1}^{N}(\eta_k^N(t))^2 dt.
\end{aligned}$$

Since δ is arbitrary, (3.8) follows from (4.6)–(4.9).

Conclusion. For the model (4.1), the MLE (3.4) is consistent and asymptotically normal with the normalizing factor $\sqrt{I_N(\theta_0)} \asymp N^{3/2}$.

Example 2

Here the system's equation is, with periodic boundary conditions,

$$\begin{aligned}
du(t,x) &= (\Delta + \theta a(x)\frac{\partial}{\partial x})u(t,x)dt + dW(t,x), \ 0 < t \leq T, \ x \in [0,1], \\
u(0,x) &= 0, \ u(t,0) = u(t,1).
\end{aligned} \tag{4.10}$$

The parameter θ is real, and the function $a = a(x) \not\equiv 0$ is an infinitely differentiable and periodic with period 1. Let $H = L_2([0,1])$, with orthonormal basis $h_0(x) = 1$, $h_{2n}(x) = \sqrt{2}\cos(2\pi nx)$, $h_{2n-1}(x) = \sqrt{2}\sin(2\pi nx)$ for $n \geq 1$. For $k \geq 0$ define the numbers

$$\mu_k = 2\pi\Big([(k-1)/2] + 1\Big),$$

where $[x]$ is the largest integer less than or equal to x. The Hilbert scale $\{H^s\}_{s\in I\!R}$ is generated by the operator Λ with $\Lambda h_k = \mu_k h_k$, $k \geq 0$. The operator $A_0 = \Delta$ satisfies $A_0 h_k = -\mu_k^2 h_k$, $k \geq 0$, and $\partial/\partial x$ satisfies

$$\frac{\partial}{\partial x} h_k = \begin{cases} -\mu_k h_{k+1}, & \text{if } k = 2n-1 \\ \mu_k h_{k-1}, & \text{if } k = 2n. \end{cases}$$

Multiplication by $a(x)$ is defined according to the formulas for the Fourier coefficients of the product of two functions, and then the operator $A_1 = a(x)\partial/\partial x$ is defined in an obvious way. The assumptions on $a(x)$ imply that

$$\|A_1 f\|_s \leq C(s)\|f\|_{s+1}$$

for all $s \in I\!R$. The operator Λ^{-1} is Hilbert-Schmidt in $H = L_2([0,1])$ and $B = 1$, therefore conditions (A1)–(A7) are fulfilled for all $\theta \in I\!R$ with $\alpha = \gamma = 1$. As a result, for every $\theta \in R$, there is a unique solution u of (4.10) satisfying

$$\mathbf{E}\sup_{0\leq t\leq T}\|u(t)\|_{-1}^2 + \mathbf{E}\int_0^T \|u(t)\|_0^2 dt < \infty.$$

The objective now is to show that (3.7) and (3.8) hold for all $\theta \in I\!R$. Again, it is assumed that $\theta = 1$. For technical reasons, equation (4.10) is rewritten as follows:

$$du = ((\Delta - 1) + (a(x)\frac{\partial}{\partial x} + 1))udt + dW.$$

As in Example 1, the following notation is introduced:

$$\xi_k(t) = \int_0^t e^{-(1+\mu_k^2)(t-s)} dW_k(s), \;\; X^N(t) = \sum_{l,k=0}^N a_{lk}\xi_l(t)\xi_k(t),$$

where $a_{lk} = \sum_{n=0}^N (A_1 h_l, h_n)_0 (A_1 h_k, h_n)_0$, and

$$\eta_k^N(t) = \int_0^t e^{-(1+\mu_k^2)(t-s)} \left(\tilde{A}_1 u^N(s), h_k\right)_0 ds,$$

where $\tilde{A}_1 = A_1 + 1$. This notation, equation (3.1), and the variation of parameters formula yield:

$$\|\Pi^N B^{-1} A_1 u^N(t)\|_0^2 = X^N(t) + 2\sum_{l,k=0}^N a_{lk}\xi_l(t)\eta_k^N(t) + \sum_{l,k=1}^N a_{lk}\eta_l^N(t)\eta_k^N(t).$$

For every $t > 0$ the random variables $\{\xi_k(t)\}_{k\geq 0}$ are independent and Gaussian with zero mean and variance

$$R_k(t) = \frac{1}{2(1+\mu_k^2)}\Big(1 - e^{-(1+\mu_k^2)t}\Big).$$

It follows that

$$\mathbf{E}\int_0^T X^N(t)dt = \sum_{k=0}^N a_{kk}\int_0^T R_k(t)dt$$

and

$$var(X^N(t)) = 2\sum_{k,l=0}^N a_{kl}^2 R_k(t)R_l(t).$$

Direct computations similar to those in Example 1 yield

$$\begin{aligned}\sum_{k=0}^N a_{kk}\int_0^T R_k(t)dt &\asymp \sum_{k,n=0}^N \frac{1}{1+\mu_k^2}(A_1h_k,h_n)_0^2 \asymp \|a\|_0^2 N \asymp N,\\ var(X^N(t)) &\leq CN,\end{aligned}$$

and

$$\lim_{N\to\infty}\mathbf{E}\int_0^T \sum_{l,k=1}^N a_{lk}\eta_l^N(t)\eta_k^N(t)dt \leq C\mathbf{E}\int_0^T \|u(t)\|_0^2 dt < \infty.$$

Now, conditions (3.7) and (3.8) of Theorem 3.1 follow in the same way as in Example 1.

Conclusion. For the model (4.10), the MLE (3.4) is consistent and asymptotically normal with normalizing factor $\sqrt{I_N(\theta_0)} \asymp N^{1/2}$.

Example 3

Let $a = a(x) \not\equiv 0$ be either as in Example 1 or in Example 2 and consider the equation

$$du(t,x) = (\Delta u(t,x) + \theta a(x)u(t,x))dt + dW(t,x),\ 0 < t \leq T \tag{4.11}$$

with zero initial conditions and zero or periodic boundary conditions. Then it follows from the results of Examples 1 and 2 that the solution u of (4.11) exists and is unique in the corresponding Hilbert scale, and

$$\mathbf{E}\sup_{0\leq t\leq T}\|u(t)\|_{-1}^2 + \mathbf{E}\int_0^T \|u(t)\|_0^2\,dt < \infty.$$

In this case $A_1 = a(x)$ and $B = 1$, so that (3.9) holds with $\alpha = \gamma = 1$. This means that for the model (4.11) the MLE (3.4) is asymptotically biased, and the MLE converges to

$$\theta_0 + \frac{\int_0^T (a(\cdot)u(t), dW(t))_0}{\int_0^T \|a(\cdot)u(t)\|_0^2\,dt},$$

as N tends to infinity.

Appendix

Proof of Lemma 3.2 For an arbitrary $\varepsilon \in (0,1)$ define
$x_n^\varepsilon := \varepsilon\int_0^T \Big(f_n(s), dW(s)\Big)_0$ and $\langle x_n^\varepsilon\rangle := \varepsilon^2\int_0^T \|f_n(s)\|_0^2 ds$. Then

$$\mathrm{P}\left(\frac{\int_0^T \Big(f_n(s), dW(s)\Big)_0}{\int_0^T \|f_n(s)\|_0^2 ds} > \varepsilon\right) = \mathrm{P}\Big(\exp\left(x_n^\varepsilon - \langle x_n^\varepsilon\rangle/2\right) > \exp\left(\langle x_n^\varepsilon\rangle/2\right)\Big),$$

and to complete the proof of the lemma it remains to use the Chebuchev inequality and note that $\sup_n \mathbf{E} \exp(x_n^\varepsilon - \langle x_n^\varepsilon \rangle/2) \leq 1$ (Theorem 5.2 and Lemma 6.1 of [7]).

□

Proof of Lemma 4.1. Note that

$$\Big(\int_0^t e^{as} f(s)ds\Big)^2 = 2\int_0^t \int_0^s e^{as} e^{au} f(u)f(s)duds.$$

If $U := \int_0^T \Big(\int_0^t e^{-a(t-s)} f(s)ds\Big)^2 dt$, then direct computations yield

$$U = 2\int_0^T \int_0^t \int_0^s e^{-a(2t-s-u)} f(u)f(s)dudsdt \leq a^{-1}\Big(\int_0^T f^2(s)ds\Big)^{1/2} U^{1/2},$$

and the result follows.

□

References

[1] M. Huebner, R. Khasminskii, and B. L. Rozovskii, Two Examples of Parameter Estimation, in *Stochastic Processes*, ed. Cambanis *et al.* (Springer, New York, 1992).

[2] M. Huebner and B. Rozovskii, On Asymptotic Properties of Maximum Likelihood Estimators for Parabolic Stochastic PDE's, *Probability Theory and Related Fields*, **103**, 143–163 (1995).

[3] S. G. Krein, Ju. I. Petunin, and E. M. Semenov. *Interpolation of Linear Operators* (American Mathematical Society, Providence, RI, 1981).

[4] R. Mikulevicius and B.L. Rozovskii, Uniqueness and Absolute Continuity of Weak Solutions for Parabolic SPDE's, *Acta Applicandae Mathematicae*, **35**, 179–192 (1994).

[5] J. B. Walsh. An Introduction to Stochastic Partial Differential Equations, In *Ecole d'été de Probabilités de Saint-Flour, XIV*, ed. P. L. Hennequin (Lecture Notes in Mathematics, volume 1180, Springer, Berlin, 1984).

[6] B. L. Rozovskii, *Stochastic Evolution Systems* (Kluwer Academic Publishers, 1990).

[7] R. Sh. Liptser and A. N. Shiryayev, *Statistics of Random Processes* (Springer, New York, 1992).

[8] R. Sh. Liptser and A. N. Shiryayev, *Theory of Martingales* (Kluwer Academic Publishers, Boston, 1989).

[9] J. Jacod and A. N. Shiryayev, *Limit Theorems for Stochastic Processes* (Springer, Berlin, 1987).