Asymptotic analysis of the sieve estimator for a class of parabolic SPDEs

M. Huebner
Michigan State University

S. Lototsky
Massachusetts Institute of Technology

Published in *Scandinavian Journal of Statistics*, Vol. 27, pp. 353–370, 2000.

Abstract

In this paper we consider the problem of estimating a coefficient of a strongly elliptic partial differential operator in stochastic parabolic equations. The coefficient is a bounded function of time. We compute the maximum likelihood estimate of the function on an approximating space (sieve) using a finite number of the spatial Fourier coefficients of the solution and establish conditions that guarantee consistency and asymptotic normality of the resulting estimate as the number of the coefficients increases. The equation is assumed diagonalizable in the sense that all the operators have a common system of eigenfunction.

Key Words: Nonparametric estimation, maximum likelihood estimation, method of sieves, convergence rate, stochastic partial differential equations.

Running headline: Sieve estimation for SPDEs

AMS subject classification: 60H15, 62G05, 62G20, 62A10

1 Introduction

The theory of statistical inference for the problem of estimating parameters in diffusion processes is well-developed, see, for example, Ibragimov and Khasminskii (1982), Kutoyants (1984a), and Barndorff-Nielsen and Sørensen (1994).

Nonparametric estimation for *ordinary* stochastic differential equations have been studied by several authors. Kutoyants (1984b) derived asymptotic properties of kernel-type estimators of the drift term in a stochastic differential equation. In a nonstationary linear diffusion model Nguyen and Pham (1982) applied Grenander's method of sieves to the problem of estimating the drift coefficient that is a function of time. They approximated the unknown function by a finite linear combination of a given system of functions and proved consistency and asymptotic normality of the estimate. In this approach, when the unknown function is approximated by a finite linear combination of known functions, it is also necessary to determine the "optimal" number of terms in the approximation under certain accuracy and cost criteria. A large literature exists about such model selection criteria in various contexts. For example, Polyak and Tsybakov (1990) and Verulava and Polyak (1988) studied this question for regressions via Mallow's C_p criterion, while Birgé and Massart (1997) consider penalized projection estimators for various families of sieves and penalties.

However, little has been done concerning nonparametric estimation for infinite dimensional systems. One general problem is estimation of a function that is a coefficient of a partial differential operator in a parabolic stochastic partial differential equation. Stochastic partial differential equations (SPDEs) often represent physical models in areas such as oceanography, physical chemistry, economics, and geostatistics. Because of these potential applications and the interesting mathematical questions arising from them there has been growing interest in the estimation of model parameters for SPDEs. Ibragimov and Khasminskii (1997) studied asymptotic properties of estimators of general functions in the small noise asymptotics when the probability measures generated by the processes corresponding to different functions are equivalent. Other inverse problems for SPDEs in the small noise asymptotics such as recovery of initial and boundary conditions are studied in Golubev and Khasminskii (1997).

If the unknown function is the coefficient of the "leading" differential operator, which is the case in many applications, then the probability measures (on the appropriate infinite-dimensional Hilbert space) generated by the processes corresponding to different functions are singular and different approaches to constructing the estimate can be used. In particular, it is possible to estimate the function even when the time interval and the noise intensity are fixed. To construct a computable estimate, one has to work with finite dimensional projections of the observation process, for example, the first N (spatial) Fourier coefficients. The dimension of the projection is then used to describe the asymptotic properties of the estimate. In parametric models, when the coefficient is just a real number, this approach was used by Huebner and Rozovskii (1995), who constructed the maximum likelihood estimate on the basis of the first N Fourier coefficients of the process, and established the conditions for consistency and asymptotic normality of the estimate in the limit $N \to \infty$.

The objective of this paper is to combine the methods used in Huebner and Rozovskii (1995) and Nguyen and Pham (1982) and construct an estimate of a coefficient that is a function of time in a model described by a stochastic parabolic equation. Suppose the process $u(t,x)$ for $t \in [0,T]$ and $x \in G \subset \mathbb{R}^d$ is governed by the following equation:

$$\begin{array}{rcl} du(t,x) & = & (A_0 + \theta_0(t)A_1)\, u(t,x)\, dt + dW(t,x), \quad t \in (0,T],\ x \in G \\ u(0,x) & = & u_0(x) \end{array}$$

with zero boundary conditions, where $W(t,x)$ a cylindrical Brownian motion in $L_2([0,T]\times G)$ and $A_0 + \theta_0(t)A_1$ is a strongly elliptic differential operator with the unknown coefficient $\theta_0(t)$. Suppose we observe finitely many Fourier coefficients $u_1(t), \ldots, u_N(t)$ for all $t \in [0,T]$. Let Θ be the set of admissible functions θ_0. We are interested in the asymptotic properties of the sieve maximum likelihood estimate $\hat{\theta}^N$ obtained by maximizing the likelihood function based on the N Fourier coefficient. The maximization is carried out over a sieve Θ_N, that is, a finite dimensional subspace of Θ. The family of spaces $\{\Theta_N,\ N \geq 1\}$ is chosen so that the approximation error decreases to zero as the the number N of observations increases. This method of constructing an estimate is called the *method of sieves* (see Grenander (1981)).

In this paper we use linear nested sieves (cf. Birgé and Massart (1997)). We assume that every function $\theta \in \Theta$ can be represented as an infinite linear combination of known functions $\{h_j, j \geq 1\}$:

$$\theta(t) = \sum_{j=1}^{\infty} \theta_j h_j(t)$$

and the functions $\{h_j\}$ are orthonormal on $[0,T]$. If we choose the sieve Θ_N to be the span of

$h_1(t), \ldots, h_{d_N}(t)$, then the *sieve maximum likelihood estimate* will be of the form

$$\hat{\theta}^N = \sum_{j=1}^{d_N} \hat{\theta}_j h_j(t).$$

This sieve maximum likelihood estimate is therefore a particular case of projection estimates first considered by Chentsov (1982). We give an explicit formula for the sieve maximum likelihood estimate using the first N Fourier coefficients of the process and establish conditions that guarantee consistency and asymptotic normality of the resulting estimate in the limit $N \to \infty$. These conditions relate the dimension d_N of the approximating spaces Θ_N to the number N of observed Fourier coefficients and the orders of the operators A_0 and A_1 in the equation. Only the asymptotical properties of the estimate are studied, and the finite-sample issues are not discussed.

The paper is organized as follows. In Section 2 we introduce the model and the basic notation. The main results on the asymptotic properties of the sieve maximum likelihood estimate, including the convergence rates, are stated in Section 3, and the results are illustrated on several examples in Section 4. The proofs of the main results are in Section 5.

2 The Model

In this section we introduce the basic notations and assumptions about the model. It is important to note that in estimation problems where the observations are generated by finite dimensional processes it is assumed that either the noise intensity decreases ($\varepsilon \to 0$) or the time interval gets larger. For our model both the noise intensity and the time interval stay fixed. The Notation $x_N \sim y_N$ used in the paper means that $\lim_{N\to\infty} x_N/y_N = c$ where $c \neq 0, \infty$.

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{0\le t\le T}, P)$ be a stochastic basis with the usual assumptions (see Jacod and Shiryayev, 1987) and G a smooth bounded domain in $I\!R^d$ or a smooth d-dimensional compact manifold (without boundary). We denote by A_0 and A_1 partial differential operators on G with complex-valued coefficients. If G is a domain, then the operators are supplemented with zero boundary conditions. We assume that

$$A_i u(x) = -\sum_{|\alpha|\le m_i} a_i^\alpha(x) u^{(\alpha)}(x),\ a_i^\alpha \in C_b^\infty(G),\ i = 1, 2, \tag{2.1}$$

with known a_i^α and where $\alpha = (\alpha_1, \dots, \alpha_d)$, $\alpha_i = 0, 1, \dots,$ $|\alpha| = \sum_{i=1}^d \alpha_i$,

$$u^{(\alpha)}(x) = \frac{\partial^{|\alpha|} u(x)}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}.$$

The observation process is governed by the following equation:

$$\begin{array}{rcl} du(t,x) & = & (A_0 + \theta_0(t) A_1)\, u(t,x)\, dt + dW(t,x), \quad t \in (0,T], x \in G, \\ u(0,x) & = & u_0(x) \end{array} \tag{2.2}$$

where $\theta = \theta(t)$ is a bounded measurable function on $[0,T]$ and $W = W(t,x)$ is a cylindrical Brownian motion, that is, a distribution-valued process so that for every $\varphi \in C_0^\infty(G)$ with $\|\varphi\|_{L_2(G)} = 1$, $(W,\varphi)(t)$ is a standard Wiener process, and for all $\varphi_1, \varphi_2 \in C_0^\infty(G)$, $E(W,\varphi_1)(t)(W,\varphi_2)(s) = \min(t,s) \cdot (\varphi_1, \varphi_2)_{L_2(G)}$ (see Walsh (1984) for more details).

A predictable process u with values in the set of distributions on $C_0^\infty(G)$ is called a solution of (2.2) if for every $\varphi \in C_0^\infty(G)$ the equality

$$(u,\varphi)(t) = (u_0,\varphi) + \int_0^t (A_0^*\varphi, u)(s) ds + \int_0^t \theta_0(s)(A_1^*\varphi, u)(s) ds + (W,\varphi)(t)$$

holds with probability one for all $t \in [0,T]$ at once, where A_i^* is the formal adjoint of A_i, that is, an operator so that

$$(A_i\phi_1, \phi_2)_{L_2(G)} = (A_i^*\phi_2, \phi_1)_{L_2(G)} \text{ for all } \phi_1, \phi_2 \in C_0^\infty(G).$$

The following assumptions will be in force throughout the paper:

(H1) There is a complete orthonormal system $\{\varphi_k\}_{k\geq 1}$ in $L_2(G)$ so that

$$A_0\varphi_k = \kappa_k\varphi_k, \quad A_1\varphi_k = \nu_k\varphi_k,$$

(H2) The eigenvalues ν_k and κ_k satisfy $|\nu_k| \sim k^{m_1/d}$ and $|\kappa_k| \sim k^{m_0/d}$ and, uniformly in $t \in [0,T]$, $\mu_k(t) := -(\kappa_k + \theta(t)\nu_k) \sim k^{2m/d}$, $2m = \max\{m_0, m_1\}$, which means that

$$\alpha_k \leq -(\kappa_k + \theta(t)\nu_k) \leq \beta_k$$

for all $0 \leq t \leq T$ and some $\alpha_k \sim \beta_k \sim k^{2m/d}$. Recall that m_0 and m_1 are the orders of the operators A_0 and A_1 (see (2.1)).

Assumptions (H1) and (H2) hold in many physical models (see, for example, Piterbarg and Rozovskii, 1997). A typical situation is when the operators A_0 and A_1 commute and either A_0 or A_1 is uniformly elliptic and formally self-adjoint. For the sake of completeness we included a more precise statement in the Appendix. More details can be found in Safarov and Vassiliev (1997).

To state the result about existence and uniqueness of the solution of (2.2) we need some additional constructions. For $f \in C_0^\infty(G)$ and $s \in I\!\!R$ define

$$\|f\|_s^2 = \sum_{k\geq 1} k^{2s/d} |(f, \varphi_k)_{L_2(G)}|^2,$$

and then define the space $H^s(G)$ as the completion of $C_0^\infty(G)$ with respect to the norm $\|\cdot\|_s$. There is a one-to-one correspondence between the elements $v \in H^s(G)$ and sequences $\{v_k\}_{k\geq 1}$ so that

$$\|v\|_s^2 = \sum_{k\geq 1} k^{2s/d} |v_k|^2 < \infty;$$

we call $\{v_k\}$ the (spatial) *Fourier coefficients* of v. The Fourier coefficients of the cylindrical Brownian motion W are $\{w_k\}_{k\geq 1}$, independent standard Brownian motions, and therefore $W \in L_2(\Omega \times (0,T); H^{-s}(G))$ for every $s > d/2$.

Proposition 2.1 *Under assumptions (H1) and (H2), if $u_0 \in L_2(\Omega; H^{-s}(G))$ for some $s > d/2$, then there is a unique solution of (2.2) that belongs to the space $L_2(\Omega \times (0,T); H^{m-s}(G)) \cap L_2(\Omega; C((0,T), H^{-s}(G)))$; the solution satisfies*

$$E \sup_{0\leq t\leq T} \|u\|_{-s}^2(t) + E\int_0^T \|u\|_{m-s}^2(t)dt \leq K(d,m,s,T,\theta_0)\left(E\|u_0\|_{-s}^2 + T\right).$$

Proof. This follows from Theorem 3.1.4 in Huebner and Rozovskii (1995).

□

Remark 2.2 *1. In equation (2.2) we can have some correlation operator B for W as long as the eigenfunctions of B are also φ_k. In that case we replace u_0 by $B^{-1}u_0$.*

2. In principle, we can consider more general models, for example, equations with other boundary conditions or other types of operators. All we need is that the operators have the properties (H1) and (H2).

3 Main Result

In this section we construct the sieve maximum likelihood estimate of θ_0 and establish conditions under which the estimate is consistent and asymptotically normal. The set of admissible functions θ_0 is a subspace of $L_2(0,T)$ and will be denoted by Θ. The set Θ is the collection of functions for which assumption (H2) holds. If A_1 is not the leading operator, then Θ is just the set of bounded measurable function on $(0,T)$. If A_1 is the leading operator, then the functions in Θ must also be positive and bounded away from zero.

We will estimate $\theta_0(t)$ on a subspace Θ_N (sieve) of Θ. The estimate is denoted by $\hat{\theta}^N(t)$.

The observations are of the form $u_1(t),\dots,u_N(t)$, where the $u_k(t)$ are the Fourier coefficients of the process $u(t,x)$:

$$\begin{aligned} du_k(t) &= -\mu_k(t)u_k(t)\,dt + dw_k(t) \\ u_k(0) &= u_{0k}; \end{aligned} \tag{3.1}$$

recall that $\mu_k(t) = -(\kappa_k + \theta_0(t)\nu_k)$.

Let $h_1, h_2, \dots$ be an orthonormal system in Θ. Let $\{\Theta_N, N \geq 1\}$ be an increasing sequence of subspaces of Θ so that Θ_N is spanned by $h_1,\dots,h_{d_N}$. Notice that the dimension d_N of the subspace Θ_N depends on the number N of the Fourier coefficients observed.

Since $\{u_k(t),\ k = 1,\dots,K\}$ is a finite dimensional diffusion process with independent components, the corresponding likelihood ratio can be computed explicitly (Liptser and Shiryayev [1977, Theorem 7.14]) and is given by

$$\begin{aligned} Z_N &= \exp\left\{\int_0^T (\theta(t)-\theta_0(t))\left(A_1u^N(t), du^N(t) - A_0u^N(t)dt\right)_{L_2(G)}\right. \\ &\qquad \left. -\frac{1}{2}\int_0^T \left(\theta^2(t)-\theta_0^2(t)\right)\left\|A_1u^N(t)\right\|^2_{L_2(G)}\,dt\right\}. \end{aligned}$$

We obtain the sieve maximum likelihood estimate by maximizing likelihood ratio on the subspace Θ_N. Then the estimate is $\hat{\theta}^N = \sum_{j=1}^{d_N}\hat{\theta}_j h_j(t)$, and the vector $\hat{\theta}^N = (\hat{\theta}_1^N,\dots,\hat{\theta}_{d_N}^N)$ is the solution of a system of linear equations

$$J(N)\hat{\theta}^N = a^N, \tag{3.2}$$

where

$$a^N = \left(\int_0^T h_j(t)\left(A_1u^N(t), du^N(t) - A_0u^N(t)\,dt\right)_{L_2(G)}\right)_{j=1,\dots,d_N}$$

and

$$J(N) = \left(\int_0^T h_i(t) h_j(t) \|A_1 u^N(t)\|^2_{L_2(G)}\, dt \right)_{i,j=1,\ldots,d_N}.$$

The matrix $J(N)$ is invertible almost surely (proof in Appendix), and therefore $\hat{\theta}^N$ can be written as

$$\hat{\theta}^N = (J(N))^{-1} a^N; \tag{3.3}$$

however, for large d_N it seems more reasonable to use other methods of solving the system (3.2), for example, Gaussian elimination. Note that, due to assumption (H2), the matrix $J(N)$ and the vector a^N can be written explicitly in terms of Fourier coefficients $\{u_k(t), k = 1, \ldots, N\}$.

To describe the asymptotic properties of the estimate we introduce the following notations.

(N1). Define $q = 2(m_1 - m)/d$. It is known from Huebner and Rozovskii (1995) that, even in the cases of a scalar parameter ($\theta_0 = const$), a consistent estimate is possible if and only if $q \geq -1$.

(N2). Define $F_{q,N}$ by

$$F_{q,N} = \begin{cases} \frac{N^{q+1}}{q+1}, & q > -1 \\ \log N, & q = -1. \end{cases}$$

Note that $\lim_{N\to\infty} F_{q,N} / \sum_{k=1}^N k^q = 1$ as long as $q \geq -1$. In the parametric case, the quantity $F_{q,N}$ determines the rate of convergence of the maximum likelihood estimate (Huebner and Rozovskii (1995)).

(N3). Assumptions about the model imply that there a limit $\lim_{k\to\infty} k^q \mu_k(t) |\nu_k|^{-2}$; the limit will be denoted by $\tilde{\theta}(t)$. There are constants c_0, c_1 so that

$$\tilde{\theta}(t) = \begin{cases} c_1 \theta_0(t), & m_0 < m_1 = 2m; \\ c_0, & m_1 < m_0 = 2m; \\ c_0 + c_1 \theta_0(t), & m_0 = m_1 = 2m. \end{cases}$$

The exact values of c_0 and c_1 can be computed using Proposition A.1 in Appendix. Assumption (H2) implies that $\tilde{\theta}(t)$ is strictly positive on $[0, T]$.

(N4) Let $\{Q_{1,N},\ N \geq 1\}$ and $\{Q_{2,N},\ N \geq 1\}$ be any sequences of real numbers so that

$$Q_{1,N} \sim \frac{\sum_{k=1}^N k^{4(m_1-m)/d}}{(\psi_N)^2}, \qquad Q_{2,N} \sim \frac{\sum_{k=1}^N k^{(2m_1-4m)/d}}{\psi_N}$$

where $\psi_N = \int_0^T E\|A_1 u^N(t)\|^2_{L_2(G)}\, dt$. The asymptotics of the above expressions depends on q; we summarize the results in Tables 1 and 2.

Table 1: Sequence $Q_{1,N}$

$q = 2(m_1 - m)/d$	$Q_{1,N}$	$1/F_{q,N}$
$q = -1$	$1/(\log N)^2$	$1/\log N$
$-1 < q < -1/2$	$1/N^{2(q+1)}$	$(q+1)/N^{q+1}$
$q = -1/2$	$\log N/N$	$1/2N^{1/2}$
$q > -1/2$	$1/N$	$(q+1)/N^{q+1}$

Table 2: Sequence $Q_{2,N}$

$q = 2(m_1 - m)/d$	$Q_{2,N}$	$1/F_{q,N}$
$q = -1$	$1/\log N$	$1/\log N$
$-1 < q < -1 + 2m/d$	$1/N^{q+1}$	$(q+1)/N^{q+1}$
$q = -1 + 2m/d$	$\log N/N^{2m/d}$	$d/2mN^{2m/d}$
$q > -1 + 2m/d$	$1/N^{2m/d}$	$(q+1)/N^{q+1}$

The main results of this paper, consistency and asymptotic normality of the sieve maximum likelihood estimate, are stated in the following two theorems.

Theorem 3.1 (Consistency). *In addition to (H1) and (H2) assume that*

(A1) $q \geq -1$ *and* $\lim_{N\to\infty} d_N = \infty$,

(A2) $\lim_{N\to\infty} d_N Q_{1,N} = 0$,

(A3) $\sup_{0\leq t\leq T} \sum_{j=1}^{d_N} |h_j(t)|^2 \leq D_N$ *and* $\lim_{N\to\infty} D_N Q_{2,N} = 0$,

(A4) u_0 *is deterministic and belongs to* $H^{m-d/2}(G)$.

Then the estimate $\hat{\theta}^N(t)$ *is consistent in probability:*

$$P - \lim_{N\to\infty} \int_0^T |\hat{\theta}^N(t) - \theta_0(t)|^2 dt = 0.$$

Given the nature of the problem, the conditions of Theorem 3.1 are a natural combination of conditions from Birgé and Massart (1997), Huebner and Rozovskii (1995), and Nguyen and Pham (1982). In particular, assumption (A1) is necessary to get a consistent estimate; assumptions (A2) and (A3) are technical and are similar to what is often assumed in the literature, see for example Birgé and Massart (1997) and Nguyen and Pham (1982); assumption (A4) is also technical and is used to reduce the general case to the case $u_0 = 0$.

Several widely used bases satisfy the first part of assumption (A3):

1. **Cosine Basis.** If

$$h_1(t) = \frac{1}{\sqrt{T}}, \qquad h_j(t) = \sqrt{\frac{2}{T}} \cos\left(\frac{\pi(j-1)}{T} t\right), \quad j > 1,$$

   then $\sup_{0 \le t \le T} \sum_{j=1}^{d_N} |h_j(t)|^2 \le C d_N$, $D_N \sim d_N$, and the second part of assumption (A3) becomes $d_N Q_{2,N} \to 0$.

2. **Legendre Polynomial Basis.** For simplicity let $(0, T)$ be $(0, 1)$. If $h_j(t)$ is the normalized Legendre polynomial

$$h_{j+1}(t) = \sqrt{2j+1}\, p_j(t), \text{ where } p_j(t) = \frac{1}{2^j j!} \frac{d^j (t^2-1)^j}{dt^j}, \quad j = 0, 1, \ldots,$$

   then, since $|p_j(t)| \le 1$ (see, for example, Devore and Lorentz (1993)), we have

$$\sup_{0 \le t \le T} \sum_{j=1}^{d_N} |h_j(t)|^2 \le \sum_{j=1}^{d_N} (2j+1) \le C d_N^2, \quad D_N \sim d_N^2,$$

   and in this case the second part of assumption (A3) becomes $d_N^2 Q_{2,N} \to 0$.

3. **Wavelet Basis.** It is shown in Birgé and Massart [1997, Section 2.2.2] that for the basis obtained by translation and dilation of a compactly supported function we have $D_N \sim d_N$ so that the second part of assumption (A3) becomes $d_N Q_{2,N} \to 0$.

Theorem 3.2 (Asymptotic Normality). *If, in addition to assumptions of Theorem 3.1,*

(A5) $\lim_{N\to\infty} d_N^2 Q_{1,N} = 0$ *and*

(A6) $\lim_{N\to\infty} F_{q,N} \sum_{i=d_N+1}^{\infty} |\theta_{0i}|^2 = 0,$

then the estimate $\hat{\theta}^N(t)$ is asymptotically normal, that is, for every deterministic $g \in L_2(0,T)$,

$$\lim_{N\to\infty} \sqrt{F_{q,N}} \int_0^T g(t)(\hat{\theta}^N(t) - \theta_0(t))dt = \mathcal{N}\left(0, 2\int_0^T |g(t)|^2 \tilde{\theta}(t)dt\right) \tag{3.4}$$

in distribution, where $\mathcal{N}(0,\sigma^2)$ is a normal random variable with mean zero and variance σ^2.

4 Examples

In this section we consider the following function space:

$$\Theta^\gamma(0,T) = \left\{\theta = \theta(t) : |\theta_j|^2 \le \frac{L}{j^{\gamma+1}}\right\},$$

where $\theta_j = \int_0^T \theta(t)h_j(t)dt$, $\gamma > 0$, and $L = L(\theta)$ is a constant. By definition, if $\theta \in \Theta^\gamma(0,T)$, then

$$\sum_{j=d_N+1}^{\infty} |\theta_{0j}|^2 \le C d_N^{-\gamma}.$$

The space $\Theta^\gamma(0,T)$ in general depends on the basis $\{h_j\}$. To give some examples, assume that $p = p_0 + p' \ge 1$ with $p_0 = 1, 2, \ldots$ and $p' \in [0,1]$. If the cosine basis is used and the even periodic extension of $\theta(t)$ belongs to $C^{p_0,p'}(I\!R)$, that is, the extension of θ is p_0 times continuously differentiable and the p_0-th derivative is Hölder continuous of order p', then $\theta \in \Theta^{2p-1}(0,T)$. Note that if θ is continuously differentiable on $[0,T]$, then $\theta \in \Theta^\gamma(0,T)$ with $\gamma \ge 1$.

Similarly, if the Legendre polynomials are used and $\theta \in C^{p_0,p'}(0,T)$, then $\theta \in \Theta^p(0,T)$. These and other related results can be found in Devore and Lorentz (1993).

Example 1. Let $G = (0,1)$ and $\Delta = \partial^2/\partial x^2$. Consider the following equation:

$$\begin{aligned} du(t,x) &= \theta_0(t)\Delta u(t,x)\,dt + (I-\Delta)^{-1/2}dW(t,x) \\ u(0,x) &= 0 \\ u(t,0) &= u(t,1) = 0, \quad t \in [0,T]. \end{aligned}$$

In this example, $d = 1$, $m = 1$ and $m_1 = 2$, so that $q = 2$, $F_{q,N} = N^3/3$, and $\tilde{\theta}(t) = \pi^2\theta_0(t)$. The spatial Fourier basis is $e_k(x) = \sqrt{2}\sin(k\pi x)$ and the spatial Fourier coefficients of the solution are

$$u_k(t) = \exp\{-k^2\pi^2\int_0^t \theta(r)\,dr\}u_{0k} + (1+k^2\pi^2)^{-1/2}\int_0^t \exp\{-k^2\pi^2\int_s^t \theta(r)\,dr\}\,dw_k(s)$$

We assume that $0 < M_1 \le \theta(t) \le M_2$ and $\theta_0 \in \Theta^\gamma(0,T)$. Then $M_1\pi^2k^2 \le \mu_k(t) \le M_2\pi^2k^2$ and assumptions (H1) and (H2) hold. Also, if the cosine basis is used, then, according to Theorem 3.2, to have a consistent and asymptotically normal estimate we need

$$\frac{d_N^2}{N} \to 0, \quad \frac{N^3}{d_N^\gamma} \to 0.$$

Therefore we can take $d_N \sim N^r$, where

$$\frac{3}{\gamma} < r < 1.$$

To have a consistent and asymptotically normal estimate using Legendre polynomials, we take $d_N \sim N^r$, where

$$\frac{3}{\gamma} < r < \frac{1}{2}.$$

Example 2. Let $(x,y) \in (-1,1)^2$ and $\Delta = \partial^2/\partial x^2 + \partial^2/\partial y^2$. Consider the following equation

$$\begin{aligned} du(t,x,y) &= (\Delta u(t,x,y) + \theta_0(t)u(t,x,y))\ dt + dW(t,x) \\ u(0,x,y) &= u_0(x,y) \in L_2(G) \end{aligned}$$

with periodic boundary conditions, so that G is a torus. In this case $d = 2$, $m = 1$, and $m_1 = 0$, so that $q = -1$, $F_{q,N} = \log N$, and, using Safarov and Vassiliev [1997, Example 1.2.3], $\tilde{\theta}(t) = \pi$.

We assume that $|\theta_0(t)| \le M$ and $\theta_0(t) \in \Theta^\gamma(0,T)$. Assumptions (H1) and (H2) are obviously fulfilled. Also, if the cosine basis is used, then, according to Theorem 3.2, to have a consistent and asymptotically normal estimate we need

$$\frac{d_N}{\log N} \to 0, \quad \frac{\log N}{d_N^\gamma} \to 0.$$

Therefore, we can take $d_N \sim (\log N)^s$, where

$$\frac{1}{\gamma} < s < 1.$$

To have a consistent and asymptotically normal estimate using the Legendre polynomials we take $d_N \sim (\log N)^s$, where

$$\frac{1}{\gamma} < s < \frac{1}{2}.$$

5 Proof of Theorems 3.1 and 3.2

First, we introduce more notations:

(N5).

$$X^N(t) = \|A_1 u^N(t)\|^2_{L_2(G)}, \quad \psi_N = \int_0^T EX^N(t)dt$$

$$\phi^N(t) = EX^N(t)/\psi_N; \quad Y^N(t) = (X^N(t) - EX^N(t))/\psi_N$$

It follows from Huebner and Rozovskii [1995, Lemma 2.1] and assumption (H1) that $\psi_N \sim F_{q,N}$.

(N6). If $\zeta = (\zeta_1, \ldots, \zeta_{d_N})$ and $\xi = (\xi_1, \ldots, \xi_{d_N})$ are vectors in $\mathbb{R}^{d_N}$ (random or deterministic), then

$$\zeta(t) = \sum_{i=1}^{d_N} \zeta_i h_i(t), \quad \|\zeta\|^2 = \sum_{i=1}^{d_N} |\zeta_i|^2 = \int_0^T |\zeta(t)|^2 dt$$

$$(\zeta, \xi) = \sum_{i=1}^{d_N} \zeta_i \xi_i = \int_0^T \zeta(t)\xi(t)dt.$$

(N7). For matrices, $\|\cdot\|$ denotes the 2-norm. For symmetric positive definite matrices it is the largest eigenvalue, and the upper bound on the square of the norm for every matrix is the sum of squares of all the entries of the matrix.

(N8). The letter C denotes a constant whose value can depend only on d, m, m_1, T, and θ_0; the value of C can be different in different places.

To simplify the presentation, we assume that the lower bound on $\mu_k(t)$ (cf. assumption (H2)) is always positive. Since the objective is the asymptotical behavior of the estimate, this assumption does not result in any loss of generality.

To investigate the asymptotic properties of the estimate (3.2), let θ_0^N be the orthogonal projection of θ_0 onto Θ_N:

$$\theta_0^N(t) = \sum_{j=1}^{d_N} \theta_{0j} h_j(t).$$

The j-th component of the vector a^N is

$$\begin{aligned} a_j^N &= \int_0^T h_j(t) \left(A_1 u^N(t), du^N(t) - A_0 u^N(t)\, dt \right)_{L_2(G)} \\ &= \int_0^T h_j(t) \left(A_1 u^N(t), dW^N(t) \right)_{L_2(G)} + \int_0^T h_j(t)\theta_0(t) X^N(t) dt. \end{aligned}$$

We rewrite the second term of the right-hand side above:

$$\sum_{i=1}^{d_N} \theta_{0i} \int_0^T h_i(t)h_j(t)X^N(t)dt + \sum_{i=d_N+1}^{\infty} \theta_{0i} \int_0^T h_i(t)h_j(t)X^N(t)dt.$$

Hence equation (3.2) can be written as

$$J(N)\left(\hat{\theta}^N - \theta_0^N\right) = b^N + c^N, \tag{5.1}$$

where

$$b^N = \left(\int_0^T h_j(t)\left(A_1 u^N(t), dW^N(t)\right)_{L_2(G)}\right)_{j=1,\ldots,d_N}$$

and

$$c^N = \left(\int_0^T h_j(t)\left(\theta_0(t) - \theta_0^N(t)\right) X^N(t)dt\right)_{j=1,\ldots,d_N}.$$

With the notation $\tilde{J}(N) = J(N)/\psi_N$ we can write (5.1) as follows:

$$\hat{\theta}^N - \theta_0^N = (\tilde{J}(N))^{-1}\left(b^N/\psi_N + c^N/\psi_N\right). \tag{5.2}$$

Note that

$$\|\hat{\theta}^N - \theta_0\|^2 = \|\hat{\theta}^N - \theta_0^N\|^2 + \sum_{i=d_N+1}^{\infty} |\theta_{0i}|^2, \tag{5.3}$$

where the second term tends to zero as $N \to \infty$ as long as $d_N \to \infty$.

Next, we look at the contribution of the initial condition.

Since $du_k = -\mu_k(t)u_k dt + dw_k$, it follows that

$$u_k(t) = u_{0k} \exp\left(-\int_0^t \mu_k(s)ds\right) + \int_0^t \exp\left(-\int_s^t \mu_k(r)dr\right) dw_k(s),$$

and then $X^N(t) = X^{1,N}(t) + X^{2,N}(t) + X^{3,N}(t)$, where

$$X^{1,N}(t) = \sum_{k=1}^N |u_{0k}|^2 \nu_k^2 \exp\left(-2\int_0^t \mu_k(s)ds\right) \quad (\mathrm{non-random}),$$

$$X^{2,N}(t) = \sum_{k=1}^N \nu_k^2 \left(\int_0^t \exp\left(-\int_s^t \mu_k(r)dr\right) dw_k(s)\right)^2,$$

$$X^{3,N}(t) = 2\sum_{k=1}^N \nu_k^2 u_{0k} \int_0^t \exp\left(-\int_s^t \mu_k(r)dr\right) dw_k(s).$$

As a result,

$$EX^N(t) = X^{1,N}(t) + EX^{2,N}(t), \qquad var(X^N(t)) = var(X^{2,N}(t)) + E|X^{3,N}(t)|^2.$$

Since by assumption (H2) $\mu_k(t) \geq Ck^{2m/d}$ for all sufficiently large k, we have

$$X^{1,N}(t) \leq C \sum_{k=1}^{N} |u_{0k}|^2 k^{2m/d-1} k^{q+1},$$

and so

$$\lim_{N\to\infty} \sup_{0\leq t\leq T} \frac{X^{1,N}(t)}{\psi_N} = 0, \tag{5.4}$$

either by the Kronecker lemma (if $q > 1$) or because $X^{1,N}(t)$ is bounded for all N uniformly in t (if $q = -1$); note that $\sum_{k\geq 1} |u_{0k}|^2 k^{2m/d-1} = \|u_0\|^2_{m-d/2}$.

Next,

$$\sup_{0\leq t\leq T} \frac{var(X^{2,N}(t))}{(\psi_N)^2} \leq CQ_{1,N}$$

(direct computations or from the proof of Lemma 2.2 in Huebner and Rozovskii, 1995), and, using assumption (H2) once again,

$$E|X^{3,N}(t)|^2 \leq \sum_{k=1}^{N} \frac{\nu_k^4 |u_{0k}|^2}{\alpha_k} \leq C \sum_{k=1}^{N} |u_{0k}|^2 k^{2m/d-1} k^{2q+1},$$

so that $E|X^{3,N}(t)|^2/(\sum_{k=1}^N k^{2q}) \to 0$, either by the Kronecker lemma (if $q > -1/2$) or because $E|X^{3,N}(t)|^2$ is bounded for all N uniformly in t (if $q \leq -1/2$). As a result,

$$\sup_{0\leq t\leq T} \frac{var(X^N(t))}{(\psi_N)^2} \leq CQ_{1,N}. \tag{5.5}$$

5.1 Consistency

In view of (5.3), it remains to show that $P - \lim_{N\to\infty} \|\hat{\theta}^N - \theta_0^N\| = 0$. Due to (5.2), it is sufficient to show that

(C1) $\|\tilde{J}^{-1}(N)\| \leq L_N$, where L_N converges to a constant in probability;

(C2) $\|b^N/\psi_N\| \to 0$ in probability as $N \to \infty$;

(C3) $\|c^N/\psi_N\| \to 0$ in probability as $N \to \infty$.

1. The matrix.

Write

$$\tilde{J}_{ij}(N) = \tilde{J}^r_{ij}(N) + \tilde{J}^d_{ij}(N)$$

where $\tilde{J}^r_{ij}(N) = \int_0^T h_i(t)h_j(t)Y^N(t)dt$ (random part) and $\tilde{J}^d_{ij}(N) = \int_0^T h_i(t)h_j(t)\phi^N(t)dt$ (deterministic part).

The norm of the random part tends to zero in probability. Indeed,

$$\sum_{i,j=1}^{d_N} E|\tilde{J}^r_{ij}(N)|^2 \leq E\sum_{i=1}^{d_N}\sum_{j\geq 1}|\int_0^T (h_i(t)Y^N(t))h_j(t)dt|^2 = \sum_{i=1}^{d_N}\int_0^T E|h_i(t)Y^N(t)|^2dt.$$

We know from (5.5) that

$$\sup_{0\leq t\leq T} E|Y^N(t)|^2 \leq CQ_{1,N}. \tag{5.6}$$

Therefore, $E\|\tilde{J}^r(N)\|^2 \leq Cd_NQ_{1,N} \to 0$ by assumption (A2).

Next, we show that the eigenvalues of $\tilde{J}^d(N)$ are uniformly bounded from below. Indeed, for all sufficiently large N,

$$\phi^N(t) \geq C\left(1 - \frac{\sum_{k=1}^N k^{2(m_1-m)/d}e^{-2\alpha_k t}}{\psi_N}\right).$$

Therefore, it is enough to show that

$$\int_0^T |\zeta(t)|^2\frac{\sum_{k=1}^N k^{2(m_1-m)/d}e^{-2\alpha_k t}}{\psi_N}dt \leq \varepsilon_N\|\zeta\|^2 \tag{5.7}$$

and $\varepsilon_N \to 0$ as $N \to \infty$. By the first part of assumption (A3), $|\zeta(t)|^2 \leq D_N\|\zeta\|^2$, and, after integrating the rest and using the second part of assumption (A3), we get (5.7).

Now, once the eigenvalues of $\tilde{J}^d(N)$ are uniformly bounded from below, we conclude that

$$\|(\tilde{J}^d(N))^{-1}\| \leq C$$

(the norm is bounded by the inverse of the smallest eigenvalue of $\tilde{J}^d(N)$), and then

$$\begin{aligned}\|\tilde{J}^{-1}(N)\| \leq \quad & \|(\tilde{J}^d(N))^{-1}\| \cdot \|(I + (\tilde{J}^d(N))^{-1}\tilde{J}^r(N))^{-1}\| \\ & \leq C/(1 - C\|\tilde{J}^r(N)\|), \qquad P - \lim\|\tilde{J}^r(N)\| = 0,\end{aligned}$$

where the last inequality follows from the Neumann series for $(I + (\tilde{J}^d(N))^{-1}\tilde{J}^r(N))^{-1}$. This completes the proof of (C1).

2. The vector b.

We have

$$E\|b^N\|^2 = \sum_{i=1}^{d_N}\int_0^T h_i^2(t)EX^N(t)dt$$

and therefore we need

$$\lim_{N\to\infty} d_N/\psi_N = 0. \tag{5.8}$$

Since $\sup_{0\le t\le T}\sum_{i=1}^{d_N}|h_i(t)|^2 \ge d_N/T$, $\psi_N \sim F_{q,N}$, and (cf. Table 2) $Q_{2,N} \ge C/F_{q,N}$, equality (5.8) follows from assumption (A3). Convergence (C2) is proved.

3. The vector c.

We have

$$E\|c^N\|^2 \le E\sum_{i\ge1}\left(\int_0^T X^N(t)(\theta_0(t)-\theta_0^N(t))h_i(t)dt\right)^2 =$$

$$E\int_0^T |\theta_0(t)-\theta_0^N(t)|^2(X^N(t))^2dt \le \sup_t E(X^N(t))^2\int_0^T|\theta_0(t)-\theta_0^N(t)|^2dt,$$

and $\sup_t E(X^N(t))^2 \le C\psi_N^2$, because $E(X^N(t))^2 \le 2\psi_N^2(EY_N^2(t)+1)$ and, according to inequality (5.6), $\sup_t EY_N^2(t)\to 0$. Thus, convergence (C3) is proved.

This completes the proof of Theorem 3.1.

5.2 Asymptotic Normality

It follows from equation (5.3) and assumption (A6) that the limiting distributions of $\sqrt{F_{q,N}}(\theta^N - \theta_0, g)$ and $\sqrt{F_{q,N}}(\theta^N - \theta_0^N, g)$ are the same.

We begin with the following result. Recall that $\phi^N(t) = EX^N(t)/\psi_N$, and (cf. notation (N3)) define

$$\phi(t) = \frac{1/\tilde\theta(t)}{\int_0^T(1/\tilde\theta(t))dt}.$$

Proposition 5.1 *Under assumption of Theorem 3.1 we have*

$$\lim_{N\to\infty}\phi^N(t) = \phi(t) \ \text{ for almost all } 0<t\le T, \text{ and } \lim_{N\to\infty}\psi_N/F_{q,N} = \int_0^T dt/(2\tilde\theta(t)).$$

Proof. Due to (5.4) and since $\phi^N(t) = X^{1,N}(t) + EX^{2,N}(t)$ we can assume that $u_0 = 0$.

Assume first that $\theta(t)$ is smooth. Integrate by parts in the expression

$$EX^N(t) = \sum_{k=1}^N \nu_k^2\int_0^t\frac{1}{2\mu_k(s)}2\mu_k(s)\exp\left(-2\int_s^t\mu_k(r)dr\right)ds$$

to get

$$EX^N(t) = \sum_{k=1}^N\frac{\nu_k^2}{2\mu_k(t)} - \sum_{k=1}^N\frac{\nu_k^2}{2\mu_k(0)}\exp\left(-2\int_0^t\mu_k(r)dr\right)$$

$$+\sum_{k=1}^{N}\frac{\nu_k^2}{2}\int_0^t\frac{\mu_k'(s)}{\mu_k^2(s)}\exp\left(-2\int_s^t\mu_k(r)dr\right)ds.$$

Now divide everything by $F_{q,N}$ and pass to the limit as $N\to\infty$; the first term on the right will give $1/(2\tilde{\theta}(t))$, the second will give zero for $t>0$, because the series converges, the third term will also give zero, because $\mu_k'(s)/\mu_k^2(s)\le C/\alpha_k$ and we get another α_k in the denominator after integration. This completes the proof under the smoothness assumption. In general, we approximate $\theta(t)$ by smooth functions in $L_2(0,T)$ norm so that the convergence is also for almost all $t\in(0,T)$.

□

Corollary 5.2 *Under assumptions of Theorem 3.1 we have*

$$P-\lim_{N\to\infty}\frac{X^N(t)}{\psi_N}=\phi(t)\ \text{ for almost all }\ t\in(0,T) \tag{5.9}$$

and

$$\lim_{N\to\infty}\frac{\psi_N}{F_{q,N}}=\int_0^T\frac{dt}{2\tilde{\theta}(t)} \tag{5.10}$$

Indeed, to prove (5.9) note that $X^N(t)/\psi_N=\phi^N(t)+Y^N(t)$ and $Y^N(t)\to 0$ in probability for all t. The proof of (5.10) is obvious.

It follows from (5.3), (5.10), and assumption (A6) that, to prove asymptotic normality in the form (3.4), it is sufficient to show that

$$\lim_{N\to\infty}\sqrt{\psi_N}(\hat{\theta}^N-\theta_0^N,g\phi)=\mathcal{N}\left(0,\|g\sqrt{\phi}\|^2\right) \tag{5.11}$$

in distribution.

Proposition 5.3 *Under the assumptions of Theorem 3.2 we have*

$$\lim_{N\to\infty}\frac{(b^N,g)}{\sqrt{\psi_N}}=\mathcal{N}\left(0,\|g\sqrt{\phi}\|^2\right)$$

in distribution.

Proof. We have

$$\frac{(b^N,g)}{\sqrt{\psi_N}}=\frac{\sum_{i=1}^{d_N}b_i^N g_i}{\sqrt{\psi_N}}=\frac{\int_0^T\left(\sum_{i=1}^{d_N}g_ih_i(t)\right),dM^N(t)}{\sqrt{\psi_N}}$$

where $M^N(t) = \int_0^t (A_1 u^N(s), dW^N(s))_{L_2(G)}$. Then

$$\tilde{M}^N(t) := \frac{\int_0^t \left(\sum_{i=1}^N g_i h_i(s)\right) dM^N(s)}{\sqrt{\psi_N}}$$

is a continuous square integrable martingale with the bracket

$$\langle \tilde{M}^N \rangle_t = \frac{\int_0^t \left(\sum_{i=1}^{d_N} g_i h_i(s)\right)^2 X^N(s) ds}{\psi_N}.$$

According to (5.9), the bracket converges for each $t > 0$ in probability to

$$\int_0^t g^2(s)\phi(s)ds,$$

the bracket of $\int_0^t g(s)\sqrt{\phi(s)}dw_s$. Thus, by the martingale central limit theorem (Jacod and Shiryayev [1987, Theorem VIII.4.17]),

$$\lim_{N\to\infty} \frac{\sum_{i=1}^{d_N} b_i^N g_i}{\sqrt{\psi_N}} = \mathcal{N}\left(0, \int_0^T g^2(s)\phi(s)ds\right)$$

in distribution (recall that $g_i = \int_0^T h_i(t)g(t)dt$).

$\square$.

Corollary 5.4 *Let $g^N \in L_2(0,T)$ be a sequence of deterministic functions with*

$$g_i^N = \int_0^T g^N(t)h_i(t)dt,$$

and suppose that

$$\lim_{N\to\infty} \sum_{i=1}^{d_N} |g_i^N|^2 = 0.$$

Then

$$P - \lim_{N\to\infty} \frac{\sum_{i=1}^{d_N} b_i^N g_i^N}{\sqrt{\psi_N}} \equiv P - \lim_{N\to\infty} \frac{(b^N, g^N)}{\sqrt{\psi_N}} = 0.$$

The proof is obvious from the previous calculations (the bracket of the corresponding martingale now tends to zero).

In what follows, $\sqrt{\psi_N}(\hat{\theta}^N - \theta_0^N)$ will be denoted by $\tilde{\theta}^N$.

According to Proposition 5.3, the expression $(\tilde{J}(N)\tilde{\theta}^N, g) \equiv (\psi_N^{-1/2} b^N, g)$ is asymptotically normal with zero mean and variance $\|g\sqrt{\phi}\|^2$. Therefore, to establish (5.11) it remains to show that under assumption (A5) we have the convergence

$$P - \lim_{N\to\infty} (\tilde{\theta}^N, \tilde{J}(N)g - \phi g) = 0. \tag{5.12}$$

We will use the following result.

Lemma 5.5 *If $g^N \in L_2(0,T)$ is deterministic and $\|g^N\| \le C$, then*

$$P - \lim_{N\to\infty} (\tilde{\theta}^N, \tilde{J}^r(N) g^N) = 0$$

and if $\|g^N\| \to 0$, then

$$P - \lim_{N\to\infty} (\tilde{\theta}^N, g^N) = 0.$$

Proof. We have

$$|(\tilde{\theta}^N, \tilde{J}^r(N)h_N)| \le C\|\tilde{\theta}^N d_N^{-1/2}\| \cdot \|\tilde{J}^r(N) d_N^{1/2}\| \le \|\tilde{J}^{-1}(N)\| \cdot \|(\psi_N d_N)^{-1/2} b^N\| \cdot \|\tilde{J}^r(N) d_N^{1/2}\| \to 0$$

in probability, because $\|\tilde{J}^{-1}(N)\| \le C/(1 - C\|\tilde{J}^r(N)\|)$ with $P - \lim_{N\to\infty} \|\tilde{J}^r(N)\| = 0$,

while $E\|(\psi_N d_N)^{-1/2} b^N\|^2 \le C$ and $E\|\tilde{J}^r(N) d_N^{1/2}\|^2 \le d_N^2 Q_{1,N} \to 0$ by assumption (A5). Next, if $\|g^N\| \to 0$, then

$$(\tilde{\theta}^N, g^N) = (\tilde{J}^d(N)\tilde{\theta}^N, (\tilde{J}^d(N))^{-1} g^N) = (b^N/\sqrt{\psi_N}, (\tilde{J}^d(N))^{-1} g^N) - (\tilde{\theta}^N, \tilde{J}^r(N)(\tilde{J}^d(N))^{-1} g^N),$$

where the first term converges to zero by Proposition 5.3, and we just saw that the second term converges to zero as well.

□

We now show that Lemma 5.5 implies (5.12). To this end, denote by Π^N the orthogonal projection on the span of $h_1, \ldots, h_{d_N}$. Since $\tilde{\theta}^N \in \Theta_N$, it is enough to show that

$$P - \lim_{N\to\infty} (\tilde{\theta}^N, \tilde{J}(N)g - \Pi^N \phi g) = 0.$$

Note that $\tilde{J}(N)$ is the matrix representation of the operator $\Pi^N X^N/\psi_N \Pi^N$, where X^N/ψ_N is the multiplication operator by the function $X^N(t)/\psi_N$. With this convention, if $\tilde{J}(N) = \tilde{J}^d(N) + \tilde{J}^r(N)$, then $\tilde{J}^d(N) = \Pi^N \phi^N \Pi^N$ and $\tilde{J}^r(N) = \Pi^N Y^N \Pi^N$.

As a result,

$$\tilde{J}(N)g - \Pi^N \phi g = \tilde{J}^r(N)g + \Pi^N(\phi^N - \phi)\Pi^N g + \Pi^N \phi(\Pi^N g - g),$$

and it remains to apply Lemma 5.5 three times.

This completes the proof of asymptotic normality.

References

Aihara, S. (1992) Regularized maximum likelihood estimate for an infinite dimensional parameter in stochastic parabolic systems, *SIAM J. Cont. Optim.* **30**, 745-764.

Barndorff-Nielsen, O. and Sørensen, M. (1994). A review of some aspects of asymptotic likelihood theory for stochastic processes. *International Statistical Review* **62**, 133-165.

Birgé. L. and Massart, P. (1997). From model selection to adaptive estimation. *Festschrift for Lucien Le Cam*, 55-87. Springer, New York.

Chentsov, N. (1982). Statistical decision rules and optimal inference, *Transl. Math. Monogr.***53**.

Devore, R. and Lorentz, G. (1993). *Constructive approximation.* Springer-Verlag, Berlin.

Genon-Catalot, V. and Jacod, J. (1993). On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Ann. Inst. H. Poincare* **29**, 119-151.

Golubev, G. and Khasminskii, R. (1997). Statistical approach to inverse boundary problems for partial differential equations. *Preprint*, Weierstrass Institut of Applied Stochastik, Berlin.

Grenander, U. (1981). *Abstract inference*, Wiley, New York.

Huebner, M., Khasminskii, R. and Rozovskii, B. (1992). Two examples of parameter estimation. In: Cambanis, Ghosh, Karandikar, Sen [ed.], *Stochastic Processes*, Springer-Verlag, New York.

Huebner, M. and Rozovskii, B. (1995). On Asymptotic Properties of Maximum Likelihood Estimators for Parabolic Stochastic PDEs. *Prob. Theory Related Fields* **103**, 143-163.

Ibragimov, I. and Khasminskii, R. (1982). *Statistical Estimation.* Springer, New York.

Ibragimov, I.and Khasminskii, R. (1997). Some nonparametric estimation problems for parabolic SPDEs. *Research report* **37**, Wayne State University.

Jacod, J. and Shiryayev, A. (1987). *Limit theorems for stochastic processes.* Springer, New York.

Kutoyants, Yu.A., (1984a). *Parameter estimation for stochastic processes*, Heldermann, Berlin.

Kutoyants, Yu.A., (1984b). On nonparametric estimation of trend coefficients in a diffusion process. Proc. Steklov Sem. *Statistics and Control of stochastic processes.* 230-250.

LeCam, L. (1986). *Asymptotic methods in statistical decision theory.* Springer, New York.

Liptser, R.Sh. and Shiryayev, A.N. (1977). *Statistics of Random Processes, volume I.* Springer, New York.

Nguyen, H. and Pham, T. (1982). Identification of nonstationary diffusion model by the method of sieves. *SIAM J. Control Optim.* **20**, 603-611.

Nussbaum, M. (1996). Asymptotic Equivalence of density estimation and Gaussian white noise. *Ann. Statist.* **24**, 2399-2430.

Piterbarg, L. and Rozovskii, B. (1997). On asymptotic properties of parameter estimation in stochastic PDE's: discrete time sampling. *Methods Math. Stat.* **6** (2), 200–223.

Polyak, B. and Tsybakov, A. (1990). Asymptotic optimality of the C_p-test for the orthogonal series estimation of regression. *Theory Probab. Appl.* **35**, 293-306.

Safarov, Yu. and Vassiliev, D. (1997). The asymptotic distribution of eigenvalues of partial differential operators. *Transl. Math. Monogr.*, **155**.

Verulava, Yu. and Polyak, B. (1988). Selection of the regression model order. *Automat. Remote Control* **49**, 1482-1494.

Walsh, J. (1984). An introduction to stochastic partial differential equations. *Lecture Notes Math.*, **1180**, 265–439.

Appendix

Proof that the matrix $J(N)$ is invertible with probability one. Assume that $J(N)$ is singular on a set $A \in \mathcal{F}$ of positive measure. Then there is a random vector $\xi(\omega) \in I\!R^{d_N}$ that is not zero on A and, also on A,

$$(\xi, J(N)\xi) \equiv \int_0^T |\xi(t)|^2 X^N(t)dt = 0,$$

where $\xi(t) = \sum_{i=1}^{d_N} \xi_i h_i(t)$ (cf. notation (N6)).

Note that $\int_0^T E(X^N(t))^{-1/3}dt < \infty$ (direct computation when $N = 1$ and u_0=0 so that $X^N(t)$ is a square of a normal random variable with zero mean and variance of order t for t near 0; when $N > 1$ and/or $u_0 \neq 0$, the value of the expectation decreases). As a result, by Hölder's inequality

$$\int_A \int_0^T |\xi(t)|^{1/2} dtdP \leq \left(\int_A \int_0^T |\xi(t)|^2 X^N(t)dtdP\right)^{1/4} \left(\int_0^T E(X^N(t))^{-1/3}dt\right)^{3/4} = 0,$$

and so

$$\|\xi\|^2 \equiv \int_0^T |\xi(t)|^2 dt = 0 \quad (P\text{-a.s. on } A),$$

which is a contradiction.

Asymptotics of the eigenvalues of partial differential operators.

As before, G is either a smooth domain in $I\!R^d$ or a smooth d-dimensional manifold. Let A be an order $2n$ differential operator on G with complex coefficients. For technical reasons we write A in the form (cf. (2.1))

$$A = \sum_{|\alpha|,|\beta| \leq n} D^\alpha(a^{\alpha\beta} D^\beta),\ a^{\alpha\beta} \in C_b^\infty(G), \tag{5.13}$$

where $D^\alpha u(x) = (-\sqrt{-1})^{|\alpha|} u^{(\alpha)}(x)$. If G is a bounded domain, then the operator A is supplemented with zero boundary conditions

$$u^{(\alpha)}|_{\partial G} = 0 \quad \text{ for all } |\alpha| \leq n-1.$$

The operator A is called *symmetric* if $a^{\alpha\beta}(x) = a^{\beta\alpha}(x)$ for all $x \in G$.

The function

$$\mathcal{P}_A(x, \xi) = \sum_{|\alpha|,|\beta|=n} a^{\alpha\beta} \xi^\alpha \xi^\beta,$$

where $\xi^\alpha = \xi_1^{\alpha_1} \ldots \xi_d^{\alpha_d}$, is called the *principal symbol* of the operator A. The operator A is called *uniformly elliptic* in G if there is a number $\delta > 0$ so that

$$\inf_{x \in G} Re\,(\mathcal{P}_A(x, \xi)) \geq \delta \sum_{|\alpha|=n} \xi^{2\alpha}$$

for all $\xi \in I\!R^d$.

Proposition A.1 (Safarov and Vassiliev [1997, Remark 1.2.2]). Let A be a symmetric operator of the form (5.13) and assume that A is uniformly elliptic in G. Then the asymptotics of the eigenvalues corresponding to the problem $Au(x) = \lambda u(x)$ is given by

$$\lambda_k = -\zeta_A k^{2n/d} + o(k^{2n/d}),$$

where

$$\zeta_A = \left(\frac{1}{(2\pi)^d} \int_{\{(x,\xi):\mathcal{P}_A(x,\xi)<1\}} dx d\xi \right)^{-2n/d}.$$

M. Huebner, Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, huebner@stt.msu.edu