Asymptotic Analysis of a Kernel Estimator for Parabolic SPDEs with Time-Dependent Coefficients

By M. Huebner and S. Lototsky*

Michigan State University and Massachusetts Institute of Technology

Published in *Annals of Applied Probability,* Vol. 10, No. 4, pp. 1–13, 2000.

Abbreviated Title: Kernel estimator for SPDEs

Abstract

In this paper we construct a kernel estimator of a time-varying coefficient of a strongly elliptic partial differential operator in a stochastic parablic equation. The equation is assumed diagonalizable, that is, all the operators have a common system of eigenfunctions. The mean-square convergence of the estimator is established. The rate of convergence is determined both by the smoothness of the true coefficient and the asymptotics of the eigenvalues of the operators in the equation.

1 Introduction

Stochastic partial differential equations arise naturally to describe spatially distributed populations (Dawson (1980)) or growth of interacting populations (De (1987)). Other applications include oceanography where tracer evolution may be described by a stochastic PDE (see Piterbarg and Rozovskii (1996) or Piterbarg (1998)).

After a suitable model is formulated for a particular application, it is necessary to estimate relevant model parameters. In models described by linear stochastic partial differential equations (SPDEs), such parameters are often the coefficients of the corresponding partial differential operators. Estimation problems for such SPDEs are entirely different from traditional problems of statistical inference when the unknown function is the coefficient of the "leading" differential operator. In

*Supported in part by NSF Grant DMS-9972016

[0]*Key words:* kernel estimator, stochastic partial differential equations.
AMS Classification: 62G05, 60H15

this case all the information about the unknown coefficient can be extracted from the observations of the solution on a finite time interval with a fixed amplitude of the random perturbation.

One method to construct a computable estimator utilizes finite dimensional projections of the observation process, for example, the first N (spatial) Fourier coefficients. The dimension of the projection is used to describe the asymptotic properties of the estimate. The number of spatial modes is also a natural asymptotic parameter from the physical point of view, as pointed out by Piterbarg (1998). In parametric models, when the coefficient is 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 $N \to \infty$. Two special cases of these results were discussed earlier by Huebner, Khasminskii and Rozovskii (1993). Parametric models for infinite dimensional systems have also been studied by Piterbarg and Rozovskii (1997) who analyzed the asymptotic properties of the maximum likelihood estimator in the discrete time sampling case. Furthermore, Lototsky and Rozovskii (1999) studied parameter estimation when the operators in the SPDEs do not commute. Mohapl (1997) constructed consistent estimators of constant coefficients occuring in a hyperbolic SPDE with observations on a grid as the number of time and number of space observations become larger. Other inverse problems for SPDEs in the small noise asymptotics such as the estimation of a source are discussed in Chow, Ibragimov and Khasminskii (1999).

In this paper we construct a kernel-type estimator for a *time-varying* parameter in a stochastic parabolic equation. We study the optimal rate of convergence of such estimators. Although the problem of nonparametric estimation for ordinary stochastic differential equations has received a lot of attention (see e.g. Ibragimov and Khasminskii (1981), Kutoyants (1984)), little has been done concerning nonparametric estimation for infinite dimensional systems. For stochastic evolution systems Ibragimov and Khasminskii (1997) studied asymptotic properties of kernel estimators of general functions in the small noise asymptotics when the probability measures generated by the processes corresponding to different functions are equivalent. Other results, for example by Aihara (1998), and Aihara and Sunahara (1988), are concerned with the problem of estimating a spatially varying parameter in stochastic diffusion equations when the observation process is finite-dimensional.

For a stochastic ordinary differential equation, Kutoyants (1984) proved mean-square convergence of a kernel-type estimator for the drift term. In this paper we utilize the methods developed by Huebner and Rozovskii (1995) and by Kutoyants (1984) to 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 $x \in G \subset I\!R^d$ is governed by the following equation:

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

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 kernel estimator of $\theta_0(t)$ as the number N of the observed Fourier coefficients increses. To simplify the analysis, it is assumed that the equation is diagonalizable, that is, the operators A_0 and A_1 have a common system of eigenfunctions. If the initial condition u_0 is not random, then the Fourier coefficients $u_1(t),\ldots,u_N(t)$ are independent Ornstein–Uhlenbeck process, and the drift of each process contains the unknown function $\theta_0(t)$ and the eigenvalues of the operators A_0, A_1.

In Kutoyants (1984), the trend coefficient in a diffusion process was estimated from the N i.i.d. copies of the process. Even though the observations u_k in our case are not identically distributed, we use a similar approach and consider the estimate $\hat\theta^N$ of θ_0 as follows:

$$\hat\theta^N(t) = \int_0^T R_{h_N}(s-t)\,dX^N(s),$$

where R is a kernel function, $R_{h_N}(s) = R(s/h_N)/h_N$ with $h_N \to 0, N \to \infty$, and X^N is a certain process constructed from the observations $u_1,\ldots,u_N$. We prove the mean-square convergence of the type

$$\lim_{N\to\infty} \sup_{\theta_0\in\Theta} \sup_{t\in[t_1,t_2]} N^\gamma E|\hat\theta^N(t) - \theta_0(t)|^2 < \infty,$$

and explicitely compute the rate $\gamma > 0$ which is determined by the parameter class Θ and the orders of the operators A_0, A_1.

The paper is organized as follows. In Section 2 we introduce the mathematical model and the basic notations. The main results on the asymptotic properties of the kernel-type estimator, including convergence rates, are proven in Section 3. An example follows in Section 4.

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 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, either a smooth bounded domain in $\mathbb{R}^d$ or a smooth d-dimensional compact manifold (without boundary). We denote by A_0 and A_1 partial differential operators on G with real coefficients. If G is a domain, then the operators are supplemented with zero boundary conditions. We assume that

$$A_i u(x) = \sum_{|\alpha|\le m_i} a_i^\alpha(x) u^{(\alpha)}(x),\ \ a_i^\alpha \in C_b^\infty(G),\ \ i = 0, 1, \tag{2.1}$$

where $\alpha = (\alpha_1, \ldots, \alpha_d)$, $\alpha_i = 0, 1, \ldots,$ $|\alpha| = \sum_{i=1}^d \alpha_i$,

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

and the functions $a_i^\alpha(x)$ are known.

The observation process is governed by the following equation:

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

where $\theta_0 = \theta_0(t)$ 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, uniformly in $t \in [0,T]$, $\mu_k(t) := -(\kappa_k + \theta_0(t)\nu_k) \sim k^{2m/d}$, $2m = \max\{m_0, m_1\}$, which means that

$$\alpha_k \leq -(\kappa_k + \theta_0(t)\nu_k) \leq \beta_k$$

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.

Assumptions (H1) and (H2) hold in many physical models (see, for example, Piterbarg and Rozovskii (1996)). 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 in Appendix a precise statement about the eigenvalues and eigenfunctions of elliptic operators. 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*

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

Proof. This follows from Theorem 3.1.4 in Rozovskii (1990).

Remark 2.2 *1. The process* W *in equation (2.2) can have an invertible correlation operator* B *as long as the eigenfunctions of* B *are also* φ_k*. We can then reduce the equation to the standard form with cylindrical Brownian motion by replacing the initial condition* u_0 *with* $B^{-1}u_0$ *and the operators* A_i*,* $i = 0, 1$*, with* $B^{-1}A_iB$*.*

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 in the equation have the properties (H1) and (H2).

3 Main Result

If $u = u(t,x)$ is a solution of (2.2) with the operators A_0, A_1 satisfying (H1), and $\{\varphi_k\}_{k\ge1}$ is the common system of eigenfunction for A_0, A_1, then $u_k(t) = (u, \varphi_k)(t)$ is a solution of

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

where $\mu_k(t) = -(\kappa_k + \theta_0(t)\nu_k)$ and κ_k, ν_k are the eigenvalues of A_0, A_1. The objective is to construct a kernel estimate of $\theta_0(t)$ for every $t \in (0,T)$ on the basis of $u_k(t)$, $t \in (0,T)$, $k = 1, \ldots, N$.

Recall that the function $R = R(t)$, $t \in I\!R$ is called a compactly supported kernel of order $K \ge 1$ if R has the following properties:

1. $R(t) = 0$ for large $|t|$,
2. $\int_{I\!R} R(t)dt = 1$,
3. $\int_{I\!R} t^j R(t)dt = 0$ for $j = 1, \ldots, K$.

For example, $R(t) = 3(3-5t^2)/8\, I(|t| \le 1)$ is a compactly supported kernel of order 3. More examples and a general procedure for constructing such kernels are presented in Devroye (1987) and Müller (1984). As usual, a scaled kernel $R(t/h_N)/h_N$ will be used with some bandwidth h_N so that $0 < h_N \to 0$, $N \to \infty$. The exact asymptotics of h_N will be specified later.

Formally, the estimate of θ_0 at point t_0 is constructed as a weighted sum of the integrals

$$\frac{1}{h_N}\int_0^T R\left(\frac{t-t_0}{h_N}\right)\frac{du_k(t) - \kappa_k u_k(t)dt}{u_k(t)}.$$

However his expression must be modified, since this integral may not be defined due to the vanishing Fourier coefficients $u_k(t)$. Let $\{v_N,\ N \ge 1\}$ be a sequence of positive real numbers so that $v_N \downarrow 0$ as $N \to \infty$. The random processes $U_{k,N} = U_{k,N}(t)$, $k = 1, \ldots, N$, $t \in (0,T)$, are defined by

$$U_{k,N}(t) = \begin{cases} 1/u_k(t), & |u_k(t)| > v_N, \\ 1/v_N, & |u_k(t)| \le v_N. \end{cases}$$

To formulate the main result we need the following weight sequence

$$F_{\nu,N} = \sum_{k=1}^{N} \nu_k,$$

where ν_k are the eigenvalues of A_1.

For every $t \in (0,T)$, we define the estimate $\hat{\theta}^N(t)$ of $\theta_0(t)$ as follows:

$$\hat{\theta}^N(t) = \frac{1}{h_N F_{\nu,N}} \sum_{k=1}^{N} \int_0^T R\left(\frac{s-t}{h_N}\right) U_{k,N}(s)(du_k(s) - \kappa_k u_k(s)ds). \tag{3.2}$$

It is clear that a consistent estimate of $\theta_0(t)$ at fixed t is possible only if the function θ_0 is sufficiently smooth. Therefore we define the class of functions under consideration as follows.

Definition 3.1 *For a positive real number β represented as $\beta = K + \alpha$, where $K \ge 0$ is an integer and $\alpha \in (0,1]$, denote by Θ_L^β the set of K times continuously differentiable functions on $(0,T)$ with the following properties:*

(P1) For all $\theta \in \Theta_L^\beta$, $|\theta^{(K)}(t) - \theta^{(K)}(s)| \le L|t-s|^\alpha$, $t, s \in (0,T)$;

(P2) There exist $C_1, C_2, N_0 > 0$ so that, with $\mu_k(t) = -(\kappa_k + \nu_k\theta(t))$, $C_1 k^{2m/d} \le \mu_k(t) \le C_2 k^{2m/d}$ for all $k > N_0$, all $t \in (0,T)$, and all $\theta \in \Theta_L^\beta$.

The second condition ensures that equation (2.2) is solvable uniformly for $\theta_0 \in \Theta^{\beta}_{L}$. Note that the value of K in the representation $\beta = K + \alpha$ is the same as the order of the kernel. The main result of this paper follows.

Theorem 3.1 *In addition to (H1) and (H2), let the following conditions be fulfilled:*

(A1) The initial condition u_0 is deterministic and belongs to $H^{-s}(G)$ for some $s > d/2$;

(A2) The orders of the operators A_0 and A_1 are such that $q := 2(m_1 - m)/d > -1$;

(A3) The eigenvalues ν_k of A_1 are such that $|F_{\nu,N}| \sim N^{1+m_1/d}$;

(A4) The function θ_0 belongs to Θ^{β}_{L} with $\beta = K + \alpha$, and R is a bounded, compactly supported kernel of order K;

(A5) $h_N \sim N^{-(q+1)/(4\beta+1)}$, $v_N \sim h_N^{2\beta} N^{-m/d}$.

Then, for every $0 < t_1 < t_2 < T$,

$$\lim_{N\to\infty} \sup_{\theta_0 \in \Theta^{\beta}_{L}} \sup_{t\in[t_1,t_2]} N^{2(q+1)\beta/(4\beta+1)} E|\hat{\theta}^N(t) - \theta_0(t)|^2 < \infty.$$

Remark 3.2 *1. From Huebner and Rozovskii (1995) it is known that the condition $q \geq -1$ (cf. assumption (A2)) is necessary to have a consistent estimate of θ_0 in our model even if θ_0 is a constant. However, unlike in the constant parameter case, a consistent estimate of the type (3.2) is not possible in the critical case $q = -1$ as can be seen by analyzing the proof below.*

2. The rate of growth of $F_{\nu,N}$ in (A3) is an assumption about the asymptotics of the eigenvalues of A_1. For example, if A_1 is a self-adjoint elliptic operator of order m_1, then $\nu_k \sim -k^{m_1/d}$ and (A3) holds.

3. Note that the rate of convergence is determined both by the assumed smoothness of the function θ and by the order of the kernel. An analysis of the proof shows that if a lower order kernel is used with the order $K_1 < K$, then the rate of convergence will be determined by $\beta = K_1 + 1$ instead of $\beta = K + \alpha$. In particular, if the coefficient is known to be infinitely differentiable, then the rate of convergence is determined by the order of the kernel used.

Proof. In the following, C denotes a positive real number depending on d, t_1, t_2, T, the operators A_0 and A_1, the kernel R, and the space Θ_L^β. In particular, C does *not* depend on the time variable $t \in [t_1, t_2]$ nor on a function $\theta_0 \in \Theta_L^\beta$. The value of C can be different in different places.

With no loss of generality, assume that R is supported in $[-1, 1]$ and N is so large that $t_1/h_N > 2$, $(T - t_2)/h_N > 1$. By (H1), we can also assume that $\mu_k(t) > Ck^{2m/d}$ for all $k \geq 1$.

We split up the difference $\hat{\theta}^N(t) - \theta_0(t)$ into three parts $\hat{\theta}^N(t) - \theta_0(t) = J_1 + J_2 + J_3$, which are then estimated seperately. Here

$$\begin{aligned} J_1 &= \frac{1}{h_N F_{\nu,N}} \sum_{k-1}^{N} \int_0^T \nu_k R\left(\frac{s-t}{h_N}\right)(\theta_0(s) - \theta_0(t))ds, \\ J_2 &= \frac{1}{h_N F_{\nu,N}} \sum_{k=1}^{N} \int_0^T R\left(\frac{s-t}{h_N}\right) U_{k,N}(s) dw_k(s), \\ J_3 &= \frac{1}{h_N F_{\nu,N}} \sum_{k=1}^{N} \int_0^T \nu_k R\left(\frac{s-t}{h_N}\right) \theta_0(s) I(|u_k(s)| \leq v_N) \left(\frac{u_k(s)}{v_N} - 1\right) ds. \end{aligned}$$

Since R has compact support and integrates to one, we have for every $t \in (t_1, t_2) \subset (0, T)$ and all sufficiently large N

$$J_1 = \int_{-t/h_N}^{(T-t)/h_N} R(s)(\theta_0(t + h_N s) - \theta_0(t))ds.$$

By the Taylor formula,

$$\begin{aligned} \theta_0(t+\tau) &= \theta_0(t) + \sum_{m=1}^{K-1} \frac{\tau^m}{m!} \theta_0^{(m)}(t) + \frac{\tau^K}{K!} \theta_0^{(K)}(t + \gamma\tau) \\ &= \theta_0(t) + \sum_{m=1}^{K} \frac{\tau^m}{m!} \theta_0^{(m)}(t) + \frac{\tau^K}{K!} (\theta_0^{(K)}(t + \gamma\tau) - \theta_0^{(K)}(t)), \quad \gamma = \gamma(\tau) \in (0, 1). \end{aligned}$$

Using a property of the kernel,

$$J_1 = \int_{-t/h_N}^{(T-t)/h_N} R(s) h_N^K \frac{s^K}{K!} (\theta_0^{(K)}(t + \gamma h_N s) - \theta_0^{(K)}(t))ds.$$

Property (P1) of the function class assures that $|\theta_0^{(K)}(t + \gamma\tau) - \theta_0^{(K)}(t)| \leq L_\alpha |\tau|^\alpha$. This implies

$$|J_1| \leq C h_N^\beta \int_{-1}^{1} |s|^\beta |R(s)| ds \leq C h_N^\beta.$$

Therefore,

$$|J_1|^2 \leq C h_N^{2\beta}.$$

The Gaussian random variable $u_k(t)$ is given by

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

with mean

$$M_k(t) = u_{0k} \exp\left(-\int_0^t \mu_k(s)ds\right)$$

and variance

$$D_k^2(t) = \int_0^t \exp\left(-2\int_s^t \mu_k(\tau)d\tau\right) ds.$$

By property (P2)

$$D_k(t) \geq Ck^{-m/d}(1-e^{-Ct}).$$

Note that

$$P(|u_k(t)| \leq v_N) \leq \frac{v_N}{D_k(t)} \quad \text{and} \quad EU_{k,N}^2(t) \leq \frac{2}{v_N D_k(t)}.$$

Indeed,

$$P(|u_k(t)| \leq v_N) = \frac{1}{\sqrt{2\pi}D_k(t)} \int_{-v_N}^{v_N} \exp\left(-\frac{(x-M_k(t))^2}{2D_k(t)^2}\right) dx \leq \frac{2v_N}{\sqrt{2\pi}D_k(t)} \leq \frac{v_N}{D_k(t)},$$

$$EU_{k,N}^2(t) = \frac{P(|u_k(t)| \leq v_N)}{v_N^2} + \frac{1}{\sqrt{2\pi}D_k(t)} \int_{|x|>v_N} \exp\left(-\frac{(x-M_k(t))^2}{2D_k^2}\right) \frac{dx}{x^2},$$

$$\frac{1}{\sqrt{2\pi}D_k(t)} \int_{|x|>v_N} \exp\left(-\frac{(x-M_k(t))^2}{2D_k^2}\right) \frac{dx}{x^2} \leq \frac{2}{\sqrt{2\pi}D_k(t)} \int_{v_N}^{+\infty} \frac{dx}{x^2} \leq \frac{1}{v_N D_k(t)}.$$

Now we estimate J_2.

$$\begin{aligned}
E|J_2|^2 &= \frac{1}{h_N^2 F_{\nu,N}^2} \sum_{k=1}^N \int_0^T R^2\left(\frac{s-t}{h_N}\right) EU_{k,N}^2(s)ds \\
&\leq \frac{C}{h_N^2 F_{\nu,N}^2 v_N} \sum_{k=1}^N \int_0^T R^2\left(\frac{s-t}{h_N}\right) \frac{ds}{D_k(s)} \\
&\leq \frac{CN^{(m-2m_1)/d-1}}{h_N v_N} \int_{-t/h_N}^{(T-t)/h_N} \frac{R^2(s)}{1-e^{-C(t+sh_N)}} ds \\
&\leq \frac{CN^{(m-2m_1)/d-1}}{h_N v_N} \int_{-1}^{1} \frac{R^2(s)}{1-e^{-Ct_1/2}} ds \\
&\leq \frac{CN^{(m-2m_1)/d-1}}{h_N v_N}.
\end{aligned}$$

To estimate J_3, we use an inequality for independent square integrable random variables ξ_k: $E\left(\sum_{k=1}^N \xi_k\right)^2 \le 2\sum_{k=1}^N var(\xi_k) + \left(\sum_{k=1}^N E\xi_k\right)^2$. Then

$$\begin{aligned}
E|J_3|^2 &\le E\left(\frac{C}{h_N F_{\nu,N}}\sum_{k=1}^N \int_0^T \nu_k R\left(\frac{s-t}{h_N}\right) I(|u_k(s)| \le v_N) ds\right)^2 \\
&\le \left(\frac{C}{h_N F_{\nu,N}}\sum_{k=1}^N \int_0^T \nu_k R\left(\frac{s-t}{h_N}\right) P(|u_k(s)| \le v_N) ds\right)^2 \\
&\quad + \frac{C}{F_{\nu,N}^2}\sum_{k=1}^N E\left(\frac{1}{h_N}\int_0^T \nu_k R\left(\frac{s-t}{h_N}\right) I(|u_k(s)| \le v_N) ds\right)^2.
\end{aligned}$$

The first term on the right is bounded by

$$\left(\frac{Cv_N}{h_N F_{\nu,N}}\sum_{k=1}^N \int_0^T \nu_k R\left(\frac{s-t}{h_N}\right)\frac{ds}{D_k(s)}\right)^2 \le Cv_N^2 N^{2m/d}.$$

The second term can be rewritten as

$$\frac{C}{F_{\nu,N}^2}\sum_{k=1}^N \frac{1}{h_N^2}\int_0^T\int_0^T \nu_k^2 R\left(\frac{s_1-t}{h_N}\right) R\left(\frac{s_2-t}{h_N}\right) E[I(|u_k(s_1)| \le v_N) I(|u_k(s_2)| \le v_N)] ds_1 ds_2. \quad (3.3)$$

Using $E[I(|u_k(s_1)| \le v_N)I(|u_k(s_2)| \le v_N)] \le E[I(|u_k(s_1)| \le v_N)] \le C/(v_N D_k(s_1))$ and repeating the arguments used to estimate $E|J_2|^2$, the expression in (3.3) can be bounded by $Cv_N N^{m/d-1}$. As a result,

$$E|\hat{\theta}^N(t) - \theta_0(t)|^2 \le C\cdot\left(h_N^{2\beta} + \frac{N^{(m-2m_1)/d-1}}{h_N v_N} + v_N N^{m/d}(v_N N^{m/d} + 1/N)\right).$$

If $h_N \sim N^{-(q+1)/(4\beta+1)}$ and $v_N N^{m/d} \sim h_N^{2\beta}$, then $E|\hat{\theta}^N(t) - \theta_0(t)|^2 \le CN^{-2\beta(q+1)/(4\beta+1)}$, which completes the proof.

4 Example

In this section we give an example to illustrate the assumptions of the main theorem. Suppose that $u = u(t,x)$, $0 < t < 1$, $0 < x < 1$, is a solution of

$$\begin{aligned}
du(t,x) &= [\theta_0(t)\Delta u(t,x) - u(t,x)]dt + dw(t,x), \\
u(0,x) &= 0, \\
u(t,0) &= u(t,1) = 0,
\end{aligned}$$

where $\theta_0(t)$ is a smooth (infinitely differentiable) function so that assumptions (H1) and (H2) hold. For example, $\theta_0(t) = 2 + \sin t$ or $\theta_0(t) = 3 - 2\cos(2t)$. Using the notations of Section 2, we have $G = (0, 1)$, $d = 1$, $A_0 = -I$ (I is the identity operator), $A_1 = \Delta$ (the Laplace operator). Note that $m_0 = 0, m_1 = 2m = 2$ and so $q = 2(m_1 - m)/d = 2$. The eigenfunctions $\varphi_k(x) = \sqrt{2}\sin(\pi kx)$, $k \geq 1$, have corresponding eigenvalues $\kappa_k = -1$, $\nu_k = -\pi^2 k^2$. The solution can be written as

$$u(t,x) = \sum_{k \geq 1} u_k(t)\varphi_k(x)$$

where $u_k(t)$ satisfies

$$du_k(t) = -[1 + \theta_0(t)\pi^2 k^2]u_k(t)dt + dw_k(t), \;\; u_k(0) = 0.$$

The series for the solution converges in $L_2\Big(\Omega \times (0,1); L_2((0,1))\Big)$.

Choose $R(t) = 3(3-5t^2)/8I(|t| \leq 1)$, a compactly supported kernel of order $K = 3$. By assumption, θ_0 is smooth, and then, according to Remark 3 after Theorem 3.1, the rate of convergence is determined by $\beta = 4$. Following Theorem 3.1, we take $F_{\nu,N} = -\pi^2 \sum_{k=1}^N k^2$, $h_N = N^{-3/17}$, $v_N = N^{-41/17}$. The estimator is

$$\hat{\theta}^N(t) = \frac{1}{h_N F_{\nu,N}} \sum_{k=1}^N R\left(\frac{s-t}{h_N}\right) U_{k,N}(s)\Big(du_k(s) + u_k(s)ds\Big),$$

where

$$U_{k,N}(t) = \begin{cases} 1/u_k(t), & |u_k(t)| > v_N, \\ 1/v_N, & |u_k(t)| \leq v_N. \end{cases}$$

By Theorem 3.1, $\sup_{t_0 \leq t \leq t_1} E|\hat{\theta}^N(t) - \theta(t)|^2 \leq C(t_0,t_1)N^{-24/17}$.

The issues not addressed in this example are the selection of the best kernel or the optimal choice of bandwidth. For example, the same asymptotical result would hold if we choose $h_N = 100N^{-3/17}$ and $v_N = 0.1N^{-41/17}$. Also, a better rate of convergence can be achieved by taking a higher order kernel, but this will also increase the computational complexity. These important and interesting finite sample issues have to be addressed in the future.

References

[1] Aihara, S. (1998). Identifications of a discontinuous parameter in stochastic parabolic systems. *Appl. Math. Optim.* **37**, 43–69.

[2] Aihara, S. and Sunahara, Y. (1988). Identification of an infinite-dimensional parameter for stochastic diffusion equations. *SIAM J. Contr. Opt.* **26**, 1062–1075.

[3] Chow, P., Ibragimov, I. and Khasminskii, R. (1999). Statistical approach to some ill-posed problems for linear partial differential equations. *Prob. Theory Related Fields* **113**, 421-441.

[4] Dawson, D. (1980) Qualitative behavior of geostochastic systems. *Stoch. Proc. Appl.* **10**, 1–31.

[5] De, S. (1987) Stochastic models of population growth and spread. *Bull. Math. Biol.* **49**, 1–11.

[6] Devroye, L. (1987). *A course in density estimation.* Birkhäuser, Boston.

[7] Huebner,M., Khasminskii, R. and Rozovskii, B. (1993). Two examples of parameter estimation for stochastic partial differential equations. In: Cambanis, S., Ghosh, J., Karandikar, R., Sen, P., eds., *Stochastic processes.* Springer Verlag, New York. 149-160.

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

[9] Ibragimov, I. and Khasminskii, R. (1997). Some estimation problems for stochastic partial differential equations. *Dokl. Acad. Nauk* **353**, 300-302.

[10] Ibragimov, I. and Khasminskii, R. (1981) *Statistical Estimation.* Springer, New York.

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

[12] Kutoyants, Yu.A., (1984). On nonparametric estimation of trend coefficients in a diffusion process. Proceedings of the Steklov Seminar *Statistics and Control of stochastic processes.* 230–250.

[13] Lototsky, S. and Rozovskii, B. (1999). Spectral asymptotics of some functionals arising in statistical inference for SPDEs. *Stoch. Proc. Appl.* **79**, 69–94.

[14] Mohapl, J. (1997) On estimation in planar Ornstein-Uhlenbeck process. *Commun. Statist.-Stochastic Models* **13**, 435-455.

[15] Müller, H.-G. (1984) Smooth optimum kernel estimators of densities, regression curves and modes. *Ann. Statist.* **12** (1984), 766-774.

[16] Piterbarg, L. (1998). Parameter estimation for stochastic partial differential equations with applications to the upper ocean transport, *Preprint*.

[17] Piterbarg, L. and Rozovskii, B. (1996). Maximum Likelihood Estimators in the Equations of Physical Oceanography. In Adler, R., Müller, P., Rozovskii, B., eds., *Stochastic Modeling in Physical Oceanography*, pages 397–421. Birkhäuser, Boston.

[18] Piterbarg, L. and Rozovskii, B. (1997). On asymptotic problems of parameter estimation in stochastic PDEs: the discrete time sampling. *Methods Math. Statist.* **6**, 200–223.

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

[20] Safarov, Yu. and Vassiliev, D. (1997). *The asymptotic distribution of eigenvalues of partial differential operators.* AMS Transl., **155**.

[21] Walsh, J. (1984). An introduction to stochastic partial differential equations, Lecture Notes in Mathematics, **1180**, 265–439.

Appendix

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))

$$A = \sum_{|\alpha|,|\beta|\le n} D^\alpha(a^{\alpha\beta}D^\beta),\ a^{\alpha\beta} \in C_b^\infty(G), \tag{4.4}$$

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

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

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

The function

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

where $\xi^\alpha = \xi_1^{\alpha_1}\dots\xi_d^{\alpha_d}$, 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

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

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 (4.4) 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

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

Department of Statistics and Probability
Michigan State University
East Lansing, MI 48824

huebner@stt.msu.edu

Department of Mathematics, Room 2–267
Massachusetts Institute of Technology
77 Mass. Av.
Cambrige, MA 02140

lototsky@math.mit.edu