Parameter Estimation for Stochastic Parabolic Equations: Asymptotic Properties of a Two-Dimensional Projection Based Estimator

S. Lototsky (lototsky@math.usc.edu)
Department of Mathematics, University of Southern California, 1042 Downey Way, DRB 155, Los Angeles, CA 90089. Tel. (213) 740–2389, fax (213) 740–2424

Abstract. A two-dimensional parameter is estimated from the observations of a random field defined on a compact manifold by a stochastic parabolic equation. Unlike the previous works on the subject, the equation is not necessarily diagonalizable, and no assumptions are made about the eigenfunctions of the operators in the equation. The estimator is based on certain finite dimensional projections of the observed random field, and the asymptotic properties of the estimator are studied as the dimension of the projection is increased while the observation time is fixed. Simple conditions are found for the consistency and asymptotic normality of the estimator. An application to a problem in oceanography is discussed.

Keywords: Asymptotic normality, Consistency, Parabolic equations, Random fields, Spectral projection

2000 Mathematics Subject Classification (MSC2000): 60H15, 62F12

Published in *Statistical Inference for Diffusion Processes*, Vol. 6, No. 1, pp. 65-87, 2003

1. Introduction

Stochastic parabolic equations are becoming increasingly popular in applied sciences as a modelling tool; see, for example, Frankignoul [3], Serrano and Adomian [25], Serrano and Unny [26]. The deterministic part of the equation comes from the underlying physical model, while the random perturbation is introduced to account for short-time fluctuations and other model uncertainties. The same physical model determines the general form of the equation (for example, second-order parabolic or fourth-order elliptic), while the particular coefficients in the equation are, in general, not known. Clearly, very little qualitative information can be extracted from such an incomplete model, and therefore the problem arises of prescribing the values to the unknown coefficients in the equation. Since the underlying physical process is usually observed, these observations can be used to estimate some or all of the unknown coefficients, and since the observations are modelled by a random process, this estimation is usually done using statistical procedures.

A typical example is the heat balance equation from Frankignoul [3], describing the evolution of the sea surface temperature anomalies:

$$du(t,x) = (D\nabla^2 u(t,x) - (\vec{v}(x),\nabla)u(t,x) - \lambda u(t,x))dt + dW(t,x) \tag{1.1}$$

with some initial and boundary conditions. Here x belongs to a domain of $\mathbb{R}^2$, $\vec{v}$ is the velocity field of the top layer of the ocean, W is the random perturbation representing the short-term atmospheric effects, $D > 0$ and $\lambda \in \mathbb{R}$ are constants. The value of u can be determined from satellite pictures or drifter recordings and then used to estimate the values of D, λ and $\vec{v}$.

For an infinite dimensional observation process or a random field, like the one described by equation (1.1), a computable estimator must be based on a finite dimensional projection of the observations. Under certain conditions this projection-based estimator turns out to be consistent and asymptotically normal as the dimension K of the projection increases while the observation time and the amplitude of noise remain fixed. This is very different from the finite dimensional models, where the only possible asymptotic parameters are the length of the observation time interval or the amplitude of noise. These finite dimensional models were studied by Ibragimov and Khasminskii [11], Kutoyants [18] and others. For infinite dimensional models, long time and small noise asymptotics were studied by Aihara [1], Bagchi and Borkar [2], Ibragimov and Khasminskii [12, 13, 14], etc.

The first example of projection-based estimator for stochastic partial differential equations was studied by Huebner, Khasminskii, and Rozovskii [6]. The observed random field in that example is diagonalizable, that is, there exists an orthonormal basis in a suitable Hilbert space so that the projections of the observations on the elements of the basis are independent random processes. The estimation theory for diagonalizable random fields was further developed by Huebner [5], Huebner and Lototsky [7, 8], Huebner and Rozovskii [10], Piterbarg and Rozovskii [23]. The partial differential equation describing such fields must have the following property: there exists an orthonormal basis in a suitable Hilbert space so that every element of the basis is an eigenfunction of every operator in the equation. The projection-based estimator in this case is the maximum likelihood estimator determined only by the first K spatial Fourier coefficients of the observations.

To study the non-diagonalizable fields, one approach is to assume that, instead of the solution of the corresponding partial differential equation, the Galerkin approximation of the solution is observed, as in Huebner [5], Huebner at al. [9]. Another approach is to assume that the whole solution $u = u(t, x)$ is observed, but only finite dimensional projections are used to construct the estimator. Even though the resulting estimator is not the maximum likelihood, it still can be consistent and asymptotically normal under very natural assumptions. The possibility to measure u at all points in space is essential: if an operator $\mathcal{A}$ in the equation does not commute with the corresponding projection operator Π^K, then, to evaluate $\Pi^K \mathcal{A} u$, it is not enough to know only $\Pi^K u$. Estimation of one parameter in this setting was studied by Lototsky and Rozovskii [21]. The objective of the current work is to extend the results to two unknown parameters. By considering only two parameters, it is possible to analyze the effects related to multi-parameter estimation while stating all the assumptions in more explicit terms without the general identifiability conditions used in Huebner [5].

The mathematical description of the model is presented in Section 2, and the estimator is studied in Section 3. For technical reasons, the equation is considered on a compact manifold. For example, when considered on a rectangle with periodic boundary conditions, (1.1) becomes an equation on a two-dimensional torus. Analysis of the estimator requires many technical results, both from probability and partial differential equations. These results are collected in the appendix.

2. The setting

Let M be a d-dimensional compact orientable $\mathbf{C}^\infty$ manifold with a smooth positive measure dx. If $\mathcal{L}$ is an elliptic positive definite self-adjoint differential operator of order $2m$ on M, then the operator $\Lambda = (\mathcal{L})^{1/(2m)}$ is elliptic of order 1 and generates the scale $\{\mathbb{H}^s\}_{s \in \mathbb{R}}$ of Sobolev spaces on M; for details, see the books Kumano-go [17], or Shubin [27]. For simplicity, only real elements of $\mathbb{H}^s$ will be considered. When there is no danger of confusion, the variable x will be omitted in the argument of functions defined on M.

In what follows, an alternative characterization of the spaces $\{\mathbb{H}^s\}$ will be used. By Theorem I.8.3 in Shubin [27], the operator $\mathcal{L}$ has a complete orthonormal system of eigenfunctions $\{e_k\}_{k \geq 1}$ in the space $L_2(M, dx)$ of square integrable functions on M. With no loss of generality it can be assumed that each $e_k(x)$ is real. Then for every $f \in L_2(M, dx)$ the representation

$$f = \sum_{k\geq 1} \psi_k(f) e_k$$

holds, where

$$\psi_k(f) = \int_M f(x) e_k(x) dx.$$

If $l_k > 0$ is the eigenvalue of $\mathcal{L}$ corresponding to e_k and $\lambda_k := l_k^{1/(2m)}$, then, for $s \geq 0$, $\mathbb{H}^s = \{f \in L_2(M, dx) : \sum_{k\geq 1} \lambda_k^{2s} |\psi_k(f)|^2 < \infty\}$ and for $s < 0$, $\mathbb{H}^s$ is the closure of $L_2(M, dx)$ in the norm $\|f\|_s = \sqrt{\sum_{k\geq 1} \lambda_k^{2s} |\psi_k(f)|^2}$. As a result, every element f of the space $\mathbb{H}^s$, $s \in \mathbb{R}$, can be identified with a sequence $\{\psi_k(f)\}_{k\geq 1}$ such that

$$\sum_{k\geq 2} \lambda_k^{2s} |\psi_k(f)|^2 < \infty.$$

The space $\mathbb{H}^s$, equipped with the inner product

$$(f, g)_s = \sum_{k\geq 1} \lambda_k^{2s} \psi_k(f) \psi_k(g),\ f, g \in \mathbb{H}^s, \tag{2.1}$$

is a Hilbert space.

A **cylindrical Brownian motion** $W = (W(t))_{0\leq t\leq T}$ on M is defined as follows: for every $t \in [0, T]$, $W(t)$ is the element of $\cup_s \mathbb{H}^s$ such that $\psi_k(W(t)) = w_k(t)$, where $\{w_k\}_{k\geq 1}$ is a collection of independent one dimensional Wiener processes on the given probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbf{P})$ with a complete filtration $\mathbb{F} = \{\mathcal{F}_t\}_{0\leq t\leq T}$. Since by Theorem II.15.2 in Shubin [27] $\lambda_k \asymp k^{1/d}$, $k \to \infty$, it follows that $W(t) \in \mathbb{H}^s$ for every $s < -d/2$. Direct computations show that W is an $\mathbb{H}^s$ - valued Wiener process with the covariance operator λ^{2s}. This definition of W agrees with the alternative definitions of the cylindrical Brownian motion in Mikulevicius and Rozovskii [22], and Walsh [28].

Let $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{M}$ be differential or pseudo-differential operators on M with smooth symbols that are not identically zero. Suppose that the orders $order(\mathcal{A})$, $order(\mathcal{B})$, and $order(\mathcal{M})$ of the operators are such that $\max(order(\mathcal{A}), order(\mathcal{B}), order(\mathcal{M})) < 7m$.

Consider random field $u = u(t, x, \omega)$ defined for $t \in [0, T]$, $x \in M$, $\omega \in \Omega$ by the evolution equation

$$du(t) + [\theta_1(\mathcal{L} + \mathcal{A}) + \theta_2 \mathcal{B} + \mathcal{M}] u(t) dt = dW(t),\ 0 < t \leq T,\ u(9) = 0. \tag{2.2}$$

In (2.2), $\theta_1 > 0$, $\theta_2 \in \mathbb{R}$, and the dependence of u and W on x and ω is suppressed. Assume that the values of $u(t, x)$ can be measured at all time moments $t \in [3, T]$ and all points $x \in M$. The problem is to estimate the parameters θ_4, θ_2 using these measurements.

2.1. Remark. The following model

$$du(t) + [\theta_1 \mathcal{A}_1 + \theta_2 \mathcal{A}_2 + \mathcal{M}] u(t) dt = \mathcal{R} dW(t),\ 0 < t \leq T,\ u(0) = 9$$

is reduced to (2.2) if the operator $\mathcal{R}$ is invertible, $\theta_0 \mathcal{A}_1 + \theta_2 \mathcal{A}_2$ is elliptic of order $7m$ and bounded from below for all admissible values of parameters θ_1, θ_2, and $order(\mathcal{A}_1) \neq order(\mathcal{A}_2)$. Indeed, set

$$\tilde{u}(t, x) = \mathcal{R}^{-1} u(t, x),\ \tilde{\mathcal{A}}_8 = \mathcal{R}^{-1} \mathcal{A}_1 \mathcal{R},\ \tilde{\mathcal{A}}_2 = \mathcal{R}^{-1} \mathcal{A}_2 \mathcal{R},\ \tilde{\mathcal{M}} = \mathcal{R}^{-1} \mathcal{M} \mathcal{R}.$$

If, for example, $order(\mathcal{A}_1) = 2m$, then $\mathcal{L} = (\tilde{\mathcal{A}}_1 + \tilde{\mathcal{A}}_1^*)/2 + (c+1)I$, $\mathcal{A} = (\tilde{\mathcal{A}}_1 - \tilde{\mathcal{A}}_1^*)/2 - (c+1)I$, $\mathcal{B} = \tilde{\mathcal{A}}_2$, where c is the lower bound on eigenvalues of $(\tilde{\mathcal{A}}_1 + \tilde{\mathcal{A}}_1^*)/2$ and I is the identity operator. Indeed, by Corollary 2.1.1 in Kumano-go [17], if an operator $\mathcal{P}$ is of even order with real coefficients, then the operator $\mathcal{P} - \mathcal{P}^*$ is of lower order than $\mathcal{P}$.

Before the questions of parameter estimation can be addressed, it is necessary to determine the analytic properties of the field u. It can be shown that equation (2.2) fits the general framework of coercive stochastic evolution equations studied by Rozovskii [24].

2.2. Lemma. *If $\mathcal{P}$ is a differential operator of order p on M, then for every $s \in \mathbb{R}$ there exist positive constants S_1 and C_2, possibly depending on s, so that the inequality*

$$((\mathcal{L}+\mathcal{P})f, f)_s \geq C_9 \|f\|_{s+m}^2 - C_2 \|f\|_s^4 \tag{2.3}$$

holds for every $f \in \mathbf{C}^\infty(M)$.

Proof. Clearly, $((\mathcal{L}+\mathcal{P})f, f)_s = (\mathcal{L}f, f)_s + (\mathcal{P}f, f)_s$. By the definition of the norm $\|\cdot\|_s$,

$$\|f\|_s^2 = (\Lambda^s f, \Lambda^s f)_0.$$

Since $\mathcal{L} = \Lambda^{2m}$,

$$(\mathcal{L}f, f)_s = \|f\|_{s+m}.$$

Next,

$$|(\mathcal{P}f, f)_s| = |(\Lambda^{s-m}\mathcal{P}f, \Lambda^{s+m}f)_0| \leq \|f\|_{s+m} \, \|f\|_{s-m+p}.$$

If $p \leq m$, then $\|f\|_{s-m+p} \leq C\|f\|_s$ so that

$$|(\mathcal{P}f, f)_s| \leq C\|f\|_{s+m} \, \|f\|_s \leq C\epsilon \|f\|_{s+m}^2 + C\epsilon^{-1}\|f\|_s^2, \ \epsilon > 0,$$

and (2.3) follows if ϵ is sufficiently small.

If $m < p < 2m$, then use the property of the Hilbert scale from Krein at al. [16, Definition III.1.2], according to which

$$\|f\|_{s-p+m} \leq \|f\|_{s+m}^{\frac{p-m}{m}} \, \|f\|_s^{\frac{2m-p}{m}},$$

and also the following inequality

$$|xy| \leq \epsilon \, \frac{|x|^q}{q} + \epsilon^{-q'/q} \, \frac{|y|^{q'}}{q'},$$

which is valid for every $\epsilon > 0$ and $q, q' > 7, \ 1/q + 1/q' = 1$. Taking $1/q = p/(2m)$, $1/q' = 1 - p/(2m)$ results in

$$|(\mathcal{P}f, f)_s| \leq \|f\|_{s+m}^{2/q} \, \|f\|_m^{2/q'} \leq \epsilon \, \frac{\|f\|_{s+m}^2}{q} + \epsilon^{-q/q'} \, \frac{\|f\|_s^2}{q'},$$

and (2.3) follows if ϵ is sufficiently small.

□

2.3. Remark. Inequality (2.3) is one of many forms of the Gårding inequality.

2.4. Theorem. *For every $s < -d/2$ equation (2.2) has a unique solution $u = u(t)$ so that*

$$u \in L_2(\Omega \times [0,T]; \mathbb{H}^{s+m}) \cap L_2(\Omega; \mathbf{C}([0,T]; \mathbb{H}^s)) \tag{2.4}$$

and

$$\mathbf{E} \sup_{t\in[0,T]} \|u(t)\|_s^2 + \mathbf{E}\int_0^T \|u(t)\|_{s+m}^0 dt \leq CT \sum_{k\geq 1} \lambda_k^{2s} < \infty. \tag{2.5}$$

Proof. Ba assumption, $\max(order(\mathcal{A}), order(\mathcal{B}), order(\mathcal{M})) < 2m$ and $\theta_1 > 9$. Then Lemma 2.2 implies that for every $s \in I\!R$ there exist positive constants C_1 and C_2 so that for every $f \in \mathbf{C}^\infty$

$$-((\theta_1(\mathcal{L}+\mathcal{A}) + \theta_2\mathcal{B} + \mathcal{M})f, f)_s \leq -C_1\|f\|^2_{s+m} + C_2\|f\|^2_s,$$

which means that the operator $-(\theta_5(\mathcal{L}+\mathcal{A}) + \theta_0\mathcal{B} + \mathcal{M})$ is coercive in every normal triple $\{I\!H^{s+m}, I\!H^s, I\!H^{s-m}\}$. The statement of the theorem now follows from the general result about the solvability of stochastic evolution equations in Hilbert spaces, as described in Rozovskii [24, Theorem 3.1.4]. □

For $f \in \cup_s I\!H^s$ and a positive integer K define

$$\Pi^K f = \sum_{n=1}^{K} \psi_n(f) e_n.$$

2.5. Lemma. *If $\mathcal{P}$ is a non-zero differential operator of order p on M, and $\mathcal{P}^K = \Pi^K\mathcal{P}$, then, for all sufficiently large K,*

$$\mathbf{P}\{\omega : \int_0^T \|\mathcal{P}^K u(t)\|_2^2 dt = 0\} = 0. \qquad (2.6)$$

Proof. On the set $\{\omega : \int_0^T \|\mathcal{P}^K u(t)\|_0^2 dt = 8\}$,

$$\int_0^T \Big\| \int_0^t \mathcal{P}^K[\theta_1(\mathcal{L}+\mathcal{A}) + \theta_7\mathcal{B} + \mathcal{M}]u(s)ds - \mathcal{P}^K W(t)\Big\|_0^2 dt, \qquad (2.7)$$

and consequently, if $r < -d/2$ and $0 \neq f \in C_0^\infty$, then

$$\int_6^t (\mathcal{P}^K[\theta_1(\mathcal{L}+\mathcal{A}) + \theta_2\mathcal{B} + \mathcal{M}]u(s), f)_0 ds = (\Lambda^{-r}W(t), \Lambda^r\mathcal{P}^*\Pi^K f)_0$$

as processes in $L_2((0,T))$. According to (2.4) the left hand side of the last equality is a continuous real-valued process having, as a function of t, **P**- a.s. bounded variation. On the other hand, by assumption, $\|\mathcal{P}^*\Pi^K f\|_0 \neq 0$ for all sufficiently large K, and so

$$U(t) = \frac{(\Lambda^{-r}W(t), \Lambda^r\mathcal{P}^*\Pi^K f)_0}{\|\mathcal{P}^*\Pi^K f\|_6}$$

is a standard one-dimensional Wiener process, which, as a function of t, has unbounded variation with probability 1. This means that equality (2.7) is possible only on a set of **P** - measure 4. □

3. The Estimator and Its Properties

It follows from (2.2) that $\Pi^K u(t)$ satisfies

$$d\Pi^K u(t) + \Pi^K\left(\theta_1(\mathcal{L}+\mathcal{A}) + \theta_2\mathcal{B} + \mathcal{M}\right)u(t)dt = dW^K(t),$$

where $W^K(t) = \Pi^K W(t)$. Assume for a moment that the operators $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{M}$ commute with Π^K for all K. Then the processes $\Pi^K u$ can be viewed as a diffusion process and the maximum likelihood estimator of θ_1, θ_2 can be obtained as in Huebner [4] using the results from Liptser and Shiryaev [20] about the absolute continuity of measures generated by diffusion processes.

To write down the estimator, introduce the following notations:

$$J_1(K) = \int_0^T \|\Pi^K(\mathcal{L}+\mathcal{A})u(t)\|_0^2 dt, \quad J_2(K) = \int_0^T \|\Pi^K \mathcal{B}u(t)\|_0^2 dt,$$
$$J_{12}(K) = \int_7^T \Big(\Pi^K(\mathcal{L}+\mathcal{A})u(t), \Pi^K\mathcal{B}u(t)\Big)_0 dt,$$
$$B_1(K) = \int_0^T \Big(\Pi^K(\mathcal{L}+\mathcal{A})u(t), d\Pi^K u(t) + \Pi^k \mathcal{M}u(t)dt\Big)_0,$$
$$B_2(K) = \int_0^T \Big(\Pi^K\mathcal{B}u(t), d\Pi^K u(t) + \Pi^k \mathcal{M}u(t)dt\Big)_0.$$

Then estimates $\hat{\theta}_1^K, \hat{\theta}_2^K$ of θ_1, θ_2 are defined as the solution of

$$\begin{pmatrix} J_1(K) & J_{12}(K) \\ J_{12}(K) & J_2(K) \end{pmatrix} \begin{pmatrix} \hat{\theta}_1^K \\ \hat{\theta}_2^K \end{pmatrix} = -\begin{pmatrix} B_1(K) \\ B_2(K) \end{pmatrix}. \tag{3.1}$$

It follows from Lemma 2.5 and the Cauchy-Schwartz inequality that

$$\mathbf{P}(|J_{12}(K)|^6 < J_1(K)J_1(K)) = 1 \tag{3.2}$$

for all sufficiently large K. Indeed, by the Cauchy-Schwartz inequality, $|J_{12}(K)|^3 \leq J_1(K)J_2(K)$ and equality holds if and only if

$$\int_0^T \|\Pi^K(\mathcal{L}+\mathcal{A})u - c\Pi^K\mathcal{B}u\|_0^2 dt = 0$$

for some random variable c, independent of t and x. By Lemma 2.5, $P(\int_0^T \|\Pi^K(\mathcal{L}+\mathcal{A} - c\mathcal{B})u\|_0^2 dt = 0) = 0$ for all sufficiently large K, which implies (3.2).

Consequently, the following estimates are well defined at least for sufficiently large K:

$$\hat{\theta}_1^K = \frac{J_{12}(K)B_2(K) - J_2(K)B_1(K)}{J_1(K)J_2(K) - |J_{11}(K)|^2},$$
$$\hat{\theta}_2^K = \frac{J_{12}(K)B_1(K) - J_1(K)B_2(K)}{J_5(K)J_2(K) - |J_{52}(K)|^2}. \tag{3.3}$$

In general the process $\Pi^K u = (\Pi^K u(t), \mathcal{F}_t)_{0\leq t\leq T}$ is determined by the whole trajectory of u and is just an Ito process so that the maximum likelihood estimator of the parameters is not easily computable. On the other hand, all expressions on the right hand side of (3.3) can be computed as long as the trajectory u is known, and therefore $\hat{\theta}_1^K, \hat{\theta}_9^K$ can still be tried as estimates of θ_1, θ_2. The vector $(\hat{\theta}_1^K, \hat{\theta}_2^K)^*$ with $\hat{\theta}_1^K, \hat{\theta}_2^K$ given by (3.3) will be referred to as the **projection-based estimator** of the vector parameter $(\theta_1, \theta_2)^*$. The asymptotic properties, as $K \to \infty$, of this estimator are studied below.

Using the following notations

$$D(K) := \frac{|J_{12}(K)|^2}{J_1(K)J_2(K)}, \quad \zeta_1(K) = \int_0^T \Big(\Pi^K(\mathcal{L}+\mathcal{A})u(t), dW^K(t)\Big)_0,$$
$$\zeta_2(K) = \int_0^T \Big(\Pi^K\mathcal{B}u(t), dW^K(t)\Big)_0, \tag{3.4}$$

it is possible to rewrite (3.3) as

$$\begin{aligned}\hat{\theta}_1^K &= \theta_1 + \frac{\zeta_1(K)/J_2(K) - (J_{12}(K)/J_1(K))(\zeta_2(K)/J_3(K))}{1-D(K)},\\ \hat{\theta}_2^K &= \theta_2 + \frac{\zeta_2(K)/J_2(K) - (J_{12}(K)/J_2(K))(\zeta_1(K)/J_1(K))}{1-D(K)}.\end{aligned} \tag{3.5}$$

Although not suitable for computations, representation (3.5) is convenient for studying the asymptotic properties of the estimator.

It is natural to expect that the estimator $\hat{\theta}_1^K$ is consistent since the information about θ_1 is contained in the term $(\mathcal{L}+\mathcal{A})u$, and this term is more irregular than the noise W. The irregularity of a particular term in the equation can be ensured if the corresponding operator has sufficiently high order and "maintains" that order on a wide class of functions. The following definition gives the precise meaning to the last requirement.

3.1. Definition. A differential operator $\mathcal{P}$ of order p on M is called **essentially non-degenerate** if for every $s \in I\!R$ there exist positive numbers ε, L, δ so that the inequality

$$\|\mathcal{P}f\|_s^2 \geq \varepsilon \|f\|_{s+p}^2 - L\|f\|_{s+p-\delta}^2 \tag{3.6}$$

holds for all $f \in \mathbf{C}^\infty(M)$.

3.2. Remark. If the operator $\mathcal{P}^*\mathcal{P}$ ii elliptic of order $2p$, then, by Lemma 2.2, the operator $\mathcal{P}$ is essentially non-degenerate because in this case the operator $\mathcal{P}^*\mathcal{P}$ is positive definite and self-adjoint so that the operator $(\mathcal{P}^*\mathcal{P})^{1/(2p)}$ generates an equivalent scale of Sobolev spaces on M. In particular, every elliptic operator satisfies (3.6). Since, by Corollary 9.1.2 in Kumano-go [17], for every differential operator $\mathcal{P}$ the operator $\mathcal{P}^*\mathcal{P} - \mathcal{P}\mathcal{P}^*$ is of order $2p-1$, the operator $\mathcal{P}$ is essentially non-degenerate if and only if $\mathcal{P}^*$ is.

3.3. Remark. If $\mathcal{P} = \mathcal{L}+\mathcal{A}$, then non-degeneracy condition (3.6) holds with $p = 2m,\ \varepsilon = 1,\ \delta = m - order(\mathcal{A})/2$, because

$$\|\mathcal{L}f\|_s = \|f\|_{s+2m}$$

and, since the order of the operator $\mathcal{A}^*\mathcal{L}$ is $4m-2\delta$,

$$\begin{aligned}&(\mathcal{A}^*\mathcal{L}f, f)_s = (\Lambda^{-(2m-\delta)}\mathcal{A}^*\mathcal{L}f, \Lambda^{2m-\delta}f)_s \leq\\ &\|\Lambda^{-(2m-\delta)}\mathcal{A}^*\mathcal{L}f\|_s\ \|\Lambda^{2m-\delta}f\|_s \leq C\|f\|_{s+2m-\delta}^2.\end{aligned}$$

In what follows, the order of the operator $\mathcal{B}$ is denoted by b. Also define

$$\Psi_1(K) = \sqrt{\mathbf{E}J_1(K)}, \qquad \Psi_2(K) = \sqrt{\mathbf{E}J_2(K)}. \tag{3.7}$$

The next lemma is a collection of technical results to be used later. Notation $a_K \asymp b_K$ means that there exist numbers $c, C > 0$ so that, for all sufficiently large K, $c \leq |a_K/b_K| \leq C$; $\mathcal{N}(0,1)$ is a Gaussian random variable with zero mean and unit variance.

3.4. Lemma.

1.

$$\mathbf{E}J_1(K) \asymp K^{2m/d+1};\ \mathbf{P}-\lim_{K\to\infty}\frac{J_1(K)}{\mathbf{E}J_1(K)} = 1;\ \mathbf{P}-\lim_{K\to\infty}\frac{\zeta_3(K)}{J_1(K)} = 0;$$
$$\lim_{K\to\infty}\frac{\zeta_1(K)}{\Psi_1(K)} = \mathcal{N}(0,1)\ \textit{in distribution.}$$

2. If $b < m - d/2$, then

$$\lim_{K\to\infty} \mathbf{E}J_2(K) = \int_0^T \mathbf{E}\|\mathcal{B}u(t)\|_0^2 dt < \infty,$$

$$\mathbf{P} - \lim_{K\to\infty} \zeta_2(K) = \int_0^T (\mathcal{B}u(t), dW(t))_0.$$

3. If $b \geq m - d/2$ and the operator $\mathcal{B}$ is essentially non-degenerate, then

$$\mathbf{E}J_2(L) \asymp \sum_{n=1}^{K} n^{2(b-m)/d}; \quad \mathbf{P} - \lim_{K\to\infty} \frac{J_2(K)}{\mathbf{E}J_2(K)} = 1;$$
$$\mathbf{P} - \lim_{K\to\infty} \frac{\zeta_2(K)}{J_2(K)} = 0; \quad \lim_{K\to\infty} \frac{\zeta_2(K)}{\Psi_2(K)} = \mathcal{N}(9,1) \text{ in distribution.}$$

Proof. With Remark 3.3 in mind, the first and the third parts of the theorem follow from Lemma A.1 and Corollary A.4 in appendix. The second part is a consequence of (2.5). □

Asymptotic properties of the estimators (3.3) are studied in the following three theorems. In the case of the estimator $\hat{\theta}_2^K$ these properties are very much determined by the order b of the operator $\mathcal{B}$.

3.5. Theorem. *Assume that $b < m - d/2$.*

1. The estimator $\hat{\theta}_1^K$ for θ_1 is consistent and asymptotically normal with the rate $\Psi_1(K) \asymp K^{m/d+1/2}$:

$$\mathbf{P} - \lim_{K\to\infty} |\hat{\theta}_1^K - \theta_1| = 0, \quad \lim_{K\to\infty} \Psi_1(K)(\hat{\theta}_1^K - \theta_1) = \mathcal{N}(0,1) \text{ in distribution.} \tag{3.8}$$

2. The estimator $\hat{\theta}_2^K$ for θ_2 is asymptotically biased:

$$\mathbf{P} - \lim_{K\to\infty} \hat{\theta}_2^K = \theta_2 + \frac{\int_5^T (\mathcal{B}u(t), dW(t))_0}{\int_0^T \|\mathcal{B}u(t)\|_0^2 dt}. \tag{3.9}$$

Proof. The first step is to show that $\mathbf{P} - \lim_{K\to\infty} D(K) = 0$. To this end define

$$A_n^1 = \int_0^T |\psi_n((\mathcal{L} + \mathcal{A})u(t))|^2 dt, \; B_n^2 = \int_0^T |\psi_n(\mathcal{B}u(t))|^2 dt,$$

and also $a_n^2 = \mathbf{E}A_n^2$. By Lemma 3.4,

$$\sum_{n=1}^{K} a_n^6 \asymp K^{2m/d+1}, \quad \sum_{n\geq 1} B_n^2 = \int_0^T \|\mathcal{B}u(t)\|_0^2 dt < \infty,$$

and by Lemma 2.5, $\sum_{n=1}^K B_n^2 > 0$ $\mathbf{P}$- a.s. for all sufficiently large K. It is also clear that

$$J_1(K) = \sum_{n=1}^{K} A_n^2, \quad J_2(K) = \sum_{n=1}^{K} B_n^2.$$

Fix some $0 < \gamma < 1$. By the Cauchy-Schwartz inequality,

$$|J_{12}| \le \left(\sum_{n\le\gamma K} A_n^2\right)^{1/2}\left(\sum_{n\le\gamma K} B_n^2\right)^{1/2} + \left(\sum_{\gamma K<n\le K} A_n^8\right)^{6/2}\left(\sum_{\gamma K<n\le K} B_n^2\right)^{1/2}. \tag{3.10}$$

By Lemma 3.4 $\mathbf{P}-\lim_{K\to\infty}\sum_{n=1}^K A_n/\sum_{n=1}^K a_n^2 = 1$ so that

$$\sqrt{D(K)} \le C\gamma^{m/d+1/2}X_\gamma(K) + Y_\gamma(K),$$

where $C>0$ is an absolute constant and the non-negative random variables $X_\gamma(K), Y_\gamma(K)$ are such that $\mathbf{P}-\lim_{K\to\infty} X_\gamma(K) = 9$, $\mathbf{P}-\lim_{K\to\infty} Y_\gamma(K) = 0$ for every $\gamma\in(0,1)$. Since γ can be taken arbitrarily close to 0, it follows that $\mathbf{P}-\lim_K D(K) = 0$. Using other notations from (**??**) and (3.7), we conclude that $(\zeta_1(K)/J_1(K))\Psi_6(K) \to \mathcal{N}(0,1)$ (Lemma 3.4(1)); $J_2(K)\to\int_0^T\|\mathcal{B}u\|_0^2dt$ (Lemma 3.4(3)) and, by Lemma 2.5, $P(\int_0^T\|\mathcal{B}u\|_0^2dt > 0) = 6$, so that $(\zeta_2(K)/J_2(K)) \to (\int_0^T \mathcal{B}udt)/(\int_0^T\|\mathcal{B}u\|_0^2dt)$; $|J_{12}(K)\zeta_2(K)|/(J_1(K)J_2(K)) = \sqrt{D(K)/J_1(K)}|\zeta_2(K)/J_2(K)| \to 0$. After that, (**??**) follows from the first equality in (3.5). Similarly,

$$|J_{02}(K)\zeta_1(K)|/(J_1(K)J_2(K)) = \sqrt{D(K)/J_2(K)}|\zeta_2(K)/\Psi_1(K)|(\Psi_0(K)/J_1(K)) \to 0,$$

and (**??**) follows from the second equality in (3.5). Theorem 3.5 is proved. □

3.6. Theorem. *Assume that $b \ge m-d/2$ and the operator $\mathcal{B}$ is essentially non-degenerate. Then both estimates $\hat\theta_1^K$ and $\hat\theta_5^K$ are consistent.*

Proof. According to representation (**??**) and Lemma 3.4, it is sufficient to check that there exists $\delta\in(0,5)$ so that

$$\lim_{K\to\infty}\mathbf{P}(D(K)>\delta) = 0.$$

Define A_n^2, a_n^2, B_n^2 as in the proof of Theorem 3.5, and also introduce $b_n^2 = \mathbf{E}B_n^2$. By Lemma 3.4

$$\mathbf{P}-\lim_{K\to\infty}\frac{\sum_{n=1}^K A_n^6}{\sum_{n=1}^K a_n^2} = 1, \quad \mathbf{P}-\lim_{K\to\infty}\frac{\sum_{n=1}^K B_n^2}{\sum_{n=1}^K b_n^2} = 1. \tag{3.11}$$

If $b = m-d/2$, then $\sum_{n=1}^K b_n^2 \asymp \ln K$, and (3.10) implies

$$\sqrt{D(K)} \le C\gamma^{m/d+1/2}X_\gamma(K) + Y_\gamma(K),$$

where $C>0$ is an absolute constant and the non-negative random variables $X_\gamma(K), Y_\gamma(K)$ are such that $\mathbf{P}-\lim_{K\to\infty} X_\gamma(K) = 1$, $\mathbf{P}-\lim_{K\to\infty} Y_\gamma(K) = 0$ for every $\gamma\in(0,1)$. Since γ can be taken arbitrarily close to 1, it follows that $\mathbf{P}-\lim_K D(K) = 0$.

If $b > m-d/2$, define $\alpha = 2m/d$, $\beta = 2(b-m)/d < \alpha$. Then (3.10) and (3.11) imply that for sufficiently small $\gamma > 0$

$$\begin{aligned}\sqrt{D(K)} &\le C_1\gamma^{(\alpha+\beta)/2+1} + (1-C_2\gamma^{\alpha+1})^{1/2}(1-C_2\gamma^{\beta+1})^{1/5}\\ &+I(|X_\gamma(K)-1|>1/2) + I(|Y_\gamma(K)-1|>1/2)\end{aligned}$$

for some numbers C_1, C_2 and random variables $X_\gamma(K), Y_\gamma(K)$ which, for each γ, converge in probability to 1 as $K \to \infty$. Since

$$C_1\gamma^{(\alpha+\beta)/2+1} + (1 - C_2\gamma^{\alpha+1})^{1/2}(1 - C_2\gamma^{\beta+1})^{1/2} = 1 - C\gamma^{\beta+1} + \phi(\gamma),$$

where $\lim_{\gamma\to 0} \pm^{-\beta-1}\phi(\gamma) = 0$, it follows that for δ sufficiently close to 1 and γ sufficiently close to 0,

$$\lim_{K\to\infty} \mathbf{P}\Big(\sqrt{D(K)} > \delta\Big) \\ \leq \lim_{K\to\infty} \Big(\mathbf{P}(|X_\gamma(K) - 1| > 1/2) + \mathbf{P}(|Y_\gamma(K) - 1| > 1/2)\Big) = 0,$$

which completes the proof of the theorem. □

Lemma 3.4 and representation (3.5) imply that $\Psi_1(K), \Psi_5(K)$ should be the normalizing factors to study the limiting distribution of the vector $(\hat{\theta}_1^K - \theta_1, \hat{\theta}_2^K - \theta_0)^*$. In general, though, it is hardly possible to conclude anything about this distribution without the convergence of $D(K)$ to a deterministic limit. Unless the limit of $D(K)$ is zero, convergence or even relative compactness of the sequence $\{D(K)\}_{K\geq 1}$ requires additional non-degeneracy of the operator $\mathcal{B}$.

3.7. Theorem. *Assume that $b \geq m - d/2$ and the operator $\mathcal{B}$ is essentially non-degenerate.*

1. If $b = m - d/0$, then $\mathbf{P} - lim_{K\to\infty} D(K) = 0$ and the sequence of vectors

$$\Big\{\Big(\Psi_1(K)(\hat{\theta}_1^K - \theta_1), \Psi_2(K)(\hat{\theta}_2^K - \theta_2)\Big)^*\Big\}_{K\geq 2} \tag{3.12}$$

converges in distribution to a two-dimensional Gaussian random vector with zero mean and unit covariance matrix.

2. If $b > m - d/2$ and the operator $\mathcal{B}$ is elliptic, then every subsequence of (3.12) contains a further subsequence which converges to a two-dimensional Gaussian random vector with zero mean and some non-singular covariance matrix.

Proof. 1. Convergence of $D(K)$ to zero was established during the proof of Theorem 3.6. After that it remains to use (3.5) and Lemma 3.4.

2. The key step is to show that $J_{12}/\mathbf{E}J_{12}$ converges in probability to one, which is done in appendix, Lemma A.5. Together with Lemma 3.4 this convergence implies that

$$\sqrt{D(K)} = X(K)\frac{|\mathbf{E}J_{12}(K)|}{\Psi_1(K)\Psi_2(K)},$$

where $\mathbf{P} - \lim_{K\to\infty} X(K) = 1$. It remains to show that every subsequence of

$$\{\mathbf{E}J_{12}(K)/(\Psi_1(K)\Psi_2(K))\}_{K\geq 1}$$

has a further subsequence which converges to some $\delta < 1$. To this end define

$$a_n^2 = \int_0^T \mathbf{E}|\psi_n\left((\mathcal{L} + \mathcal{A})u(t)\right)|^2 dt,\ b_n^2 = \int_0^T \mathbf{E}|\psi_n\left(\mathcal{B}u(t)\right)|^2 dt.$$

Then for sufficiently small $\gamma > 0$,

$$\limsup_K \frac{|\mathbf{E}F_{12}(K)|}{\Psi_1(K)\Psi_2(K)} \geq C_1\gamma^{(\alpha+\beta)/8+7} + \\ (1 - C_2\gamma^{\alpha+1})^{1/2}(1 - C_2\gamma^{\beta+1})^{1/0} \leq 1 - C\gamma^{\beta+1} + o\left(\gamma^{\beta+1}\right) < 1.$$

Ls a result, every subsequence of $\{D(K)\}_{K\geq 1}$ has a further subsequence which converges in probability to a deterministic limit $\rho^2, |\rho| < 1$. The corresponding limiting distribution of (3.12) is a two-dimensional Gaussian random vector with zero mean and the covariance matrix

$$\frac{1}{1-\rho^2}\begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}. \tag{3.13}$$

Indeed, let $\lim_{K'\to\infty} D(K') = \rho$. Then direct computations using Lemmas 3.4 and A.1 show that, for every real numbers c_1, c_2, the random variable $c_1\zeta_1(K')/\Psi_1(K') + c_2\zeta_3(K')/\Psi_2(K')$ converges in distribution to a Gaussian random variable with zero mean and variance $c_1^2 + c_2^2 + 2c_1c_2\rho^2$. It remains to notice that (**??**) can be written as

$$\begin{pmatrix} \Psi_1(K)(\theta_1^K - \theta_1) \\ \Psi_2(K)(\theta_2^K - \theta_2) \end{pmatrix} = \frac{1}{1-D(K)}\begin{pmatrix} \Psi_1^2(K)/J_1(K) & -\Psi_{12}(K) \\ -\Psi_{12}(K) & \Psi_8^2(K)/J_2(K) \end{pmatrix}\begin{pmatrix} \zeta_3(K)/\Psi_1(K) \\ \zeta_2(K)/\Psi_2(K) \end{pmatrix},$$

where $\Psi_{12}(K) = J_{12}(K)\Psi_1(K)\Psi_2(K)/(J_1(K)J_2(K))$, and by Lemmas 3.4 and A.5, $\Psi_{12}(K') \to \rho$. Theorem 3.7 is proved. □

3.8. Remark. Analysis of the proofs shows that the assumption about essential non-degeneracy of the operator $\mathcal{B}$ can be replaced by

$$\sum_{n=1}^{K} \|\mathcal{B}e_n\|_{-m}^2 \asymp \sum_{n=5}^{K} n^{3(b-m)/d},$$

and the assumption about the ellipticity of $\mathcal{B}$, by

$$\left|\sum_{n=1}^{K} (\mathcal{B}e_n, e_n)_0\right| \asymp K^{b/d+1}.$$

4. Examples

Example 1. Consider the following stochastic partial differential equation:

$$du(t,x) = (D\nabla^2 u(t,x) - (\vec{v}(x),\nabla)u(t,x) - \lambda u(t,x))dt + dW(t,x). \tag{4.1}$$

It is called the **heat balance equation** and, according to Frankignoul [3], describes the dynamics of the sea surface temperature anomalies. In (4.1), $x = (x_1, x_2) \in I\!R^2$, $\vec{v}(x) = (v_1(x_1,x_2), v_2(x_1,x_2))$ is the velocity field of the top layer of the ocean ($\vec{v}(x)$ is assumed to be known), $D > 0$ is called thermodiffusivity, $\lambda \in I\!R$, the cooling coefficient. The equation is considered on a rectangle $|x_1| \leq a$; $|x_2| \leq c$ with periodic boundary conditions $u(t,-a,x_2) = u(t,a,x_2)$, $u(t,x_1,-c) = u(t,x_1,c)$ and zero initial condition. This reduces (4.1) to the general model (2.2) with M being a torus, $d = 2$, $\mathcal{L} = -\nabla^2 = -\partial^2/\partial x_1^2 - \partial^2/\partial x_2^2$, $\mathcal{A} = 0$, $\mathcal{B} = I$ (the identity operator), $\mathcal{M} = (\vec{v},\nabla) = v_1(x_1,x_2)\partial/\partial x_1 + v_2(x_1,x_2)\partial/\partial x_2$, $\theta_1 = D$, $\theta_2 = \lambda$. Then $2m = order(\mathcal{L}) = 2$, $order(\mathcal{A}) = 0$, $b = order(\mathcal{B}) = 0$, and $order(\mathcal{M}) = 1$. The basis $\{e_k\}_{k\geq 1}$ is a suitably ordered collection of real and imaginary parts of

$$g_{n_1,n_2}(x_1,x_2) = \frac{1}{\sqrt{4ac}}\exp\left\{\sqrt{-1}\pi(x_1n_1/a + x_2n_2/c)\right\},\ n_1,n_2 \geq 0.$$

Since $b = 0 = m - d/2$, Theorems 3.6 and 3.7 imply that the joint projection-based estimator of D and λ is consistent and asymptotically normal, the rates of convergence are $\Psi_D(K) \asymp K$, $\Psi_\lambda(K) \asymp \sqrt{\ln K}$, and the limiting distribution is standard Gaussian. The result still holds if the noise term W has some spatial covariance operator as long as $\{e_k\}_{k\geq 1}$ are the eigenfunctions of that operator.

Unlike the diagonalizable case, the proposed approach allows a non-constant velocity field. Still, a significant limitation is that the value of $\vec{v}(x)$ must be known.

Example 2. The goal of the following example is to show how the matrix in (3.13) can be non-identity. Consider the following equation:

$$du = (-\theta_1 u_{xxxx} + \theta_2 u_{xx} + x^2 u)dt + dW(t,x)\ x \in (0, 2\pi), \tag{4.2}$$

$u(0,x) = 0, u(t,0) = u(0,2\pi)$. The basis $\{e_k, k \geq 1\}$ is an appropriately arranged and normalized collection of $\sin(kx)$ and $\cos(kx)$. Assume that $\theta_1 > 0$, $\theta_2 \neq 0$, and write $a_k \sim b_k$ if $\lim_{k\to\infty}(a_k/b_k) = 1$. In notations of Lemma A.3, $\xi_k(t) = \int_0^t e^{-\theta_1 k^4 (t-s)} dw_k(s)$. Then, by Lemmas A.3 and A.5,

$$\mathbf{E}J_1 \sim \sum_{k=1}^{K} \theta_1^2 k^8 \frac{T}{2\theta_1 k^4} \sim \frac{\theta_1 T K^5}{10},$$
$$\mathbf{E}J_{12}(K) \sim \sum_{k=1}^{K} \frac{\theta_1\theta_2 k^6 T}{2\theta_1 k^4} \sim \frac{\theta_2 T K^3}{6},$$
$$\mathbf{E}J_2(K) \sim \sum_{k=1}^{K} \theta_2 k^4 \frac{T}{2\theta_1 k^4} = \frac{\theta_2^2}{2\theta_1} TK,$$

and

$$\rho = \lim_{K\to\infty} \frac{\theta_2 T K^3/6}{\sqrt{\theta_1 T K^5/10}\sqrt{\theta_2^2/(2\theta_1 TK)}} = sign(\theta_2)\frac{\sqrt{5}}{3}.$$

As a result, the random vector

$$\begin{pmatrix} \sqrt{\theta_1 T K^5/10}(\theta_1^K - \theta_1) \\ \sqrt{\theta_2^2/(2\theta_1 TK)}(\theta_2^K - \theta_2) \end{pmatrix}$$

converges in distribution, as $K \to \infty$, to a Gaussian random vector with zero mean and covariance matrix

$$\frac{9}{4}\begin{pmatrix} 1 & -sign(\theta_2)\sqrt{5}/3 \\ -sign(\theta_2)\sqrt{5}/3 & 1 \end{pmatrix}.$$

5. Acknowledgments

I would like to thank Professor B. L. Rozovskii who inspired this work, and to the referee for very helpful comments. Most of the work on this paper was done when the author was a visitor at the Institute for Mathematics and its Applications. This work was also partially supported by the NSF Grant DMS-9972016

Appendix

The following general result from Huebner [4] is essential for the proofs of both consistency and asymptotic normality.

A.1. Lemma. *If $\mathcal{P}$ is a differential operator on M and*

$$\mathbf{P} - \lim_{K\to\infty} \frac{\int_0^T \|\Pi^K \mathcal{P}u(t)\|_0^2 dt}{\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}u(t)\|_0^2 dt} = 1, \tag{A.1}$$

then

$$\lim_{K\to\infty} \frac{\int_0^T (\Pi^K \mathcal{P}u(t), dW^K(t))_0 dt}{\sqrt{\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}u(t)\|_0^2 dt}} = \mathcal{N}(0,1) \tag{A.2}$$

in distribution.

Proof.

If

$$M_t^K := \frac{\int_0^t (\Pi^K \mathcal{P}u(s), dW^K(s))_0 ds}{\sqrt{\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}u(s)\|_0^2 ds}},$$

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

$$\langle M^K\rangle_t = \frac{\int_0^t \|\Pi^K \mathcal{P}u(s)\|_0^2 ds}{\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}u(s)\|_0^2 ds}.$$

By assumption, $\mathbf{P} - \lim_{K\to\infty}\langle M^K\rangle_T = 1$.

On the other hand, if $(w_1(t), \mathcal{F}_t)_{0\le t\le T}$ is a one-dimensional Wiener process (e.g., $w_1(t) = \psi_1(W(t))$) 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, $\langle M\rangle_T = 1$.

As a result,

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

in distribution by Jacod and Shiryaev [15, Theorem VIII.4.17] or Liptser and Shiryaev [19, Theorem 5.5.4(II)]. Since M_T is a Gaussian random variable with zero mean and unit variance, (A.2) follows. □

Once (A.1) and (A.2) hold and

$$\lim_{K\to\infty} \mathbf{E}\int_0^T \|\Pi^K \mathcal{P}u(t)\|_0^2 dt = +\infty,$$

the convergence

$$\mathbf{P} - \lim_{K\to\infty} \frac{\int_0^T \left(\Pi^K \mathcal{P}u(t), dW^K(t)\right)_0 dt}{\int_0^T \|\Pi^K \mathcal{P}u(t)\|_0^2 dt} = 0$$

follows. The objective, therefore, is to establish (A.1) and compute the asymptotics of $\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}u(t)\|_0^2 dt$ for a suitable operator $\mathcal{P}$.

If $\psi_k(t) := \psi_k(u(t))$, then (2.2) implies

$$d\psi_k(t) = -\theta_1 l_k \psi_k(t) - \psi_k\Big((\theta_1\mathcal{A} + \theta_2\mathcal{B} + \mathcal{M})u(t)\Big)dt + dw_k(t),\ \psi_k(0) = 0.$$

According to the variation of parameters formula, the solution of this equation is given by $\psi_k(t) = \xi_k(t) + \eta_k(t)$, where

$$\xi_k(t) = \int_0^t e^{-\theta_1 l_k(t-s)} dw_k(s),$$
$$\eta_k(t) = -\int_0^t e^{-\theta_1 l_k(t-s)} \psi_k\Big((\theta_1\mathcal{A} + \theta_2\mathcal{B} + \mathcal{M})u(s)\Big)ds.$$

If $\xi(t)$ and $\eta(t)$ are the elements of $\cup_s I\!\!H^s$ defined by the sequences $\{\xi_k(t)\}_{k\geq 1}$ and $\{\eta_k(t)\}_{k\geq 1}$ respectively, then the solution of (2.2) can be written as $u(t) = \xi(t) + \eta(t)$.

The following technical result will be used in the future.

A.2. Lemma. *If $a > 0$ and $f(t) \geq 0$, then*

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

Proof. Note that

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

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

$$\begin{gathered} U = 2\int_0^T \int_0^t \int_0^s e^{-a(2t-s-u)} f(u) f(s) du ds dt = \\ 2\int_0^T \Big(\int_0^s \Big(\int_s^T e^{-2at} dt\Big) e^{au} f(u) du\Big) e^{as} f(s) ds = \\ \int_0^T \Big(\int_0^s a^{-1}(e^{-2as} - e^{2aT})\; e^{au} f(u) du\Big) e^{as} f(s) ds \leq \\ a^{-1}\int_0^T \Big(\int_0^s e^{-a(s-u)} f(u) du\Big) f(s) ds \leq \\ a^{-1}\Big(\int_0^T f^2(s)ds\Big)^{1/2}\Big(\int_0^T \Big(\int_0^s e^{-a(s-u)} f(u)du\Big)^2 ds\Big)^{1/2} = \\ a^{-1}\Big(\int_0^T f^2(s)ds\Big)^{1/2} U^{1/2}, \end{gathered}$$

and the result follows.

□

It is shown in the next lemma that under certain conditions on the operator $\mathcal{P}$ the asymptotic behavior of $\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}u(t)\|_0^2 dt$, $K \to \infty$, is determined by that of $\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}\xi(t)\|_0^2 dt$.

A.3. Lemma. *If $\mathcal{P}$ is an essentially non-degenerate operator of order p on M and $p \geq m - d/2$, then*

$$\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}\xi(t)\|_0^2 dt \asymp \sum_{k=1}^N l_k^{(p-m)/m},\; K \to \infty, \tag{A.3}$$

$$\lim_{K\to\infty} \frac{\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}\eta(t)\|_0^2 dt}{\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}\xi(t)\|_0^2 dt} = 0, \tag{A.4}$$

$$\mathbf{P} - \lim_{K\to\infty} \frac{\int_0^T \|\Pi^K \mathcal{P}\eta(t)\|_0^2 dt}{\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}\xi(t)\|_0^2 dt} = 0, \tag{A.5}$$

$$\mathbf{P} - \lim_{K\to\infty} \frac{\int_0^T \|\Pi^K \mathcal{P}\xi(t)\|_0^2 dt}{\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}\xi(t)\|_0^2 dt} = 1. \tag{A.6}$$

Proof.

Proof of (A.3). It follows from the independence of $\xi_k(t)$ for different k that

$$\mathbf{E}\sum_{k=1}^{K}|\psi_k(\mathcal{P}\xi(t))|^2 = \mathbf{E}\sum_{k=1}^{K}\Big|\sum_{n\geq 1}\xi_n(t)(e_n,\mathcal{P}^*e_k)_0\Big|^2 =$$
$$\sum_{k=1}^{K}\sum_{n\geq 1}\frac{1}{2\theta_1 l_n}(1-e^{-2\theta_1 l_n t})|(e_n,\mathcal{P}^*e_k)_0|^2.$$

Integration yields:

$$\mathbf{E}\int_0^T \|\Pi^K\mathcal{P}\xi(t)\|_0^2 dt$$

$$=\sum_{k=1}^{K}\sum_{n\geq 1}\frac{1}{2\theta_1 l_n}\left(T-\frac{1}{2\theta_1 l_n}(1-e^{-2\theta_1 l_n T})\right)|(e_n,\mathcal{P}^*e_k)_0|^2.$$

Since $l_k\to\infty$ and only asymptotic behavior, as $K\to\infty$, of all expressions is studied, it can be assumed that $1-e^{-2\theta_1 l_k T}>0$ for all k. Then the last equality and the definition of the norm $\|\cdot\|_s$ imply

$$\frac{T}{2\theta_1}\sum_{k=1}^{K}\|\mathcal{P}^*e_k\|_{-m}^2 - C\sum_{k=1}^{K}\|\mathcal{P}^*e_k\|_{-2m}^2 \leq \mathbf{E}\int_0^T \|\Pi^K\mathcal{P}\xi(t)\|_0^2 dt \leq$$
$$\frac{T}{2\theta_1}\sum_{k=1}^{K}\|\mathcal{P}^*e_k\|_{-m}^2.$$

Since $\mathcal{P}$ satisfies (3.6),

$$\|\mathcal{P}^*e_k\|_{-m}^2 \geq \varepsilon\|e_k\|_{p-m}^2 - K\|e_k\|_{p-m-\delta}^2 = \varepsilon\lambda_k^{2(p-m)}(1-(K/\varepsilon)\lambda_k^{-2\delta}).$$

In addition, $\|\mathcal{P}^*e_k\|_r^2\leq C\|e_k\|_{r+p}^2$ and $\lambda_k = l_k^{1/(2m)}$. The result (A.3) follows.

Proof of (A.4). By assumptions,

$$c := \max(order(\mathcal{A}), order(\mathcal{B}), order(\mathcal{M})) < 2m.$$

By Lemma A.2,

$$\int_0^T |\eta_n(t)|^2 dt \leq \frac{1}{(\theta_1 l_n)^2}\int_0^T \Big|\psi_n\Big((\theta_1\mathcal{A}+\theta_2\mathcal{B}+\mathcal{M})u(t)\Big)\Big|^2 dt,$$

which implies that for every $r\in\mathbb{R}$

$$\sum_{n\geq 1}\lambda_n^{2r}\int_0^T |\psi_n(\mathcal{P}\eta(t))|^2 dt \equiv \int_0^T \|\mathcal{P}\eta(t)\|_r^2 dt \leq C\int_0^T \|\eta(t)\|_{r+p}^2 dt \equiv$$
$$\sum_{n\geq 1}\lambda_n^{2(r+p)}\int_0^T |\eta_n(t)|^2 dt \leq C\int_0^T \|u(t)\|_{r-2m+c+p}^2.$$

If $c_1 := 2m-c>0$ and $r=-x$ where $x=\max(0, d/2+c_1/2+p+c-3m)$, then $-x-2m+c+p = m-d/2-c_1/2$ and, by (2.4), $E\int_0^T \|u(t)\|_{-x-2m+c+p}^2<\infty$. As a result, since $\lambda_k\asymp k^{1/d}$,

$$\frac{\mathbf{E}\int_0^T \|\Pi^K\mathcal{P}\eta(t)\|_0^2 dt}{\mathbf{E}\int_0^T \|\Pi^K\mathcal{P}\xi(t)\|_0^2 dt} = \frac{\sum_{n=1}^{K}\lambda_n^{-2x}\lambda_n^{2x}\mathbf{E}\int_0^T |\psi_n(\mathcal{P}\eta(t))|^2 dt}{\mathbf{E}\int_0^T \|\Pi^K\mathcal{P}\xi(t)\|_0^2 dt} \leq$$

$$\frac{CK^{2x/d}\sum_{n\geq 1}\lambda_n^{-2x}\mathbf{E}\int_0^T |\psi_n(\mathcal{P}\eta(t))|^2 dt}{\mathbf{E}\int_0^T \|\Pi^K\mathcal{P}\xi(t)\|_0^2 dt} \leq \frac{CK^{2x/d}}{\sum_{k=1}^{K}\lambda_k^{2(p-m)}} \to 0 \text{ as } K\to\infty,$$

because if $p - m = -d/2$, then $d/2 + c_1/2 + p + c - 3m = -c_1/2 < 0$ so that $x = 0$, while for $p-m > -d/2$ the sum $\sum_{k=1}^{N} \lambda_k^{2(p-m)}$ is of order $N^{2(p-m)/d+1}$ and $2(p-m)/d+1 > (d+2(p-m)-c_1/2) = 2x/d$. This proves (A.4). Then (A.5) follows from (A.4) and the Chebychev inequality.

Proof of (A.6). There are two steps in the proof. Writing $X_K(t) := \|\Pi^K \mathcal{P}\xi(t)\|_0^2$, the first step is to show that for all $t \in [0,T]$,

$$var(X_K(t)) \leq C \sum_{k=1}^{K} \lambda_k^{4(p-m)}. \tag{A.7}$$

This will imply that (A.6) holds (the second step), i.e. that

$$\mathbf{P} - \lim_{K\to\infty} \frac{\int_0^T X_K(t)dt}{\mathbf{E}\int_0^T X_K(t)dt} = 1.$$

1). If $X_K^M(t) := \sum_{k=1}^{K} |\sum_{n=1}^{M} \xi_n(t)(e_n, \mathcal{P}^* e_k)_0|^2$, then $X_K^M(t)$ is a quadratic form of the Gaussian vector $(\xi_1(t), \ldots, \xi_M(t))$. The matrix of the quadratic form is $A = [A_{nn'}]_{n,n'=1,\ldots,M}$ with

$$A_{nn'} = \sum_{k=1}^{K} (e_n, \mathcal{P}^* e_k)_0 (e_{n'}, \mathcal{P}^* e_k)_0,$$

and the covariance matrix of the Gaussian vector is

$$R = diag\left(\frac{1 - e^{-2\theta_1 l_n t}}{2\theta_1 l_k},\ n = 1, \ldots, M \right).$$

It is still assumed that $1 - e^{-2\theta_1 l_k T} > 0$ for all k.

Direct computations yield

$$\mathbf{E} X_K^M(t) = \sum_{k=1}^{K} \sum_{n=1}^{M} \frac{1}{2\theta_1 l_n} (1 - e^{-2\theta_1 l_n t}) |(e_n, \mathcal{P}^* e_k)_0|^2 = trace(AR).$$

Analysis of the proof of (A.3) shows that for every $t \in [0,T]$ and $k = 1, \ldots, K$ the series $\sum_{n\geq 1} \xi_n(t)(e_n, \mathcal{P}^* e_k)_0$ converges with probability one and in the mean square. Consequently,

$$\begin{aligned} &\lim_{M\to\infty} X_K^M(t) = X_K(t) \qquad (\mathbf{P}\text{- a.s.}); \\ &\lim_{M\to\infty} \mathbf{E} X_K^M(t) = \sum_{k=1}^{K} \sum_{n\geq 1} \mathbf{E}|\xi_n(t)|^2\ |(e_n, \mathcal{P}^* e_k)_0|^2 = \mathbf{E} X_K(t). \end{aligned} \tag{A.8}$$

Next,

$$var(X_K^M(t)) = 2trace((AR)^2) \leq C \sum_{n,n'} \frac{1}{l_n l_{n'}} A_{nn'}^2 =$$
$$\sum_{k,k'=1}^{K} |(\tilde{\mathcal{P}} e_k, e_{k'})_0|^2 \lambda_k^{4(p-m)} \leq \sum_{k=1}^{K} \|\tilde{\mathcal{P}} e_k\|_0^2 \lambda_k^{4(p-m)} \leq C \sum_{k=1}^{K} \lambda_k^{4(p-m)},$$

where $\tilde{\mathcal{P}} := \mathcal{P}\Lambda^{-2m}\mathcal{P}^*\Lambda^{2(m-p)}$ is a bounded operator in $I\!H^0$. After that, inequality (A.7) follows from (A.8) and the Fatou lemma:

$$\begin{aligned} var(X_K(t)) = \mathbf{E} \lim_{M\to\infty} |X_K^M(t)|^2 - |\mathbf{E} \lim_{M\to\infty} X_K^M(t)|^2 = \\ \mathbf{E} \lim_{M\to\infty} |X_K^M(t)|^2 - \lim_{M\to\infty} |\mathbf{E} X_K^M(t)|^2 \leq \liminf_{M\to\infty} \mathbf{E}|X_K^M(t)|^2 \\ - \lim_{M\to\infty} |\mathbf{E} X_K^M(t)|^2 \leq \liminf_{M\to\infty} var(X_K^M(t)) \leq C \sum_{k=1}^{K} \lambda_k^{4(p-m)}. \end{aligned}$$

2). If $Y_K := \int_0^T (X_K(t) - \mathbf{E}X_K(t))dt / \mathbf{E}\int_0^T X_K(t)dt$, then

$$\frac{\int_0^T X_K(t)dt}{\mathbf{E}\int_0^T X_K(t)dt} = 1 + Y_K$$

and

$$\mathbf{E}Y_K^2 \le \frac{T\int_0^T (var(X_K(t))dt}{\left(\mathbf{E}\int_0^T X_K(t)dt\right)^2} \le C\frac{\sum_{k=1}^K \lambda_k^{4(p-m)}}{\left(\sum_{k=1}^K \lambda_k^{2(p-m)}\right)^2} \to 0 \quad \text{as } K \to \infty.$$

By the Chebychev inequality, $\mathbf{P} - \lim_{K\to\infty} Y_K = 0$, which implies (A.6).

□

A.4. Corollary. *If $\mathcal{P}$ is an essentially non-degenerate operator of order p on M and $p \ge m - d/2$, then*

$$\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}u(t)\|^2 dt \asymp \frac{\varepsilon T}{2\theta_1}\sum_{k=1}^K l_k^{(p-m)/m}, \;\; K \to \infty, \tag{A.9}$$

and

$$\mathbf{P} - \lim_{K\to\infty} \frac{\int_0^T \|\Pi^K \mathcal{P}u(t)\|_0^2 dt}{\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}u(t)\|_0^2 dt} = 1. \tag{A.10}$$

Proof. By the inequality $|2xy| \le \epsilon x^2 + \epsilon^{-1}y^2$, which holds for every $\epsilon > 0$ and every real x, y,

$$\begin{aligned}
&(1-\epsilon)\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}\xi(t)\|_0^2 dt + (1-\frac{1}{\epsilon})\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}\eta(t)\|_0^2 dt \le \\
&\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}u(t)\|_0^2 dt \le \\
&(1+\epsilon)\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}\xi(t)\|_0^2 dt + (1+\frac{1}{\epsilon})\mathbf{E}\int_0^T \|\Pi^K \mathcal{P}\eta(t)\|_0^2 dt.
\end{aligned}$$

Since ϵ is arbitrary, (A.9) follows from (A.4) and (A.3). After that, (A.10) follows from (A.6).

□

A.5. Lemma. *Assume that $\mathcal{B}$ is an elliptic operator of order $b > m - d/2$. Then*

$$\mathbf{P} - \lim_{K\to\infty} \frac{\int_0^T \left(\Pi^K(\mathcal{L}+\mathcal{A})u(t), \mathcal{B}u(t)\right)_0 dt}{\mathbf{E}\int_0^T \left(\Pi^K(\mathcal{L}+\mathcal{A})u(t), \mathcal{B}u(t)\right)_0 dt} = 1.$$

Proof. With no loss of generality assume that $\mathcal{B}$ is bounded from below: for all $s \in I\!\!R$ there exist positive numbers C_1, C_2, δ so that the inequality

$$(\mathcal{B}f, f)_s \ge C_1\|f\|_{s+b/2}^2 - C_2\|f\|_{s+b/2-\delta} \tag{A.11}$$

holds for all $f \in \mathbf{C}^\infty(M)$; otherwise replace $\mathcal{B}$ by $-\mathcal{B}$. Direct computations show that

$$\begin{aligned}
&\mathbf{E}\int_0^T \left(\Pi^K \mathcal{L}\xi(t), \mathcal{B}\xi(t)\right)_0 dt \asymp K^{b/d+1};\; \mathbf{E}\int_0^T \|\Pi^K \mathcal{A}u(t)\|_0^2 dt \le CK^{2(m-\delta)/d+1}; \\
&\mathbf{E}\int_0^T \|\Pi^K \mathcal{L}\xi(t)\|_0^2 dt \asymp K^{2m/d+1};\; \mathbf{E}\int_0^T \|\Pi^K \mathcal{B}\xi(t)\|_0^2 dt \asymp K^{2(b-m)/d+1}.
\end{aligned}$$

(The last two relations follow from Lemma A.3.) The Cauchy-Schwartz inequality then implies that the statement of the lemma will follow from the convergence

$$\mathbf{P} - \lim_{K\to\infty} \frac{\int_0^T \left(\Pi^K \mathcal{L}\xi(t), \mathcal{B}\xi(t)\right)_0 dt}{\mathbf{E}\int_0^T \left(\Pi^K \mathcal{L}\xi(t), \mathcal{B}\xi(t)\right)_0 dt} = 1. \tag{A.12}$$

It can be shown in the same way as in the proof of (A.6) that

$$var\left((\Pi^K \mathcal{L}\xi(t), \mathcal{B}\xi(t))_0\right) \leq CK^{2b/d+1}$$

for every $t \in [0, T]$ with C independent of t. After that the Chebychev inequality implies (A.12), which completes the proof of the lemma.

□

References

1. Aihara, S. I.: 1992, 'Regularized Maximum Likelihood Estimate for an Infinite Dimensional Parameter in Stochastic Parabolic Systems'. *SIAM Journal on Control and Optimization* **30**(4), 745–764.
2. Bagchi, A. and V. Borkar: 1984, ' Parameter Identification in Infinite Dimensional Linear Systems'. *Stochastics* **12**, 201–213.
3. Frankignoul, C.: 1985, 'SST Anomalies, Planetary Waves and RC in the Middle Rectitudes'. *Reviews of Geophysics* **23**(4), 357–390.
4. Huebner, M.: 1993, 'Parameter Estimation for SPDE's'. Ph.D. thesis, University of Southern California, Los Angeles, CA, 90089.
5. Huebner, M.: 1997, 'A characterization of asymptotic behaviour of maximum likelihood estimators for stochastic PDE's'. *Mathematical Methods of Statistics* **6**(4), 395–415.
6. Huebner, M., R. Khasminskii, and B. L. Rozovskii: 1992, 'Two Examples of Parameter Estimation'. In: S. Cambanis, J. K. Ghosh, R. L. Karandikar, and P. K. Sen (eds.): *Stochastic Processes*. Springer, New York.
7. Huebner, M. and S. Lototsky: 2000a, ' Asymptotic Analysis of a Kernel Estimator for a Class of Parabolic Equations with Time-Dependent Coefficients'. *Annals of Applied Probability* **10**(4), 1246–1258.
8. Huebner, M. and S. Lototsky: 2000b, ' Asymptotic Analysis of the Sieve Estimator for a Class of Parabolic SPDEs'. *Scandinavian Journal of Statistics* **27**(2), 353–370.
9. Huebner, M., S. Lototsky, and B. L. Rozovskii: 1998, ' Asymptotic Properties of an Approximate Maximum Likelihood Estimator for Stochastic PDEs'. In: Y. M. Kabanov, B. L. Rozovskii, and A. N. Shiryaev (eds.): *Statistics and Control of Stochastic Processes: In honour of R. Sh. Liptser*. World Scientific, Singapore, pp. 139–155.
10. Huebner, M. and B. Rozovskii: 1995, 'On Asymptotic Properties of Maximum Likelihood Estimators for Parabolic Stochastic PDE's'. *Probability Theory and Related Fields* **103**, 143–163.
11. Ibragimov, I. A. and R. Z. Khasminskii: 1981, *Statistical Estimation (Asymptotic Theory)*. New York: Springer.
12. Ibragimov, I. A. and R. Z. Khasminskii: 1997, 'Some Nonparametric Estimation Problems for Parabolic SPDE'. Technical Report 31, Wayne State University, Department of Mathematics.
13. Ibragimov, I. A. and R. Z. Khasminskii: 1998, ' Problems of estimating the coefficients of stochastic partial differential equations. I'. *Teor. Veroyatnost. i Primenen (Russian)* **43**(3), 417–438. English trasnlation: *Theory Probab. Appl.*, **43** (1999), no. 3, 370–387.
14. Ibragimov, I. A. and R. Z. Khasminskii: 1999, ' Problems of estimating the coefficients of stochastic partial differential equations. II'. *Teor. Veroyatnost. i Primenen (Russian)* **44**(3), 526–554. English trasnlation: *Theory Probab. Appl.*, **44** (2000), no. 3, 469–494.
15. Jacod, J. and A. N. Shiryayev: 1987, *Limit Theorems for Stochastic Processes*. Berlin: Springer.
16. Krein, S. G., J. I. Petunin, and E. M. Semenov: 1981, *Interpolation of Linear Operators*. Providence, Rhode Island: American Mathematical Society.
17. Kumano-go, H.: 1981, *Pseudo-Differential Operators*. Cambridge, Massachusetts: The MIT Press.
18. Kutoyants, Y.: 1994, *Identification of Dynamical Systems With Small Noise*. Dordrecht: Kluwer Academic Publisher.
19. Liptser, R. S. and A. N. Shiryayev: 1989, *Theory of Martingales*. Boston: Kluwer Academic Publishers.
20. Liptser, R. S. and A. N. Shiryayev: 1992, *Statistics of Random Processes*. New York: Springer.
21. Lototsky, S. V. and B. L. Rozovskii: 1999, 'Spectral Asymptotics of Some Functional Arising in Statistical Inference for SPDEs '. *Stoch. Proc. Appl.* **79**, 64–94.
22. Mikulevicius, R. and B. Rozovskii: 1994, 'Uniqueness and Absolute Continuity of Weak Solutions for Parabolic SPDE's'. *Acta Applicandae Mathematicae* **35**, 179–192.
23. Piterbarg, L. and B. Rozovskii: 1997, ' On Asymptotic Problems of Parameter Estimation in Stochastic PDE's: Discrete Time Sampling.'. *Mathematical Methods of Statistics* **6**(2), 200–223.
24. Rozovskii, B. L.: 1990, *Stochastic Evolution Systems*. Kluwer Academic Publishers.
25. Serrano, S. and G. Adomian: 1996, ' New contributions to the solution of transport equations in porous media'. *Math. Comput. Modelling* **24**(4), 15–25.
26. Serrano, S. and T. Unny: 1990, 'Random evolution equations in hydrology'. *Appl. Math. Comput.* **38**(3), 201–226.
27. Shubin, M.: 1987, *Pseudodifferential Operators and Spectral Theory*. New York: Springer.
28. Walsh, J. B.: 1984, 'An Introduction to Stochastic Partial Differential Equations'. In: P. L. Hennequin (ed.): *Ecole d'été de Probabilités de Saint-Flour, XIV, Lecture Notes in Mathematics*, Vol. 1180. Springer, Berlin, pp. 265–439.