Optimal Filtering of Stochastic Parabolic Equations

S. V. Lototsky

Abstract. An estimation problem is considered for a stochastic parabolic equation with an unknown random coefficient. The additional randomness in the coefficient generalizes a popular estimation problem that has been extensively studied in recent years. The filter estimate of the coefficient is constructed from a finite-dimensional projection of the solution of the equation. Under certain conditions this estimate is approximated using a generalized Kalman-Bucy filter whose filter variance tends to zero as the dimension of the projection increases.

In: S. Albeverio, Z-M. Ma, and M. Roeckner (editors), Recent Developments in Stochastic Analysis and Related Topics (Proceedings of the First Sino-German Conference on Stochastic Analysis, August 29–September 3, 2002, Beijing, China), pp. 330–353, World Scientific, 2004.

1. Introduction

Stochastic partial differential equations (SPDEs) are becoming more and more popular as a modelling tool in various branches of applied science. Hydrology [**38, 39**], mathematical finance [**37**], physical oceanography [**7, 32**], and population biology [**5, 6**] are some of the areas currently using SPDE-based models. Numerous other examples of applications of SPDEs can be found in the books [**4, 10, 23**] and in the classical paper [**40**].

Successful utilization of any equation as a modelling tool requires rigorous results about existence, uniqueness, and regularity properties of the solution under sufficiently general assumptions. Even though the whole topic of SPDEs is relatively young, there are already many comprehensive studies of the analytical properties of both linear and non-linear equations, for examples, [**10, 23, 24, 34, 35, 40**]. Still, these analytic results are only the first step toward efficient practical use of SPDEs. Indeed, every time a real-life process is represented by an equation, only the general form of the equation is known and the details must be determined by reconciling the model with the observations of the process. In other words, an inverse problem must be solved to find, on the basis of the observations, the coefficients, free terms, and, sometimes, initial and boundary conditions in the equation.

For stochastic equations, inverse problems are usually solved by methods of statistical inference, using the observations as the input of a suitable estimator. The key mathematical question is the asymptotic behavior of the estimator, that is, whether the estimator approaches the true value, and how fast, as more and more of the observations become available. In particular, long time asymptotic assumes increase of the observation time, and small noise asymptotic, decrease of the noise intensity in the equation.

The first works on statistical inference for SPDEs [**1, 2, 27**] studied estimators in the long time asymptotic. Later, models with small observation noise were introduced and studied [**8, 12, 18, 19, 20, 21**]. While some of the papers address estimation of the initial condition and free force, the majority of the research has

---

2000 *Mathematics Subject Classification.* Primary 60G35; Secondary 60H15, 62M20.

Key words and phrases. Bayes Estimator, Conditionally Gaussian Process, Degenerate Equation, Kalman-Bucy Filter, Kushner Equation, Ricatti Equation.

The work was partially supported by the Sloan Research Fellowship and by the ARO Grant DAAD19-02-1-0374.

been on parameter estimation. The following model has become especially popular:

$$du(t,x) = (\mathcal{A}_0 + \theta(t,x)\mathcal{A}_1)u(t,x)dt + \varepsilon dW(t,x),\ 0 < t \leq T,\ u|_{t=0} = u_0, \tag{1.1}$$

where $\theta = \theta(t,x)$ is an unknown coefficient, $\mathcal{A}_0$ and $\mathcal{A}_1$ are known differential operators so that $\mathcal{A}_0 + \theta\mathcal{A}_1$ is elliptic for all admissible values of θ, and W is a space-time white noise. The equation is considered on a smooth compact manifold or in a smooth bounded domain with some boundary conditions. Long time asymptotic means $T \to \infty$, and small noise, $\varepsilon \to 0$.

Under certain conditions, a consistent estimator of θ in (1.1) is possible even if T and ε are fixed. Any computable estimator must be based on a finite dimensional projection of the solution of (1.1), and, if the order of the operator $\mathcal{A}_1$ is sufficiently high, then the estimator can approach the true value as the dimension of the projection increases even if the observation time and the noise intensity remain fixed. This asymptotic behavior, known as spectral asymptotic, is much more interesting than either long time or small noise, and has no analogs in finite-dimensional setting. The projection-based, or spectral, estimators were first introduced in [**13**], and further studied in [**17**], for the model of the type (1.1) with one unknown scalar parameter $\theta(t,x) = \theta_0 \in \mathbb{R}$ and with commuting operators $\mathcal{A}_0$, $\mathcal{A}_1$. The commutativity assumption ensures that the SPDE (1.1) is diagonalizable, that is, can be reduced to a system of uncoupled ordinary differential equations. Further work in that direction included analysis of maximum likelihood-type estimators for several scalar coefficients [**11**], sieve and kernel estimators for time-dependent coefficients $\theta = \theta(t)$ [**14, 15**], and Bayes-type estimators [**3**]. Non-diagonalizable models were also studied [**16, 30**].

In all the above works, the unknown coefficient θ was assumed deterministic. Additional randomness in θ not only makes the model more general but also poses new and interesting mathematical challenges. When $\theta = \theta(t)$ is random, the corresponding estimation problem becomes the problem of filtering, with the solution u of the SPDE (1.1) being the observation process. There are two classical filtering models, linear Gaussian and nonlinear diffusion, that have been extensively studied from both theoretical and applied points of view. Filtering problem for equation (1.1) does not fall into either of the categories. Indeed, to have existence and uniqueness of solution of (1.1), the unknown process θ must be uniformly bounded and therefore modelled by a nonlinear diffusion equation with degenerating coefficients, while the right-hand side of the observation process u is linear in u. As a result, new constructions and technical tools are necessary to carry out the analysis.

In the current paper, the filtering problem is studied for equation (1.1) with $\varepsilon = 1$, fixed $T > 0$, and a diffusion process $\theta = \theta(t)$ as the unknown coefficient. The underlying SPDE is assumed diagonalizable, and a finite-dimensional projection of the solution represents the observation process. The unknown coefficient is modelled by an Ito equation with coefficients degenerating at the end points of some interval. The special form of degeneracy ensures that the process never leaves the interval, while the corresponding filtering equations have a unique solution in a certain weighted function space. The filtering density satisfies a non-linear Kushner-type equation, and admits an alternative representation as a normalized solution of a linear Zakai-type equation. Under certain assumptions, an approximation of the optimal filter is constructed using a generalization of the Kalman-Bucy filter. The same condition on the order of the operators as in [**17**] ensures that, for every $T > 0$, the solution of the corresponding Ricatti equation, representing the filter variance, tends to zero as the dimension of the observation process increases.

Section 2 presents the basic existence, uniqueness, and regularity results for equation (1.1) under the assumption that the process θ is compactly supported. The filtering problem is studied in Section 3. The main result is that the filtering density exists and is a smooth function with the same support as θ. The proof is based on recent results about solvability of degenerate parabolic equations in domains [**29**]. The spectral asymptotic of the filter is studied in Section 4 when θ can be approximated, in a certain sense, by a Gaussian process.

2. Stochastic Parabolic Equations With Random Coefficients

Let G be either a smooth bounded domain in $\mathbb{R}^d$ or a smooth compact d-dimensional manifold without boundary and with a smooth positive measure. Denote by $C_0^\infty(G)$ the collection of infinitely differentiable, compactly supported, complex-valued functions on G. Let $\mathcal{A}_0$ and $\mathcal{A}_1$ be differential or pseudo-differential operators on $C_0^\infty(G)$. If G is a bounded domain, then, to simplify the presentation, all operators will be considered with zero boundary conditions.

On a stochastic basis $\mathbb{F}=(\Omega,\mathcal{F},\{\mathcal{F}_t\}_{t\ge 0},\mathbb{P})$ with the usual assumptions (see, for example, [**22**]), consider a cylindrical Brownian motion $W=W(t,x)$. In other words, W is a random process with values in the set $\mathcal{D}'(G)$ of distributions on G so that, for every $\varphi\in C_0^\infty(G)$ with $\|\varphi\|_{L_2(G)}=1$, $(W,\varphi)(t)$ is a standard Wiener process on $\mathbb{F}$, and for all $\varphi_1,\varphi_2\in C_0^\infty(G)$, $\mathbb{E}(W,\varphi_1)(t)(W,\varphi_2)(s)=\min(t,s)\cdot(\varphi_1,\varphi_2)_{L_2(G)}$.

For a predictable random process $\theta=\theta(t)$ on $\mathbb{F}$ and a $\mathcal{D}'(G)$-valued random variable u_0, consider the following equation:

$$\begin{aligned} du(t,x) &= (\mathcal{A}_0+\theta(t)\mathcal{A}_1)\,u(t,x)\,dt+dW(t,x), \quad t\in(0,T],\ x\in G,\\ u(0,x) &= u_0(x). \end{aligned} \tag{2.1}$$

Definition 1. *A predictable process u with values in $\mathcal{D}'(G)$ is called a solution of (2.1) if and only if, for every $\varphi\in C_0^\infty(G)$, the equality*

$$(u,\varphi)(t)=(u_0,\varphi)+\int_0^t(\mathcal{A}_0^*\varphi,u)(s)ds+\int_0^t\theta(s)(\mathcal{A}_1^*\varphi,u)(s)ds+(W,\varphi)(t)$$

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

$$(\mathcal{A}_i\phi_1,\phi_2)_0=(\mathcal{A}_i^*\phi_2,\phi_1)_0 \text{ for all } \phi_1,\phi_2\in C_0^\infty(G).$$

Remark 2.1. *It is possible to consider a more general noise process W with a correlation operator $\mathcal{B}$. As long as $\mathcal{B}$ is invertible, the corresponding equation is reduced to (2.1) by applying the operator $\mathcal{B}^{-1}$ to every term.*

Definition 2. *Equation (2.1) is called diagonalizable if and only if the following conditions hold:*

D1: *There is a complete orthonormal system $\{h_k,\ k\ge 1\}$ in $L_2(G)$ so that*

$$\mathcal{A}_0h_k=\kappa_kh_k,\quad \mathcal{A}_1h_k=\nu_kh_k.$$

D2: *There exist positive finite limits $\lim_{k\to\infty}|\nu_k|k^{-m_1/d}$ and $\lim_{k\to\infty}|\kappa_k|k^{-m_0/d}$, where m_i is the order of the operator $\mathcal{A}_i$.*

D3: *There exist positive real numbers c_1,c_2 so that, for all $t\in[0,T]$ and $\omega\in\Omega$,*

$$-c_1<\liminf_{k\to\infty}\frac{(\kappa_k+\theta(t)\nu_k)}{k^{2m/d}},\ \limsup_{k\to\infty}\frac{(\kappa_k+\theta(t)\nu_k)}{k^{2m/d}}<-c_2, \tag{2.2}$$

where $2m=\max(m_0,m_1)$.

Conditions D1–D3 hold in many physical models (see, for example, [**32, 33**]). A typical situation is when the operators $\mathcal{A}_0$ and $\mathcal{A}_1$ commute and either $\mathcal{A}_0$ or $\mathcal{A}_1$ is uniformly elliptic and formally self-adjoint. More details can be found in [**36**].

For $\varphi\in C_0^\infty(G)$ define $\varphi_k=\int_G\varphi(x)h_k(x)dx$. Conditions D1–D3 imply that, for every $n>0$, there exists a $C=C(\varphi,n)>0$ so that $|\varphi_k|\le Ck^{-n}$. Therefore, for every $\gamma\in\mathbb{R}$, we can define the norm $\|\cdot\|_\gamma$ on $C_0^\infty(G)$ as follows:

$$\|\varphi\|_\gamma^2=\sum_{k\ge1}k^{2\gamma/d}|\varphi_k|^2.$$

Let H^γ be the completion of $C_0^\infty(G)$ with respect to the norm $\|\cdot\|_\gamma$. An element $v\in H^\gamma$ is then identified with a sequence $\{v_k,\ k\ge 1\}$ of complex numbers so that $\|v\|_\gamma^2=\sum_{k\ge1}k^{2\gamma/d}|v|^2<\infty$. The numbers v_k are called *Fourier coefficients* of v, and $v\in H^\gamma$ is represented as a formal Fourier series $v(x)=\sum_{k\ge1}v_kh_k(x)$.

Remark 2.2. *A similar construction of the Hilbert spaces H^γ can be carried out when equation (2.1) is not diagonalizable; see* [**30**] *for details.*

THEOREM 2.3. *Assume that equation (2.1) is diagonalizable and* $u_0 \in L_2(\Omega; H^r)$ *for some* $r < -d/2$. *Then there is a unique solution* u *of (2.1). This solution belongs to the space* $L_2(\Omega \times (0,T); H^{m+r}) \bigcap L_2(\Omega; C((0,T), H^r))$ *and satisfies*

$$\mathbb{E} \sup_{0\le t\le T} \|u\|_r^2(t) + \mathbb{E}\int_0^T \|u\|_{m+r}^2(t)dt \le K(C_1, C_2, d, m, r, T)\left(\mathbb{E}\|u_0\|_r^2 + T\right),$$

where m *is from condition D3.*

Proof. It is know (see, for example, [**24**]) that $W(t,x) = \sum_{k\ge1} w_k(t)h_k(x)$, where $w_k,\ k\ge 1$, are independent Wiener processes. As a result, for every $r < -d/2$, the process W is an H^r-valued continuous square integrable martingale with quadratic variation $\langle W\rangle_t = t\sum_{k\ge1} k^{2r/d}$.

Next, conditions D1–D3 imply that, for every $r \in \mathbb{R}$, there exist positive numbers C_1, C_2 so that, for all $\varphi \in C_0^\infty(G)$, $t \in [0,T]$, and $\omega \in \Omega$,

$$\Re\Big(((\mathcal{A}_0 + \theta(t)\mathcal{A}_1)\varphi, \varphi)_r \Big) + C_1\|\varphi\|_{r+m}^2 \le C_2\|\varphi\|_r^2, \tag{2.3}$$

where $\Re(\cdot)$ is the real part of the expression. Indeed, it follows from (2.2) that

(1) There exists a positive number a_1 so that, all $k\ge1$ and all $t\in[0,T],\ \omega\in\Omega$,
$$|\kappa_k + \theta(t)\nu_k)| \le a_1 k^{2m/d}.$$
(2) There exists an integer k_0 and a positive number a_2 so that, for all $k > k_0$ and all $t \in [0,T],\ \omega\in\Omega$,
$$\Re(\kappa_k + \theta(t)\nu_k) \le -a_2 k^{2m/d}.$$

If $\varphi(x) = \sum_{k\ge1}\varphi_k h_k(x)$, then $(\mathcal{A}_0 + \theta(t)\mathcal{A}_1)\varphi(x) = \sum_{k\ge1}(\kappa_k + \theta(t)\nu_k)\varphi_k h_k(x)$ and

$$\Re\Big(((\mathcal{A}_0 + \theta(t)\mathcal{A}_1)\varphi, \varphi)_r \Big) = \sum_{k\ge1}\Re\Big(\kappa_k + \theta(t)\nu_k\Big)|\varphi_k|^2 k^{2r/d} \le -a_2\sum_{k\ge1}|\varphi_k|^2 k^{2(r+m)/d}$$
$$+(a_1+a_2)k_0^{2m/d}\sum_{k\ge1}|\varphi_k|^2 k^{2r/d} = -a_2\|\varphi\|_{r+m}^2 + (a_1+a_2)k_0^{2m/d}\|\varphi\|_r^2.$$

The statement of the theorem now follows from Theorem 3.1.4 and Remark 3.4.9 in [**35**]. □

The Fourier coefficients $u_k = u_k(t),\ k\ge1$, of the solution of (2.1) satisfy the following uncoupled system of stochastic ordinary differential equations

$$\begin{aligned} du_k(t) &= (\kappa_k + \theta(t)\nu_k)u_k(t)\,dt + dw_k(t),\ 0 < t \le T,\\ u_k(0) &= u_{0,k}. \end{aligned} \tag{2.4}$$

As a result, under conditions D1–D3, the infinite collection of equations (2.4) is equivalent to (2.1), and the solution of (2.1) can be written as a Fourier series

$$u(t,x) = \sum_{k\ge1} u_k(t)h_k(x), \tag{2.5}$$

converging in the corresponding Hilbert space. Note that condition D1 implies equations (2.4) for the Fourier coefficients of u, while conditions D2 and D3 ensure the appropriate convergence of the Fourier series (2.5).

Example. The following equation is a one-dimensional version of the heat balance equation from physical oceanography [**7, 32**]:

$$du(t,x) = (u_{xx} - \theta(t)u_x(t,x) - u(t,x))dt + dW(t,x),\ 0 < t \le T,\ 0 < x < 1, \tag{2.6}$$

with periodic boundary conditions. In this example, $G = S^1$, a circle, $\mathcal{A}_0 = \frac{d^2}{dx^2} - 1$, $\mathcal{A}_1 = -\frac{d}{dx}$, so that $m_0 = 2m = 2$, $m_1 = 1$. With a suitable ordering, $h_k(x) = e^{2\pi i k x}$, where $i = \sqrt{-1}$, so that $\kappa_k = -4\pi^2k^2 - 1$, $\nu_k = -2\pi i k$, and condition (2.2) holds as long as there exist real number a_θ, b_θ so that $\theta(t) \in [a_\theta, b_\theta]$ for all $t\in[0,T]$ and $\omega\in\Omega$.

Alternatively, one can consider

$$du(t,x) = (\theta(t)u_{xx} - u_x(t,x) - u(t,x))dt + dW(t,x),\ 0 < t \le T,\ 0 < x < 1, \tag{2.7}$$

with periodic boundary conditions. Then $\mathcal{A}_0 = -\frac{d}{dx} - 1$, $\mathcal{A}_1 = \frac{d^2}{dx^2}$, so that $m_0 = 1$, $m_1 = 2$. In this case, condition (2.2) holds as long as there exist *positive* numbers a_θ and b_θ so that $\theta(t) \in [a_\theta, b_\theta]$ for all $t \in [0,T]$ and $\omega \in \Omega$.

□

REMARK 2.4. *While it is possible to construct a path-wise solution of equation (2.1) if the numbers C_1 and C_2 in condition (2.3) are random and not uniformly bounded, this path-wise construction complicates the further analysis and is not discussed in the current paper.*

3. Optimal Nonlinear Filtering of Diagonalizable Equations

Consider the problem of estimating the random process θ from the observations of the first N Fourier coefficients (2.4) of the solution of the diagonalizable equation (2.1). The solution of this estimation problem is, of course, impossible without additional assumptions about the process θ.

Condition (2.2) implies that the process θ is uniformly bounded: there exist real numbers a_θ, b_θ so that, for all $\omega \in \Omega$,

$$\inf_{0\le t\le T} \theta(t) \ge a_\theta,\quad \sup_{0\le t\le T} \theta(t) \le b_\theta. \tag{3.1}$$

If $\mathcal{A}_1$ is not the leading operator, that is, if $m_1 < m_0 = 2m$, then there are no further restrictions on the numbers a_θ, b_θ. If $\mathcal{A}_1$ is the leading operator, that is, $m_0 < m_1 = 2m$, then (2.2) implies $a_\theta > 0$, that is, θ must be uniformly positive. Finally, if $m_0 = m_1 = 2m$, then the bounds on a_θ, b_θ will be determined by the asymptotic behavior of the sequences $k^{-2m/d}\nu_k$ and $k^{-2m/d}\kappa_k$; this asymptotic behavior depends on the particular operators (see [**36**] for details).

A possible model for θ is the Ito diffusion equation:

$$d\theta(t) = B(t,\theta(t))dt + r(t,\theta(t))dV(t), \tag{3.2}$$

where B and r are sufficiently regular functions and, for simplicity, the Wiener process V is independent of W. The initial choice of the functions B and r might not guarantee that (3.1) holds. Below is a general procedure for modifying equation (3.2) to ensure that condition (3.1) holds.

Let $\rho = \rho(x)$ be a smooth, compactly supported function on $\mathbb{R}$ so that

1. There exist finite nonzero limits $\lim_{x\to a_\theta} \rho(x)/(x - a_\theta)$ and $\lim_{x\to b_\theta} \rho(x)/(b_\theta - x)$;
2. $\rho(x) > 0$ for $x \in (a_\theta, b_\theta)$;
3. $\rho(x) = 1$ on $[a_\theta + \delta, b_\theta - \delta]$ for some sufficiently small $\delta > 0$.

Such a function exists and can be constructed by appropriately mollifying the characteristic function of the interval $[a_\theta, b_\theta]$. Note that $\rho(a_\theta) = \rho(b_\theta) = 0$, while the first derivative of ρ at those points is not zero.

Consider the following modification of equation (3.2):

$$d\theta(t) = \rho(\theta(t))B(t,\theta(t))dt + \rho(\theta(t))r(t,\theta(t))dV(t) \tag{3.3}$$

with some initial condition θ_0, independent of V and W.

PROPOSITION 3.1. *Assume that, for $0 \le t \le T$ and $x \in [a_\theta, b_\theta]$, the functions $B = B(t,x)$ and $r = r(t,x)$ are deterministic, bounded, and Lipschitz continuous in x, uniformly in (t,x). Then equation (3.3) has a unique strong solution for every square-integrable initial condition. If the initial condition θ_0 is a random variable whose distribution is supported in $[a_\theta, b_\theta]$, then the solution of (3.3) satisfies (3.1).*

Proof. Recall that ρ is a smooth compactly supported function. Assumptions about B and r imply that the coefficients in (3.3) are bounded and uniformly Lipschitz continuous in x. The first statement of the proposition is then a consequence of the general solvability theorem for the Ito equations (see, for example, Theorem 5.2.1 in [**31**]). The second statement of the proposition is the consequence of the uniqueness of solution of (3.3). Indeed, by assumption on the function ρ, $\rho(a_\theta) = \rho(b_\theta) = 0$, which means that the constant functions $\theta(t) = a_\theta$ and $\theta(t) = b_\theta$ satisfy (3.3). Therefore, each solution of (3.3) starting in $[a_\theta, b_\theta]$ will stay in that interval for all $t > 0$. □

The above proposition shows that (3.3) is an acceptable model of the coefficient process θ. The filtering problem can now be stated for the unobserved state process θ and the observation process $u_1, \ldots, u_N$:

$$\begin{aligned} d\theta(t) &= \rho(\theta(t))B(t,\theta(t))dt + \rho(\theta(t))r(t,\theta(t))dV(t), \\ du_k(t) &= (\kappa_k + \theta(t)\nu_k)u_k(t)dt + dw_k(t), \ k = 1, \ldots, N. \end{aligned} \tag{3.4}$$

The filtering problem for (3.4) consists in computing the conditional density of $\theta(t)$ given the observations up to time t. It is known [**35**, Chapter 6] that, under certain regularity assumptions, the conditional density in the diffusion filtering model satisfies a nonlinear stochastic parabolic equation, also know as Kushner's equation. Alternatively, the density can be computed by normalizing the solution of the linear Zakai equation. While the usual regularity assumptions for diffusion filtering models are not satisfied for (3.4), the corresponding equations can still be derived and studied.

Recall that the filtering density for (3.4) is a random field $\Pi = \Pi(t,x)$ so that, for every bounded measurable function $F = F(x)$,

$$\mathbb{E}(F(\theta(t))|u_k(s), \ k = 1, \ldots, N; \ 0 < s \le t) = \int_{\mathbb{R}} f(x)\Pi(t,x)dx.$$

Under condition (3.1), it is natural to expect Π to be supported in $[a_\theta, b_\theta]$ for all t.

Let

$$(\mathcal{L}f)(t,x) = \rho(x)B(t,x)f'(x) + \frac{1}{2}\rho^2(x)r^2(t,x)f''(x)$$

be the generator of θ. If the functions B and r are sufficiently smooth in x, then the adjoint $\mathcal{L}^*$ of $\mathcal{L}$ is defined by

$$(\mathcal{L}^*f)(t,x) = -\frac{\partial}{\partial x}\left(\rho(x)B(t,x)f(x)\right) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left(\rho^2(x)r^2(t,x)f(x)\right).$$

THEOREM 3.2. *Let the following conditions be fulfilled:*
1. The functions B and r are infinitely differentiable in x on $[a_\theta, b_\theta]$ so that each derivative with respect to x is uniformly bounded as a function of t and x.
2. There exists an $\varepsilon > 0$ so that $r^2(t,x) \ge \varepsilon$ for all $t \in [0,T]$ and $x \in [a_\theta, b_\theta]$.
3. The Wiener process V is independent of W.
4. The initial condition θ_0 is independent of V and W and has a density $\Pi_0 \in C_0^\infty((a_\theta, b_\theta))$.

Then the filtering density $\Pi = \Pi(t,x)$ for (3.4) exists and has the following properties:

(1) *For every $t \in [0,T]$ and $\mathbb{P}$–a.a. $\omega \in \Omega$, the support of Π is $[a_\theta, b_\theta]$ and the function Π is infinitely differentiable with respect to x with all the derivatives vanishing at points a_θ and b_θ.*
(2) *The function Π is a path-wise solution of the non-linear equation*

$$d\Pi(t,x) = (\mathcal{L}^*\Pi)(t,x)dt + \sum_{k=1}^N \Big((\kappa_k + x\nu_k)\Pi(t,x) - \bar{H}_k(t)\Pi(t,x)\Big)u_k(t)(du_k(t) - \bar{H}_k(t)dt) \tag{3.5}$$

with initial condition Π_0, where $\bar{H}_k(t) = \int_{\mathbb{R}^d}(\kappa_k + x\nu_k)\Pi(t,x)dx$.

Proof. While (3.5) is the formal Kushner equation for (3.4), the coefficients in (3.4) do not satisfy several technical assumptions that are traditionally used in the literature to derive the equation and study its properties. Specifically, the operator $\mathcal{L}$ is not uniformly elliptic (because of the function ρ) and the observation functions $H_k(x,y) = (\kappa_k + \nu_k x)y$ are not bounded in either x or y. This difficulty is resolved by using approximations and certain results about solvability of stochastic parabolic equations in weighted spaces.

For $M = 1, 2, \ldots$, define the stopping time

$$\tau_M = \inf\left\{ t > 0 : \sum_{k=1}^{N} |u_k(t)| > M \right\};$$

if $\sup_{0<t<T} \sum_{k=1}^{N} |u_k(t)| \le M$, we set $\tau_M = T$. Note that $\lim_{M\to\infty} \tau_M = T$ with probability one. It follows from Theorem 3.2, Remark 3.4, and Theorem 4.2 in [**29**] that equation (3.5) has a unique solution Π_M on the random interval $[0, \tau_M]$ in a certain weighted space. This solution is the filtering density for (3.4) on $[0, \tau_M]$. The support and smoothness properties of Π_M follow from the embedding theorems for the weighted spaces [**28**]. If $M_2 > M_1$, then, by uniqueness, $\Pi_{M_1} = \Pi_{M_2}$ on $[0, \tau_{M_1}]$. We then set $\Pi(t, x) = \Pi_M(t, x)$ on $[0, \tau_M]$. □

THEOREM 3.3. *Under the assumptions of Theorem 3.2, the filtering density* Π *can be represented as*

$$\Pi(t,x) = \frac{p(t,x)}{\int_{\mathbb{R}} p(t,x)dx}, \tag{3.6}$$

and the random field p *has the following properties:*

(1) *For every* $t \in [0,T]$ *and* $\mathbb{P}$*–a.a.* $\omega \in \Omega$*, the support of* p *is* $[a_\theta, b_\theta]$ *and the function* p *is infinitely differentiable with respect to* x *with all the derivatives vanishing at points* a_θ *and* b_θ
(2) *The function* p *is a path-wise solution of the linear equation*

$$dp(t,x) = (\mathcal{L}^* p)(t,x)dt + \sum_{k=1}^{N} (\kappa_k + x\nu_k) p(t,x) u_k(t) du_k(t)$$

with initial condition Π_0*.*

This theorem is proved in the same way as Theorem 3.2.

REMARK 3.4.
1. The smoothness conditions on the coefficients B *and* r *can be relaxed, with the appropriate loss of smoothness for* Π *and* p*.*

2. A more general form of equation (3.3) can be considered to include some, or all, of w_k *from the observations, and the results similar to Theorems 3.2 and 3.3 can be established for the corresponding filtering problem.*

4. Linear Filtering

The objective of this section is to study the following modification of the filtering problem (3.4):

$$\begin{aligned} d\theta(t) &= a(t)\theta(t)dt + b(t)dV(t), \\ du_k(t) &= (\kappa_k + \theta(t)\nu_k)u_k(t)dt + dw_k(t),\ k = 1, \ldots, N, \end{aligned} \tag{4.1}$$

$a = a(t)$ and $b = b(t)$ are measurable and bounded functions on $[0, T]$. If θ_0 is a Guassian random variable, then the solution $\theta = \theta(t)$ of the equation

$$d\theta(t) = a(t)\theta(t)dt + b(t)dV(t),\ 0 < t \le T,\ \theta(0) = \theta_0, \tag{4.2}$$

is a Gaussian process. While such a process cannot appear as a coefficient in (2.1) (condition (3.1) is not satisfied), the finite system (4.1) has a unique strong solution and the corresponding optimal filter has a much simpler structure than the one provided by Theorem 3.2. Moreover, this filter can be considered an approximate solution of the filtering problem for

$$\begin{aligned} d\theta(t) &= \rho(\theta(t))a(t)\theta(t)dt + \rho(\theta(t))b(t)dV(t), \\ du_k(t) &= (\kappa_k + \theta(t)\nu_k)u_k(t)dt + dw_k(t),\ k = 1, \ldots, N. \end{aligned} \tag{4.3}$$

Indeed, if $\sup_{0<t<T} |b(t)|$ is small, then the trajectories of (4.2) are close to the solution of the deterministic equation $\dot{x}(t) = a(t)x(t)$. With a suitable choice of the function $a = a(t)$ and initial condition θ_0, sufficiently many realizations of the process (4.2) will stay inside the interval $[a_\theta + \delta, b_\theta - \delta]$ for all $t \in [0, T]$, and, for such realizations, the solutions of (4.1) and (4.3) coincide.

We therefore consider (4.1) under the following assumptions:

- **(L1):** The functions $a = a(t)$ and $b = b(t)$ are measurable and bounded on $[0, T]$;
- **(L2):** The Wiener processes V and w_k, $k = 1, \dots, N$, are independent;
- **(L3):** The initial conditions $(\theta_0, u_{0,1}, \dots u_{0,N})$ are independent of the Wiener processes V and w_k, $k = 1, \dots, N$.
- **(L4):** $\mathbb{E}(\theta_0^4 + \sum_{k=1}^N u_{0,k}^4) < \infty$.
- **(L5):** The conditional distribution of θ_0 given $u_{0,1}, \dots u_{0,N}$ is $\mathbb{P}$-a.s. Gaussian.

It turns out that under these assumptions the conditional distribution of θ given the observations u_k is Gaussian, and the best mean-square estimate of $\theta(t)$ given $u_k(s)$ can be computed from a generalized Kalman-Bucy filter.

THEOREM 4.1. *Under assumptions (L1)–(L5), the conditional distribution of $\theta(t)$ given $\mathcal{F}^u_{N,t} = \sigma(u_k(s),\ k = 1, \dots, N,\ 0 \le s \le t)$, is $\mathbb{P}$-a.s. Gaussian with parameters*

$$\hat\theta_N(t) = \mathbb{E}(\theta(t)|\mathcal{F}^u_{N,t}), \quad \gamma_N(t) = \mathbb{E}\Big((\theta(t) - \hat\theta_N(t))^2|\mathcal{F}^u_{N,t}\Big).$$

The functions $\hat\theta_N(t)$ and $\gamma_N(t)$ satisfy the following system of equations:

$$\begin{aligned} d\hat\theta_N(t) &= a(t)\hat\theta_N(t) + \gamma_N(t)\sum_{k=1}^N \nu_k u_k(t)\Big(du_k - (\kappa_k u_k(t) + \nu_k u_k(t)\hat\theta_N(t))dt\Big),\\ \dot\gamma_N(t) &= 2a(t)\gamma_N(t) + b^2(t) - \gamma_N^2(t)\sum_{k=1}^N \nu_k^2 u_k^2(t), \end{aligned} \tag{4.4}$$

with initial conditions

$$\hat\theta_N(0) = \mathbb{E}(\theta_0|u_{0,1}, \dots u_{0,N}), \quad \gamma_N(0) = \mathbb{E}\Big((\theta_0 - \hat\theta_N(0))^2|u_{0,1}, \dots u_{0,N}\Big).$$

Proof. The result essentially follows from Theorems 8.1 in [**25**], and 12.6 and 12.7 in [**26**]. The main condition to verify is

$$\int_0^T \mathbb{E}|u_k(t)\theta(t)|^2 dt < \infty. \tag{4.5}$$

Note that θ is a Gaussian process independent of all w_k; $\mathbb{E}(\theta(t)) = \mathbb{E}(\theta_0)A(t)$; $\mathbb{E}|\theta(t)|^2 = A^2(t)\left(\mathbb{E}|\theta_0|^2 + \int_0^t A^{-2}(s)b^2(s)ds\right)$, where $A(t) = \exp\left(\int_0^t a(s)ds\right)$. Also, $u_k(t) = u_{0,k}A_k(t) + \int_0^t (A_k(t)/A_k(s))dw_k(s)$, where $A_k(t) = \exp\left(\int_0^t (\kappa_k + \nu_k\theta(s))ds\right)$. Then both $\mathbb{E}|\theta(t)|^4$ and $\mathbb{E}|u_k(t)|^4$ exist and are continuous functions of t, so that (4.5) holds. □

As was mentioned above, Theorem 4.1 and equations (4.4) can provide an approximate solution of the filtering problem for the nonlinear model (4.3). The central question in the study of equations (4.4) is the asymptotic behavior of the filter variance $\gamma_N(t)$ as more and more observations become available. When the observations come from a diagonalizable SPDE (2.1), there are two ways to increase the amount of information from the observations: to increase t or to increase N. Computing $\lim_{t\to\infty}\gamma_N(t)$ for fixed N requires no additional knowledge about equation (2.1) and is completely analogous to the corresponding problem for ordinary differential equations (see, for example, [**26**, Chapter 14]). Computing $\lim_{N\to\infty}\gamma_N(t)$ for fixed t has no analogs in the literature and will be discussed next.

THEOREM 4.2. *Consider filtering problem for (4.3) under the following assumptions:*

(1) *The distribution of the random variable θ_0 is supported in $[a_\theta, b_\theta]$;*
(2) *$u_0 \in L_2(H^{-d/2})$ is deterministic.*

Define $\gamma_N(t)$ according to (4.4) with some $\gamma_N(0) > 0$. If

$$q = \frac{2(m_1 - m)}{d} \ge -1, \tag{4.6}$$

then, with probability one, $\lim_{N\to\infty}\gamma_N(t_0)=0$ *for every* $0<t_0\le T$*. If, in addition,*

$$\inf_{0\le t\le T}|b(t)|>0,$$

then, for every $0<t_0\le T$*, there exists, with probability one, a finite positive limit* $\lim_{N\to\infty}\psi_N\gamma_N(t_0)$ *, where*

$$\psi_N=\begin{cases}\sqrt{\ln N}, & q=-1\\ N^{(q+1)/2}, & q>-1.\end{cases} \tag{4.7}$$

The proof of this theorem is based on the following two results about the Ricatti equation.

LEMMA 4.3. *Assume that* $x=x(t)\ge 0$ *is a solution of* $\dot{x}(t)=\alpha_1(t)x(t)-\beta_1(t)x^2(t)+\gamma_1(t)$*, and* $y=y(t)\ge 0$ *is a solution of* $\dot{y}(t)=\alpha_2(t)y(t)-\beta_2(t)y^2(t)+\gamma_2(t)$ *so that* $x(t_0)\ge y(t_0)\ge 0$*, and, for all* $t\ge t_0$*,* $\alpha_1(t)\ge\alpha_2(t)$*,* $\beta_2(t)\ge\beta_1(t)\ge 0$*,* $\gamma_1(t)\ge\gamma_2(t)\ge 0$*. Then* $x(t)\ge y(t)$ *for all* $t\ge t_0$*.*

Proof. If $z(t)=x(t)-y(t)$, then direct computations show that $\dot{z}(t)=A(t)z(t)+F(t)$, where $F(t)\ge 0,\ t\ge t_0$. Since $z(t_0)\ge 0$, it follows that $z(t)\ge 0$ for all $t\ge t_0$. □

LEMMA 4.4. *Let* $y_N=y_N(t)$*,* $N\ge 1$*,* $t\ge 0$*, be a positive solution of*

$$\dot{y}_N(t)=\alpha y_N(t)-\beta^2\psi_N^2y_N^2(t)+\gamma, \tag{4.8}$$

where $\alpha\in\mathbb{R}$*,* $\beta>0$*,* $\gamma>0$*, and* ψ_N *is defined in (4.7). Then, for every* $t>0$*,*

$$\lim_{N\to\infty}\psi_N y_N(t)=\sqrt{\gamma}/\beta.$$

Proof. This result follows from the explicit formula for $y_N(t)$ [**31**, Example 6.2.12]. The formula can easily be derived by observing that the equation for y_N is separable. □

Proof of Theorem 4.2. Since for each $t\in(0,T]$ the random variables $u_k(t)$, $k=1,\dots,N$, are conditionally independent given θ, the strong law of large numbers implies that, under condition (4.6), for every $t_0>0$, there exists, with probability one, a finite positive limit $\lim_{N\to\infty}\psi_N^{-1}\sum_{k=1}^N\nu_k^2u_k^2(t_0)$; see the proof of Lemma 2.2 in [**17**] for details. By Lemma 4.3, there exist a non-negative function y_N satisfying (4.8) and an integer-valued random variable μ satisfying $\mathbb{P}(0<\mu<\infty)=1$ so that, for all $N>\mu$, $\gamma_N(t_0)\le y_N(t_0)$. By Lemma 4.4 we conclude that $\lim_{N\to\infty}\gamma_N(t_0)=0$ with probability one. If $|b(t)|^2\ge\varepsilon>0$ for all t, then, by Lemma 4.3, there exist non-negative functions $y_N,\tilde{y}_N$, both satisfying equations of the type (4.8), and a positive integer-valued random variable μ satisfying $\mathbb{P}(\mu<\infty)=1$ so that, for all $N>\mu$, $\tilde{y}_N(t_0)\le\gamma_N(t_0)\le y_N(t_0)$. By Lemma 4.4 we conclude that $\lim_{N\to\infty}\psi_N\gamma_N(t_0)$ exists and is positive. Theorem 4.2 is proved. □

Theorem 4.2 shows that, under condition (4.6), the variance of the linear filter tends to zero as more and more of the spatial Fourier coefficients of the solution of (2.1) are included in the observation process. The question remains open whether a similar result holds for the non-linear filter. In the example at the end of Section 2, condition (4.6) holds for both equations; for equation $du=\Delta u-\vec{v}\cdot\nabla u+\theta(t)u$ with a known vector $\vec{v}$, this condition holds if and only if $d\ge 2$ [**17**].

Equations (4.4) can be solved explicitly if $a(t)=b(t)=0$. In that case, $\theta(t)=\theta_0$ and

$$\begin{aligned}\hat{\theta}_N(T)&=\frac{m_0+\gamma_0\sum_{k=1}^N\int_0^T\nu_ku_k(t)(du_k(t)-\kappa_ku_k(t))dt}{1+\gamma_0\sum_{k=1}^N\int_0^T\nu_k^2u_k^2(t)dt},\\ \gamma_N(T)&=\frac{\gamma_0}{1+\gamma_0\sum_{k=1}^N\int_0^T\nu_k^2u_k^2(t)dt},\end{aligned} \tag{4.9}$$

where $m_0=\mathbb{E}(\theta_0)$, $\gamma_0=var(\theta_0)$, and u_0 is assumed deterministic. If (4.6) holds, then direct computations show that $\lim_{N\to\infty}\hat{\theta}_N(T)=\theta_0$ with probability one. Similarly, $\lim_{N\to\infty}\psi_N^2\gamma_N(T)$ has a positive finite limit with probability one, which does not contradict Theorem 4.2. For a constant coefficient θ, expression (4.9) is an example of the Bayes estimator.

It is known from [**17**] that, if $\theta(t) = \theta$, a real number, then the maximum likelihood estimator $\hat{\theta}^N$ of θ is given by

$$\hat{\theta}^N = \frac{\sum_{k=1}^N \int_0^T \nu_k u_k(t)(du_k(t) - \kappa_k u_k(t))dt}{\sum_{k=1}^N \int_0^T \nu_k^2 u_k^2(t)dt}. \tag{4.10}$$

By comparing (4.9) with (4.10), we conclude that, for every $\alpha < 1$, $\lim_{N\to\infty} \psi_N^\alpha(\hat{\theta}_N(T) - \hat{\theta}^N) = 0$ with probability one and in every $L_q(\Omega)$. The results of [**17**] then imply that the Bayes estimator (4.9) is consistent, both with probability one and in every $L_q(\Omega)$, asymptotically normal with rate ψ_N, and asymptotically efficient.

References

[1] S. I. Aihara. Regularized Maximum Likelihood Estimate for an Infinite Dimensional Parameter in Stochastic Parabolic Systems. *SIAM Journal on Control and Optimization*, 30(4):745–764, 1992.
[2] A. Bagchi and V. Borkar. Parameter Identification in Infinite Dimensional Linear Systems. *Stochastics*, 12:201–213, 1984.
[3] J. P. N. Bishwal. The Bernstein–von Mises Theorem and Spectral Asymptotics of Bayes Estimators for Parabolic SPDEs. *J Australian Math. Society, Series A*, 72(2):287–298, 2001.
[4] R. Carmona and B. Rozovskii (editors). *Stochastic Partial Differential Equations: Six Perspetives.* AMS, Providence, RI, 1999.
[5] D. Dawson. Qualitative behavior of geostochastic systems. *Stoch. Proc. Appl.*, 10:1–31, 1980.
[6] S. De. Stochastic models of population growth and spread. *Bull. Math. Biol.*, 49:1–11, 1987.
[7] C. Frankignoul. SST Anomalies, Planetary Waves and RC in the Middle Rectitudes. *Reviews of Geophysics*, 23(4):357–390, 1985.
[8] G. K. Golubev and R. Z. Khasminskii. A Statistical Approach to Some Inverse Problems For Partial Differential Equations. *Problemy Peredachi Informatsii (Russian)*, 35(2):51–66, 1999. English translation in *Problems Inform. Transmission*, **35** (1999), no. 2, 136–149.
[9] T. Hida, H-H. Kuo, J. Potthoff, and L. Sreit. *White Noise.* Kluwer Academic Publishers, Boston, 1993.
[10] H. Holden, B. Øksendal, B. Ubøe, and T. Zhang. *Stochastic Partial Differential Equations.* Birkhauser, Boston, 1996.
[11] M. Huebner. A characterization of asymptotic behaviour of maximum likelihood estimators for stochastic PDE's. *Mathematical Methods of Statistics*, 6(4):395–415, 1997.
[12] M. Huebner. Asymptotic Properties of the Maximum Likelihood Estimator for Stochastic PDEs Disturbed by Small Noise. *Statistical Inference for Stochastic Processes*, 2:57–68, 1999.
[13] M. Huebner, R. Khasminskii, and B. L. Rozovskii. Two Examples of Parameter Estimation. In S. Cambanis, J. K. Ghosh, R. L. Karandikar, and P. K. Sen, editors, *Stochastic Processes*, pages 149–160. Springer, New York, 1992.
[14] M. Huebner and S. Lototsky. Asymptotic Analysis of a Kernel Estimator for a Class of Parabolic Equations with Time-Dependent Coefficients. *Annals of Applied Probability*, 10(4):1246–1258, 2000.
[15] M. Huebner and S. Lototsky. Asymptotic Analysis of the Sieve Estimator for a Class of Parabolic SPDEs. *Scandinavian Journal of Statistics*, 27(2):353–370, 2000.
[16] M. Huebner, S. Lototsky, and B. L. Rozovskii. Asymptotic Properties of an Approximate Maximum Likelihood Estimator for Stochastic PDEs. In Yu. M. Kabanov, B. L. Rozovskii, and A. N. Shiryaev, editors, *Statistics and Control of Stochastic Processes: In honour of R. Sh. Liptser*, pages 139–155. World Scientific, Singapore, 1998.
[17] M. Huebner and B. Rozovskii. On Asymptotic Properties of Maximum Likelihood Estimators for Parabolic Stochastic PDE's. *Probability Theory and Related Fields*, 103:143–163, 1995.
[18] I. A. Ibragimov and R. Z. Khasminskii. Some Estimation Problems in Infinite Dimensional White Noise. In D.Pollard, E.Torgersen, and G.Yang, editors, *Festschrift for Lucien Le Cam*, pages 259–274. Springer, Berlin, 1997.
[19] I. A. Ibragimov and R. Z. Khasminskii. Some estimation problems for stochastic partial differential equations. *Dokl. Akad. Nauk (Russian)*, 353(3):300–302, 1997.
[20] I. A. Ibragimov and R. Z. Khasminskii. Problems of estimating the coefficients of stochastic partial differential equations. I. *Teor. Veroyatnost. i Primenen (Russian)*, 43(3):417–438, 1998. English trasnlation: *Theory Probab. Appl.*, **43** (1999), no. 3, 370–387.
[21] I. A. Ibragimov and R. Z. Khasminskii. Problems of estimating the coefficients of stochastic partial differential equations. II. *Teor. Veroyatnost. i Primenen (Russian)*, 44(3):526–554, 1999. English trasnlation: *Theory Probab. Appl.*, **44** (2000), no. 3, 469–494.
[22] J. Jacod and A. N. Shiryayev. *Limit Theorems for Stochastic Processes.* Springer, Berlin, 1987.
[23] G. Kallianpur and X. Xiong. *Infinite Dimensional Stochastic Differential Equations*, volume 26 of *Lecture Notes–Monographs.* Institute of Mathematical Statistics, 1995.
[24] N. V. Krylov. *Introduction to the Theory of Diffusion Processes.* American Mathematical Society, Providence, RI, 1995.
[25] R. Sh. Liptser and A. N. Shiryayev. *Statistics of Random Processes, I: Geberal Theory.* Springer, New York, 2000.
[26] R. Sh. Liptser and A. N. Shiryayev. *Statistics of Random Processes, II: Applications.* Springer, New York, 2001.
[27] W. Loges. Girsanov's Theorem in Hilbert Space and an Application to Statistics of Hilbert Space Valued Stochastic Differential Equations. *Stoc. Proc. Appl.*, 17:243–263, 1984.

[28] S. V. Lototsky. Sobolev Spaces with Weights in Domains and Boundary Value Problems for Degenerate Elliptic Equations. *Methods and Applications of Analysis*, 7(1):195–204, 2000.

[29] S. V. Lototsky. Linear Stochastic Parabolic Equations, Degenerating at the Boundary of the Domain. *Electronic Journal of Probability*, 6, 2001. Paper number 24, http://www.math.washington.edu/~ejpecp/EjpVol6/paper24.abs.html.

[30] S. V. Lototsky and B. L. Rozovskii. Spectral Asymptotics of Some Functional Arising in Statistical Inference for SPDEs . *Stoc. Proc. Appl.*, 79:64–94, 1999.

[31] B. K. Oksendal. *Stochastic Differential Equations : an Introduction with Applications, 5th ed.* Springer, Berlin, 1998.

[32] L. Piterbarg and B. Rozovskii. Maximum Likelihood Estimators in the Equations of Physical Oceanography. In R. J. Adler et al., editor, *Stochastic Modelling in Physical Oceanography*, pages 397–421. Birkhäuser, Boston, 1996.

[33] L. Piterbarg and B. Rozovskii. On Asymptotic Problems of Parameter Estimation in Stochastic PDE's: Discrete Time Sampling. *Mathematical Methods of Statistics*, 6(2):200–223, 1997.

[34] G. Da Prato and J. Zabczyk. *Stochastic Equations in Infinite Dimensions.* Cambridge University Press, Cambridge, 1992.

[35] B. L. Rozovskii. *Stochastic Evolution Systems.* Kluwer Academic Publishers, 1990.

[36] Yu. Safarov and D. Vassiliev. *The Asymptotic Distribution of Eigenvalues of Partial Differential Operators*, volume 155 of *Transl. Math. Monogr.* AMS, Providence, RI, 1997.

[37] P. Santa-Clara and D. Sornette. The Dynamics of the Forward Interest Rate Curve with Stochastic String Shocks. *The Review of Financial Studies*, 14(1):149–185, 2001. On the Web: http://xxx.lanl.gov/abs/cond-mat/9801321.

[38] S. Serrano and G. Adomian. New contributions to the solution of transport equations in porous media. *Math. Comput. Modelling*, 24(4):15–25, 1996.

[39] S. Serrano and T. Unny. Random evolution equations in hydrology. *Appl. Math. Comput.*, 38(3):201–226, 1990.

[40] J. B. Walsh. An Introduction to Stochastic Partial Differential Equations. In P. L. Hennequin, editor, *Ecole d'été de Probabilités de Saint-Flour, XIV, Lecture Notes in Mathematics*, volume 1180, pages 265–439, Springer, Berlin, 1984.

Current address: Department of Mathematics, USC, Los Angeles, CA 90089, USA

E-mail address: lototsky@math.usc.edu

URL: http://math.usc.edu/~lototsky