Parameter estimation in diagonalizable bilinear stochastic parabolic equations

Igor Cialenco

Department of Applied Mathematics, Illinois Institute of Technology

10 West 32nd Str, Bld E1, Room 208, Chicago, IL 60616, USA

(igor@math.iit.edu, http://math.iit.edu/∼igor)

Sergey V. Lototsky *

Department of Mathematics, University of Southern California

3620 S. Vermont Avenue, KAP 108, Los Angeles, CA 90089, USA

(lototsky@math.usc.edu, http://math.usc.edu/∼lototsky)

Abstract. A parameter estimation problem is considered for a stochastic parabolic equation with multiplicative noise under the assumption that the equation can be reduced to an infinite system of uncoupled diffusion processes. From the point of view of classical statistics, this problem turns out to be singular not only for the original infinite-dimensional system but also for most finite-dimensional projections. This singularity can be exploited to improve the rate of convergence of traditional estimators as well as to construct completely new closed-form exact estimator.

Keywords: Regular models, singular models, multiplicative noise, SPDE.

AMS 2000: Primary 62F12; Secondary 60H15

1. Introduction

In the classical statistical estimation problem, the starting point is a family $\mathbf{P}^\theta$ of probability measures depending on the parameter θ belonging to some subset Θ of a finite-dimensional Euclidean space. Each $\mathbf{P}^\theta$ is the distribution of a random element. It is assumed that a realization of one random element corresponding to one value $\theta = \theta_0$ of the parameter is observed, and the objective is to estimate the values of this parameter from the observations.

* SVL acknowledges support from the NSF CAREER award DMS-0237724.

The intuition is to select the value θ corresponding to the random element that is *most likely* to produce the observations. A rigorous mathematical implementation of this idea leads to the notion of the regular statistical model [10]: the statistical model (or estimation problem) $\mathbf{P}^\theta,\ \theta \in \Theta$, is called regular, if the following two conditions are satisfied:

– there exists a probability measure $\mathbf{Q}$ such that all measures $\mathbf{P}^\theta$ are absolutely continuous with respect to $\mathbf{Q}$;

– the density $d\mathbf{P}^\theta/d\mathbf{Q}$, called the likelihood ratio, has a special property, called local asymptotic normality.

If at least one of the above conditions is violated, the problem is called singular.

In regular models, the estimator $\widehat{\theta}$ of the unknown parameter is constructed by maximizing the likelihood ratio and is called the maximum likelihood estimator (MLE). Since, as a rule, $\widehat{\theta} \neq \theta_0$, the consistency of the estimator is studied, that is, the convergence of $\widehat{\theta}$ to θ_0 as more and more information becomes available. In all known regular statistical problems, the amount of information can be increased in one of two ways: (a) Increasing the sample size, for example, the observation time interval (large sample asymptotic); (b) reducing the amplitude of noise (small noise asymptotic). The asymptotic behavior of $\widehat{\theta}$ in both cases is well-studied. It is also known that many other estimators in regular models are asymptotically equivalent to the MLE.

While all regular models are in a sense the same, each singular model is different. Sometimes, it is possible to approximate a singular model with a sequence of regular models. For each regular model, an MLE is constructed, and then in the limit one can often get the true value of the parameter while both the sample size and the noise amplitude are fixed. Some singular models cannot be approximated by a sequence of regular models and admit estimators that have nothing to do with the MLE [14]. In this paper, Section 4, we introduce a completely new type of such estimators for a large class of singular models.

Infinite-dimensional stochastic evolution equations, that is, stochastic evolution equations in infinite-dimensional spaces, are a rich source of statistical problems, both regular and singular. A typical example is the Itô equation

$$\begin{cases} du(t) + (\mathcal{A}_0 + \theta\mathcal{A}_1)u(t)dt = f(t)dt + \sum_{j\geq 1} (\mathcal{M}_j u(t) + g_j(t))dW_j(t), \\ u(0) = u_0, \end{cases} \tag{1.1}$$

where $t \in [0,T]$, $\mathcal{A}_0$, $\mathcal{A}_1$, $\mathcal{M}_j$ are linear operators, f, g_j are adapted processes, W_j are independent Wiener processes, and θ is the unknown parameter belonging to an open subset of the real line. The underlying assumption is that the solution u exists, is unique, and can be observed as an infinite-dimensional object for all $t \in [0,T]$. Depending on the operators in the equation, the estimation model can be regular, a singular limit of regular problems, or completely singular.

If $\mathcal{A}_0, \mathcal{A}_1$, $\mathcal{M}_j$ are partial differential or pseudo-differential operators, (1.1) becomes a stochastic partial differential equation (SPDE), which is becoming increasingly popular for modelling various phenomena in fluid mechanics [25], oceanography [21], temperature anomalies [4, 22], finance [3, 5, 6], and other domains. Various estimation problems for different types of SPDEs have been investigated by many authors: [1, 2, 7, 8, 9, 11, 12, 13, 17, 18, 19, etc.].

Depending on the stochastic part, (1.1) is classified as follows:

- equation with **additive noise**, if $\mathcal{M}_j = 0$ for all j;

- equation with **multiplicative noise** (or bilinear equation) otherwise.

Depending on the operators, (1.1) is classified as follows:

- **Diagonalizable equation**, if the operators $\mathcal{A}_0$, $\mathcal{A}_1$, and $\mathcal{M}_j$, $j \geq 1$, have a common system of eigenfunctions h_k, $k \geq 1$, and this system is an orthonormal basis in a suitable Hilbert space.

- **Non-diagonalizable equation** otherwise.

A diagonalizable equation is reduced to an infinite system of *uncoupled* one-dimensional diffusion processes; these processes are the Fourier coefficients of the solution in the basis h_k. As a result, while somewhat restrictive as a modelling tool, diagonalizable equations are an extremely convenient object to study estimation problems and often provide the benchmark results that carry over to more general equations.

The parameter estimation problem for a diagonalizable equation (1.1) with additive space-time white noise (that is, $g_j = h_j$ and $\mathcal{M}_j = 0$ for all j) was studied for the first time by Huebner, Khasminskii, and Rozovskii [8], and further investigated in [7, 8, 9, 23]. The main feature of this problem is that every N-dimensional projection of the equation leads to a regular statistical problem, but the problem can become singular in the limit $N \to \infty$ (a singular limit of regular problems); when this happens, the dimension N of the projection becomes a natural asymptotic parameter of the problem. Once the diagonalizable model is well-understood, extensions to more general equations can be considered ([18, 19]).

This paper is the first attempt to investigate the estimation problem for infinite-dimensional bilinear equations. Such models are often completely singular, that is, cannot be represented as a limit of regular models. We consider the more tractable situation of diagonalizable equations. In Section 2 we provide the necessary background on stochastic evolution equations, with emphasis on diagonalizable bilinear equations. The maximum likelihood estimator (MLE) and its modifications for diagonalizable bilinear equations are studied in Section 3. We give sufficient conditions on operators $\mathcal{A}_0, \mathcal{A}_1, \mathcal{M}$ that ensure consistency and asymptotic normality of the MLE. We also demonstrate that the MLE in this model is not always the best estimator, which, for a singular model is not at all surprising. Section 4 emphasizes the point even more by introducing a `closed-form exact estimator`. Due to the specific structure of stochastic term, for a large class of infinite-dimensional systems with finite-dimensional noise, one can get the *exact* value of the unknown parameter after a finite number of arithmetic manipulations with the observations. The very existence of such estimators in these models is rather remarkable and has no analogue in classical statistics.

As an illustration, let θ be a positive number, W a standard Wiener process, and consider the Itô equation

$$du(t,x) - \theta u_{xx}(t,x)dt = u(t,x)dW(t),\ t > 0,\ x \in (0,\pi), \tag{1.2}$$

with zero boundary conditions. If $h_k(x) = \sqrt{2/\pi}\,\sin(kx)$, $k \geq 1$, and

$$u_k(t) = \int_0^\pi u(t,x)h_k(x)dx\ ,$$

then

$$u(t,x) = \sum_{k\geq 1} u_k(t)h_k(x) \tag{1.3}$$

and each u_k is a geometric Brownian motion:

$$u_k(t) = u_k(0) - \int_0^t k^2 u_k(s)ds + \int_0^t u_k(s)dW(s).$$

We assume that $u_k(0) \neq 0$ for all $k \geq 1$. In Sections 3 and 4 we establish the following result.

Theorem 1.1. *If the solution of equation* (1.2) *is observed in the form* (1.3)*, then the parameter θ can be computed in each of the following ways:*

(E1) $\theta = \lim_{T\to\infty} \left(\frac{1}{k^2 T} \ln \frac{u_k(0)}{u_k(T)} - \frac{1}{2k^2} \right)$ *for every $k \geq 1$;*

(E2) $\theta = \lim_{k\to\infty} \frac{1}{k^2 T} \ln \frac{u_k(0)}{u_k(T)}$ *for every $T > 0$;*

(E3) $\theta = \frac{1}{T(k^2 - n^2)} \ln \frac{u_n(T) u_k(0)}{u_k(T) u_n(0)}$ *for every $T > 0$ and $n \neq k$.*

Both (E1) and (E2) are essentially the same maximum likelihood estimator, but the infinite-dimensional nature of the equation makes it possible to study this estimator in two different asymptotic regimes. (E3) is a closed-form exact estimator. While it is most likely to be the best choice for this particular problem, we show in Section 4 that computational complexity of closed-form exact estimators can dramatically increase with the number of Wiener processes driving the equation, while the complexity of the MLE is almost unaffected by this number. The result is another unexpected feature of closed-form exact estimators: ever though they produce the exact value of the parameter, they are not always the best choice computationally.

2. Stochastic Parabolic Equations

In this section we introduce the diagonalizable stochastic parabolic equation depending on a parameter and study the main properties of the solution.

Let $\mathbf{H}$ be a separable Hilbert space with the inner product $(\cdot,\cdot)_0$ and the corresponding norm $\|\cdot\|_0$. Let Λ be a densely-defined linear operator on $\mathbf{H}$ with the following property: there exists a positive number c such that $\|\Lambda u\|_0 \geq c\|u\|_0$ for every u from the domain of Λ. Then the operator powers Λ^γ, $\gamma \in \mathbb{R}$, are well defined and generate the spaces $\mathbf{H}^\gamma$: for $\gamma > 0$, $\mathbf{H}^\gamma$ is the domain of Λ^γ; $\mathbf{H}^0 = \mathbf{H}$; for $\gamma < 0$, $\mathbf{H}^\gamma$ is the completion of $\mathbf{H}$ with respect to the norm $\|\cdot\|_\gamma := \|\Lambda\cdot\|_0$ (see for instance Krein at al. [15]). By construction, the collection of spaces $\{\mathbf{H}^\gamma,\ \gamma \in \mathbb{R}\}$ has the following properties:

– $\Lambda^\gamma(\mathbf{H}^r) = \mathbf{H}^{r-\gamma}$ for every $\gamma, r \in \mathbb{R}$;

– For $\gamma_1 < \gamma_2$ the space $\mathbf{H}^{\gamma_2}$ is densely and continuously embedded into $\mathbf{H}^{\gamma_1}$: $\mathbf{H}^{\gamma_2} \subset \mathbf{H}^{\gamma_1}$ and there exists a positive number c_{12} such that $\|u\|_{\gamma_1} \le c_{12}\|u\|_{\gamma_2}$ for all $u \in \mathbf{H}^{\gamma_2}$;

– for every $\gamma \in \mathbb{R}$ and $m > 0$, the space $\mathbf{H}^{\gamma-m}$ is the dual of $\mathbf{H}^{\gamma+m}$ relative to the inner product in $\mathbf{H}^\gamma$, with duality $\langle\cdot,\cdot\rangle_{\gamma,m}$ given by

$$\langle u_1, u_2\rangle_{\gamma,m} = (\Lambda^{\gamma-m}u_1, \Lambda^{\gamma+m}u_2)_0, \text{ where } u_1 \in \mathbf{H}^{\gamma-m},\ u_2 \in \mathbf{H}^{\gamma+m}.$$

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ be a stochastic basis with the usual assumptions, and let $\{W_j,\ j \ge 1\}$ be a collection of independent standard Brownian motions on this basis. Consider the following Itô equation

$$\begin{cases} du(t) + (\mathcal{A}_0 + \theta\mathcal{A}_1)u(t)dt = f(t)dt + \sum\limits_{j\ge1}(\mathcal{M}_j u(t) + g_k(t))dW_j(t),\ 0 < t \le T, \\ u(0) = u_0 \end{cases} \tag{2.1}$$

where $\mathcal{A}_0$, $\mathcal{A}_1$, $\mathcal{M}_j$ are linear operators, f and g_k are adapted process, and θ is a scalar parameter belonging to an open set $\Theta \subset \mathbb{R}$.

Definition 2.1.
(a) Equation (2.1) *is called an* `equation with additive noise` *if $\mathcal{M}_j = 0$ for all $j \ge 1$. Otherwise,* (2.1) *is called an* `equation with multiplicative noise` *(also known as a* `bilinear equation`*).*
(b) Equation (2.1) *is called* `diagonalizable` *if the operators $\mathcal{A}_0$, $\mathcal{A}_1, \mathcal{M}_j$, $j \ge 1$, have a common system of eigenfunctions $\{h_k,\ k \ge 1\}$ such that $\{h_k,\ k \ge 1\}$ is an orthonormal basis in $\mathbf{H}$ and each h_k belongs to every $\mathbf{H}^\gamma$.*
(c) Equation (2.1) *is called* `parabolic in the triple` $(\mathbf{H}^{\gamma+m}, \mathbf{H}^\gamma, \mathbf{H}^{\gamma-m})$ *if*

∘ *the operator $\mathcal{A}_0 + \theta\mathcal{A}_1$ is uniformly bounded from $\mathbf{H}^{\gamma+m}$ to $\mathbf{H}^{\gamma-m}$ for $\theta \in \Theta$: there exists a positive real number C_1 such that*

$$\|(\mathcal{A}_0 + \theta\mathcal{A}_1)v\|_{\gamma-m} \le C_1\|v\|_{\gamma+m} \tag{2.2}$$

for all $\theta \in \Theta$, $v \in \mathbf{H}^{\gamma+m}$;

∘ *There exists a positive number δ and a real number C such that, for every $v \in \mathbf{H}^{\gamma+m}$, $\theta \in \Theta$,*

$$-2\langle(\mathcal{A}_0 + \theta\mathcal{A}_1)v, v\rangle_{\gamma,m} + \sum_{j\geq 1} \|\mathcal{M}_j v\|_\gamma^2 + \delta\|v\|_{\gamma+m}^2 \leq C\|v\|_\gamma^2. \tag{2.3}$$

Remark 2.2. (a) Note that (2.2) and (2.3) imply uniform continuity of the family of operators $\mathcal{M}_j$, $j \geq 1$ from $\mathbf{H}^{\gamma+m}$ to $\mathbf{H}^\gamma$; in fact,

$$\sum_{j\geq 1} \|\mathcal{M}_j v\|_\gamma^2 \leq 2C_1\|v\|_{\gamma+m}^2 + C\|v\|_\gamma^2.$$

(b) If equation (2.1) is parabolic, then condition (2.3) implies that

$$\langle(2\mathcal{A}_0 + 2\theta\mathcal{A}_1 + CI)v, v\rangle_{\gamma,m} \geq \delta\|v\|_{\gamma+m}^2,$$

where I is the identity operator. The Cauchy-Schwartz inequality and the continuous embedding of $\mathbf{H}^{\gamma+m}$ into $\mathbf{H}^\gamma$ then imply

$$\|(2\mathcal{A}_0 + 2\theta\mathcal{A}_1 + CI)v\|_\gamma \geq \delta_1\|v\|_\gamma$$

for some $\delta_1 > 0$ uniformly in $\theta \in \Theta$. As a result, we can take $\Lambda = (2\mathcal{A}_0 + 2\theta^*\mathcal{A}_1 + CI)^{1/(2m)}$ for some fixed $\theta^* \in \Theta$.

From now on, if equation (2.1) *is parabolic and diagonalizable, we will assume that the operator* Λ *has the same eigenfunctions as the operators* $\mathcal{A}_0,\ \mathcal{A}_1,\ \mathcal{M}_j$*; by Remark 2.2, this leads to no loss of generality.*

Example 1. (a) For $0 < t \leq T$ and $x \in (0, 1)$, consider the equation

$$du(t,x) - \theta\, u_{xx}(t,x)dt = u_x(t,x)dW(t) \tag{2.4}$$

with periodic boundary conditions; $u_x = \partial u/\partial x$. Then $\mathbf{H}^\gamma$ is the Sobolev space on the unit circle (see, for example, Shubin [26, Section I.7]) and $\Lambda = \sqrt{I - \mathbf{\Delta}}$, where $\mathbf{\Delta}$ is the Laplace operator on $(0, 1)$ with periodic boundary conditions. Direct computations show that equation (2.4) is diagonalizable; it is parabolic if and only if $2\theta > 1$.

(b) Let G be a smooth bounded domain in $\mathbb{R}^d$. Let $\mathbf{\Delta}$ be the Laplace operator on G with zero boundary conditions. It is known (for example, from Shubin [26]), that

1. the eigenfunctions $\{h_k,\ k \geq 1\}$ of $\mathbf{\Delta}$ are smooth in G and form an orthonormal basis in $L_2(G)$;

2. the corresponding eigenvalues $\sigma_k,\ k \geq 1$, can be arranged so that $0 < -\sigma_1 \leq -\sigma_2 \leq \ldots$, and there exists a number $c > 0$ such that $|\sigma_k| \sim ck^{2/d}$, that is,

$$\lim_{k\to\infty} |\sigma_k| k^{-2/d} = c.$$

We take $\mathbf{H} = L_2(G)$, $\Lambda = \sqrt{I - \mathbf{\Delta}}$, where I is the identity operator. Then $\|\Lambda u\|_0 \geq \sqrt{1-\sigma_1}\|u\|_0$ and the operator Λ generates the Hilbert spaces $\mathbf{H}^\gamma$, and, for every $\gamma \in \mathbb{R}$, the space $\mathbf{H}^\gamma$ is the closure of the set of smooth compactly supported function on G with respect to the norm

$$\left(\sum_{k\geq 1}(1+k^2)^\gamma |\varphi_k|^2\right)^{1/2}, \quad \text{where } \varphi_k = \int_G \varphi(x) h_k(x)dx,$$

which is an equivalent norm in $\mathbf{H}^\gamma$. Let θ and σ be real numbers. Then the stochastic equation

$$du - \theta \mathbf{\Delta} u dt = \Lambda u\, dW \tag{2.5}$$

is

– always diagonalizable;

– parabolic in $(\mathbf{H}^{\gamma+1}, \mathbf{H}^\gamma, \mathbf{H}^{\gamma-1})$ for every $\gamma \in \mathbb{R}$ if and only if $2\theta > 1$.

Indeed, we have $\mathcal{A}_0 = 0$, $\mathcal{A}_1 = -\mathbf{\Delta}$, $\mathcal{M}_1 = \Lambda$, $\mathcal{M}_j = 0$, $j \geq 2$, and

$$-2\theta\langle \mathcal{A}_1 v, v\rangle_{\gamma,1} = -2\theta\|v\|^2_{\gamma+1} + 2\theta\|u\|^2_\gamma,$$

and so (2.3) holds with $\delta = 2\theta - 1$ and $C = 2\theta$.

Remark 2.3. Taking in (2.1) $\mathbf{H} = L_2(G)$, where G is a smooth bounded domain in $\mathbb{R}^d$, and $\mathcal{A}_0 = -\mathbf{\Delta}$, $\mathcal{A}_1 = I$, $\mathcal{M}_j u = h_k(x)u(x)$, $g_k = h_k(x)g(t,x)$, we get a bilinear equation driven by space-time white noise. Direct analysis shows that this equation is not diagonalizable. Moreover, the equation is parabolic if and only if $d = 1$, that is, when G is an interval; for details, see the lecture notes by Walsh [27].

For a diagonalizable equation, the parabolicity condition (2.3) can be expressed in terms of the eigenvalues of the operators in the equation.

Theorem 2.4. *Assume that equation* (2.1) *is diagonalizable, and*

$$\mathcal{A}_0 h_k = \rho_k h_k,\ \mathcal{A}_1 h_k = \nu_k h_k,\ \mathcal{M}_j h_k = \mu_{jk} h_k.$$

With no loss of generality (see Remark 2.2)*, we also assume that*

$$\Lambda h_k = \lambda_k h_k.$$

Then equation (2.1) *is parabolic in the triple* $(\mathbf{H}^{\gamma+m}, \mathbf{H}^{\gamma}, \mathbf{H}^{\gamma-m})$ *if and only if there exist positive real numbers* δ, C_1 *and a real number* C_2 *such that, for all* $k \geq 1$ *and* $\theta \in \Theta$*,*

$$\lambda_k^{-2m} |\rho_k + \theta \nu_k| \leq C_1; \tag{2.6}$$

$$-2(\rho_k + \theta\nu_k) + \sum_{j\geq 1} |\mu_{jk}|^2 + \delta \lambda_k^{2m} \leq C_2. \tag{2.7}$$

Proof. We show that, for a diagonalizable equation, (2.6) is equivalent to (2.2) and (2.7) is equivalent to (2.3). Indeed, note that for every $\gamma,\ r \in \mathbb{R}$,

$$\|h_k\|_{\gamma+r} = \|\Lambda^r h_k\|_\gamma = \lambda_k^r \|h_k\|_\gamma.$$

Then (2.6) is (2.2) with $v = h_k$, and (2.7) is (2.3) with $v = h_k$. Since both (2.6) and (2.7) are uniform in k and the collection $\{h_k,\ k \geq 1\}$ is dense in every $\mathbf{H}^\gamma$, the proof of the theorem is complete. □

The following is the basic existence/uniqueness/regularity result for parabolic equations; for the proof, see Rozovskii [24, Theorem 3.2.1].

Theorem 2.5. *Assume that equation* (2.1) *is parabolic in the triple* $(\mathbf{H}^{\gamma+m}, \mathbf{H}^{\gamma}, \mathbf{H}^{\gamma-m})$ *and*

1. *the initial condition* u_0 *is deterministic and belongs to* $\mathbf{H}^\gamma$*;*
2. *the process* $f = f(t)$ *is* $\mathcal{F}_t$*-adapted with values in* $\mathbf{H}^{\gamma-m}$ *and*
$$\mathbb{E}\int_0^T \|f(t)\|^2_{\gamma-m} dt < \infty;$$
3. *each process* $g_k = g_k(t)$ *is* $\mathcal{F}_t$*-adapted with values in* $\mathbf{H}^\gamma$ *and*
$$\sum_{j\geq 1} \mathbb{E}\int_0^T \|g_j(t)\|^2_\gamma < \infty.$$

Then there exists a unique $\mathcal{F}_t$-adapted process $u = u(t)$ with the following properties:

– $u \in L_2(\Omega; L_2((0,T); \mathbf{H}^{\gamma+m}) \bigcap L_2(\Omega; C((0,T); \mathbf{H}^{\gamma}))$;

– *u is a* `solution` *of* (2.1), *that is, the equality*

$$u(t) + \int_0^t (\mathcal{A}_0 + \theta\mathcal{A}_1)u(s)ds = u_0 + \int_0^t f(s)ds + \sum_{j\geq 1}(\mathcal{M}_j u(s) + g_k(s))dW_j(s).$$

holds in $\mathbf{H}^{\gamma-m}$ for all $t \in [0,T]$ on the same set $\Omega' \subset \Omega$ of probability one;

– *There exists a positive real number C_0 depending only on T and the numbers C, δ in* (2.3) *such that*

$$\mathbb{E}\sup_{0<t<T}\|u(t)\|_\gamma^2 + \mathbb{E}\int_0^T \|u(t)\|_{\gamma+m}^2 dt \leq C_0\left(\|u_0\|_\gamma^2 + \mathbb{E}\int_0^T \|f(t)\|_{\gamma-m}^2 dt + \sum_{j\geq 1}\mathbb{E}\int_0^T \|g_j(t)\|_\gamma^2\right)$$

Corollary 2.6. Assume that equation (2.1) is parabolic and diagonalizable. Then, under the assumptions of Theorem 2.5 we have

$$u(t) = \sum_{k=1}^\infty u_k(t)h_k \text{ and } \sum_{k=1}^\infty \lambda_k^{2\gamma}\mathbb{E}|u_k(t)|^2 < \infty,\ t \in [0,T], \tag{2.8}$$

where $u_k(t) = (\Lambda^\gamma u(t), h_k)_0$ satisfies

$$du_k(t) = \Big((\rho_k + \theta\nu_k)u_k(t) + f_k(t)\Big)dt + \sum_{j=1}^n (\mu_{jk}u_k(t) + g_k(t))dW_j(t), \tag{2.9}$$

with $u_k(0) = (\Lambda^\gamma u_0, h_k)_0,\ f_k(t) = \langle \Lambda^\gamma f(t), h_k\rangle_{0,m},\ g_k(t) = (\Lambda^\gamma g(t), h_k)_0$.

3. Maximum Likelihood Estimators

With $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$, $W_j,\ j \geq 1$, and $\{\mathbf{H}^r,\ r \in \mathbb{R}\}$ as in the previous section, consider the stochastic Itô equation

$$du(t) + (\mathcal{A}_0 + \theta\mathcal{A}_1)u(t)dt = \sum_{j\geq 1}\mathcal{M}_j u(t)dW_j(t),\ 0 < t \leq T,\ u(0) = u_0. \tag{3.1}$$

We assume that

- equation (3.1) is parabolic in the triple $(\mathbf{H}^{\gamma+m};\mathbf{H}^{\gamma},\mathbf{H}^{\gamma-m})$ for some $\gamma\in\mathbb{R}$, $m>0$;
- equation (3.1) is diagonalizable;
- $u_0\in\mathbf{H}^{\gamma}$.
- The solution of (3.1) is observed (can be measured without errors) for all $t\in[0,T]$.

The objective is to estimate the real number θ from the observations $u(t),\ t\in[0,T]$.

Even though whole random field u can be observed, the actual computations can be performed only on a finite-dimensional projection of u. By Corollary 2.6, we have

$$u(t)=\sum_{k=1}^{\infty}u_k(t)h_k, \tag{3.2}$$

$$u_k(t)+\int_0^t(\rho_k+\theta\nu_k)u_k(s)ds=(\Lambda^{\gamma}u_0,h_k)_0+\int_0^t u_k(s)\sum_{j\geq 1}\mu_{jk}dW_j(s), \tag{3.3}$$

Thus, a finite collection of the Geometric Brownian motions u_k is a natural finite-dimensional projection of u.

To simplify certain formulas, we will use the following notations:

$$M_k=\sum_{j\geq 1}|\mu_{jk}|^2,\ \eta_k=\frac{M_k}{\nu_k^2}. \tag{3.4}$$

3.1. Maximum Likelihood Estimator (MLE)

Let $u_{k_1},\dots,u_{k_N}$ be a finite collection of diffusion processes (3.3). For each $\theta\in\Theta$, the vector $U_N=(u_{k_1},\dots,u_{k_N})$ generates a measure on the space of continuous $\mathbb{R}^N$-valued functions. If these measures are absolutely continuous with respect to some convenient reference measure, then the MLE of θ will be the value maximizing the corresponding density given the observations. The choice of the reference measure is dictated, among other factors, by the possibility to find a closed-form expression of the density. For diffusion processes with a parameter in the drift, the standard choice is the measure generated by the process with a fixed value of the parameter, for example, the true value θ_0. Analysis of the relevant conditions for mutual absolute continuity, as given, for example, in the book by Liptser and Shiryaev [16, Theorem 7.16], demonstrates that

- if $N{=}1$, then the measures generated by u_k for different values of θ are mutually absolutely continuous, and the density with respect to the measure corresponding to the true parameter θ_0 is

$$L_k(\theta,\theta_0)=\exp\Big\{-\int_0^T\frac{\nu_k(\theta-\theta_0)}{M_k}\frac{du_k}{u_k}-\frac{\rho_k\nu_k(\theta-\theta_0)\,T}{M_k}-\frac{(\nu_k)^2(\theta^2-\theta_0^2)\,T}{2M_k}\Big\}. \quad (3.5)$$

- For $N>1$, the measures are typically mutually singular and so is the resulting estimation problem. We will see later how to exploit this singularity and gain a computational advantage over the straightforward MLE.

Thus, observation of a single process $u_k(t),\ 0\le t\le T$, provides an MLE $\widehat{\theta}_k$ of θ; by (3.5),

$$\widehat{\theta}_k=-\frac{1}{\nu_k T}\int_0^T\frac{du_k}{u_k}-\frac{\rho_k}{\nu_k}. \quad (3.6)$$

By Itô's Lemma,

$$d\ln(u_k)=\frac{du_k}{u_k}-\frac{1}{2}M_k dt,$$

and hence from (3.6) we get

$$\widehat{\theta}_k=\frac{1}{\nu_k T}\ln\frac{u_k(0)}{u_k(T)}-\frac{M_k}{2\nu_k}. \quad (3.7)$$

Notice that, by uniqueness of solution of equation (3.3), the function $u_k(t)$ cannot change sign and so $u_k(0)/u_k(T)>0$. From (3.6) and (3.3) we have the following alternative representation of the MLE:

$$\widehat{\theta}_k=\theta_0-\frac{1}{\nu_k T}\sum_{j\ge 1}\mu_{jk}W_j(T); \quad (3.8)$$

in particular,

$$\mathbb{E}(\widehat{\theta}_k-\theta_0)^2=\frac{\eta_k}{T} \quad (3.9)$$

and $\sqrt{T/\eta_k}\,(\widehat{\theta}_k-\theta_0)$ is a standard Gaussian random variable *for every* $T>0$ *and* $k\ge 1$.

All properties of the MLE (3.7) now follow directly from (3.8) and (3.9) and are summarized below.

Theorem 3.1. *Assume that equation* (3.1) *is diagonalizable, parabolic in the triple* $(\mathbf{H}^{\gamma+m};\mathbf{H}^{\gamma},\mathbf{H}^{\gamma-m})$ *for some* $\gamma\in\mathbb{R}$, $m>0$, *and* $u_0\in\mathbf{H}^{\gamma}$. *Then*

1. *For every $k \geq 1$ and $T > 0$, $\widehat{\theta}_k$ is an unbiased estimator of θ_0.*

2. *For every $k \geq 1$, as $T \to \infty$, $\widehat{\theta}_k$ converges to θ_0 with probability one and $\sqrt{T}(\widehat{\theta}_k - \theta_0)$ converges in distribution to a Gaussian random variable with zero mean and variance η_k.*

3. *If, in addition,*
$$\lim_{k \geq 1} \eta_k = 0, \tag{3.10}$$
then, for every $T > 0$, as $k \to \infty$, $\widehat{\theta}_k$ converges to θ_0 with probability one and $(\widehat{\theta}_k - \theta_0)/\sqrt{\eta_k}$ converges in distribution to a Gaussian random variable with zero mean and variance $1/T$.

Remark 3.2. Conditions (2.7) and (3.10) are, in general, not connected. Indeed, let $\Lambda = \sqrt{I - \mathbf{\Delta}}$, where $\mathbf{\Delta}$ is the Laplace operator on a smooth bounded domain in $\mathbb{R}^d$ with zero boundary conditions. Then equation

$$du - (\mathbf{\Delta} u - \theta u)udt = \Lambda u dW(t)$$

satisfies (2.3), but does not satisfy (3.10): in this case, $\lim_{k\to\infty} \eta_k = \infty$. Similarly, equation

$$du - (\theta \mathbf{\Delta} u - u)dt = (I - \mathbf{\Delta})^{3/4} u dW(t)$$

does not satisfy (2.3) for any θ, but satisfies (3.10). We remark that the solution of this last equation can be constructed in special weighted Wiener chaos spaces that are much larger than $L_2(\Omega; L_2((0,T); \mathbf{H}^\gamma))$; see [20].

Example 2. Let us consider the following modification of equation (2.5) from Example 1(b):
$$du - (\mathbf{\Delta} u + \theta u)dt = \sum_{j\geq 1} (1 - \mathbf{\Delta})^{-j/2}\, u\, dW_j(t) \ .$$
We have $\nu_k = 1$, $\rho_k = -\sigma_k > 0$, where σ_k are the eigenvalues of $\mathbf{\Delta}$, and so $\rho_k \sim ck^{2/d}$; $\mu_{jk} = (1 + \rho_k)^{-j}$ and
$$M_k = \sum_{j\geq 1} \frac{1}{(1+\rho_k)^j} = \frac{1}{\rho_k} \to 0, \ k \to \infty.$$
By Theorem 3.1 the maximum likelihood estimator $\widehat{\theta}_k$ of θ is
$$\widehat{\theta}_k = \frac{1}{T} \ln \frac{u_k(0)}{u_k(T)} + \frac{1}{2\sigma_k}$$

and

$$\mathbb{E}(\widehat{\theta}_k - \theta_0)^2 \sim cT^{-1}k^{-2/d}$$

3.2. Modifications of the MLE

By Theorem 3.1, the MLE (3.7) can be consistent and asymptotically normal either in the limit $T \to \infty$ or in the limit $k \to \infty$. An increase of T always improves the quality of the estimator by reducing the variance; if (3.10) holds, then the variance of the estimator can be further reduced by using u_k with the largest available value k.

The natural question is whether the quality of the estimator can be improved even more by using more than one process u_k. This question is no longer of statistical nature: as equation (3.3) shows, each u_k contains essentially the same stochastic information. More precisely, the sigma-algebra generated by each $u_k(t),\ t \in [0,T]$ coincides with the sigma-algebra generated by $\mu_{jk}W_j(t),\ j \geq 1,\ t \in [0,T]$ (some of μ_{jk} can, in principle, be zeroes). Moreover, as was mentioned above, the statistical estimation model for θ, involving two or more processes u_k, is singular. In what follows, we will see how to use this singularity to gain computational advantage over (3.7).

The problem can now be stated as follows: given a sequence of numbers $\widehat{\theta}_k$ such that $\lim_{k\to\infty} \widehat{\theta}_k = \theta_0$, can we transform it into a sequence $\widetilde{\theta}_k$ such that

$$\lim_{k\to\infty} \widetilde{\theta}_k = \theta_0,\ \limsup_{k\to\infty} \frac{|\widetilde{\theta}_k - \theta_0|}{|\widehat{\theta}_k - \theta_0|} < 1. \tag{3.11}$$

If (3.11) holds, it is natural to say that $\widetilde{\theta}_k$ converges to θ_0 faster than $\widehat{\theta}_k$. Accelerating the convergence of a sequence is a classical problem in numerical analysis. The main features of this problem are (a) There are many different methods to accelerate the convergence, and (b) the effectiveness of every method varies from sequence to sequence.

We will investigate two methods:

1. Weighted averaging;

2. Aitken's $\triangle^2$ method.

Theorem 3.3 (Weighted averaging). *Let $\beta_k,\ k \geq 1$, be a sequence of non-negative numbers and*

$$\sum_{k\geq 1} \beta_k = +\infty.$$

Define the weighted averaging estimator $\widehat{\theta}_{(N)}$ by

$$\widehat{\theta}_{(N)} = \frac{\sum\limits_{k=1}^{N} \beta_k \widehat{\theta}_k}{\sum\limits_{k=1}^{N} \beta_k}. \tag{3.12}$$

Then

1. *For every $N \geq 1$ and $T > 0$, $\widehat{\theta}_{(N)}$ is an unbiased estimator of θ_0.*

2. *For every $N \geq 1$, as $T \to \infty$, $\widehat{\theta}_{(N)}$ converges to θ_0 with probability one and $\sqrt{T}(\widehat{\theta}_{(N)} - \theta_0)$ converges in distribution to a Gaussian random variable with zero mean and variance*
$$V_N = \sum_{j\geq 1} \left(\frac{\sum_{k=1}^{N} (\beta_k \mu_{jk}/\nu_k)}{\sum_{k=1}^{N} \beta_k} \right)^2. \tag{3.13}$$

3. *If, in addition,* (3.10) *holds then, for every $T > 0$, as $N \to \infty$, $\widehat{\theta}_{(N)}$ converges to θ_0 with probability one.*

Proof. By (3.8),

$$\widehat{\theta}_{(N)} = \theta_0 + \frac{\sum\limits_{j\geq 1} \left(\sum\limits_{k=1}^{N} (\beta_k \mu_{jk}/\nu_k) \right) W_j(T)}{T \sum\limits_{k=1}^{N} \beta_k}, \tag{3.14}$$

from which the first two statement of the theorem follow. For the last statement, we combine (3.12) with the Toeplitz lemma: if $\lim_{k\to\infty} a_k = a$ and $\beta_k > 0$, then

$$\lim_{N\to\infty} \frac{\sum_{k=1}^{N} \beta_k a_k}{\sum_{k=1}^{N} \beta_k} = a.$$

□

The behavior of V_N/η_N, as $N \to \infty$ can be just about anything. Take $\rho_k = 0$, $\mu_{jk} = 0$, $j > 1$. Then,

– With $\beta_k = 1/k$, $\nu_k = k^2$, and $\mu_{1k} = k$, we get $\eta_N = 1/N^2$ and

$$\frac{V_N}{\eta_N} \sim \frac{\alpha N^2}{\ln^2 N} \to \infty,\ N \to \infty.$$

for some $\alpha > 0$; recall that, for $a_n, b_n > 0$, notation $a_n \sim b_n$ means

$$\lim_{n\to\infty}(a_n/b_n) = 1.$$

– With $\beta_k = k$, $\nu_k = k^2$, and $\mu_{1k} = k$, we get $\eta_N = 1/N^2$ and

$$V_N \sim 4\eta_N > \eta_N$$

– With $\beta_k = 1$, $\nu_k = k^2$, and $\mu_{1k} = (-1)^k k$, we get $\eta_N = 1/N^2$ and

$$V_N \sim (\ln^2 2)\eta_N < \eta_N$$

– With $\beta_k = 1$, $\nu_k = k$, and $\mu_{1k} = (-1)^k\sqrt{k}$, we get $\eta_N = 1/N$ and

$$\frac{V_N}{\eta_N} \sim \frac{\beta}{N} \to 0,\ N \to \infty.$$

Next, we consider **Aitken's $\triangle^2$ method**. This method consists in transforming a sequence $A = \{a_n,\ n \geq 1\}$ to a sequence

$$b_n(A) = a_n - \frac{(a_{n+1} - a_n)^2}{a_{n+2} - 2a_{n+1} + a_n}.$$

The main result concerning this method is that if $\lim_{n\to\infty} a_n = a$ and

$$\lim_{n\to\infty} \frac{|a_{n+1} - a|}{|a_n - a|} = \lambda \in (0,1), \tag{3.15}$$

then $\lim_{n\to\infty} b_n(A) = a$ and

$$\lim_{n\to\infty} \frac{|b_n(A) - a|}{|a_n - a|} = 0.$$

That is, the sequence $b_n(A)$ converges to the same limit a but faster.

Accordingly, under the condition (3.10), we define

$$\widetilde{\theta}_k = \widehat{\theta}_k - \frac{(\widehat{\theta}_{k+1} - \widehat{\theta}_k)^2}{\widehat{\theta}_{k+2} + 2\widehat{\theta}_{k+1} - \widehat{\theta}_k}, \tag{3.16}$$

with a hope that

$$\lim_{k\to\infty} \frac{\mathbb{E}(\widetilde{\theta}_k - \theta_0)^2}{\mathbb{E}(\widehat{\theta}_k - \theta_0)^2} < 1. \tag{3.17}$$

In general, there is no guarantee that this will be the case because typically $\eta_k \sim \alpha k^{-\delta}$ for some $\alpha > 0$ and $\delta > 0$, and so, if we set

$$a_k = \mathbb{E}(\widehat{\theta}_k - \theta_0)^2,$$

we get by Theorem 3.1

$$\lim_{n\to\infty} \frac{|a_{n+1} - a|}{|a_n - a|} = 1.$$

Direct investigation of the sequence $\widetilde{\theta}_k$ is possible if there is only one Wiener process $W = W(t)$ driving the equation, that is, $\mu_{jk} = 0$ for $j \geq 2,\ k \geq 1$. In this case, (3.8) shows that

$$\widetilde{\theta}_k = \theta_0 + \frac{W(T)}{T}\left(r_k - \frac{(r_{k+1} - r_k)^2}{r_{k+2} - 2r_{k+1} + r_k}\right), \tag{3.18}$$

where $r_k = \mu_{1k}/\nu_k$. Then direct computations show that

– if $r_k \sim \alpha k^{-\delta}$, α, $\delta > 0$, then

$$\frac{\mathbb{E}(\widetilde{\theta}_k - \theta_0)^2}{\mathbb{E}(\widehat{\theta}_k - \theta_0)^2} \sim \frac{1}{(\delta + 1)^2}.$$

– if $r_k = (-1)^k/k$, then

$$\frac{\mathbb{E}(\widetilde{\theta}_k - \theta_0)^2}{\mathbb{E}(\widehat{\theta}_k - \theta_0)^2} \sim \frac{c}{k^2},\ c > 0.$$

For more than one Wiener process, we find

$$\widetilde{\theta}_k = \theta_0 + \frac{\xi_k^2}{\zeta_k},$$

where (ξ_k, ζ_k) is a two-dimensional Gaussian vector with known distribution. The analysis of this estimator, while possible, is technically much more difficult and will require many additional assumptions on μ_{jk}. We believe that this analysis falls outside the scope of this paper, and we present here only some numerical results. We suppose that Fourier coefficients u_k satisfy (3.3) with $\nu_k = k, \rho_k = 0,\ \mu_{jk} = (-1)^k/(k+j)$, the noise term is driven by $n = 10$ Wiener processes, and the true value of the parameter $\theta_0 = 1$. From (3.7) we note that the estimates $\hat{\theta}_k$ can be calculated if we only know the value of $\log(u_k(T)/u_k(0))$, rather than the whole path $u_k(t),\ 0 \leq t \leq T$. Using the closed-form solution of equation (3.3) $u_k(t) = u_k(0)\exp(-(\theta_0\nu_k + \sum_j \mu_{jk}^2/2)t + \sum_j \mu_{jk}W_j(t))$, we simulate $\log(u_k(T)/u_k(0))$ directly, without applying some discretization schemes to the process $u_k(t)$. Three type of estimates are presented in Figure 1. The obtained numerical results are consistent with above theoretical results: Aitken's $\triangle^2$ method performs the best, Weighted Averages Estimates with $\beta_k = k$ perform better than simple estimates.

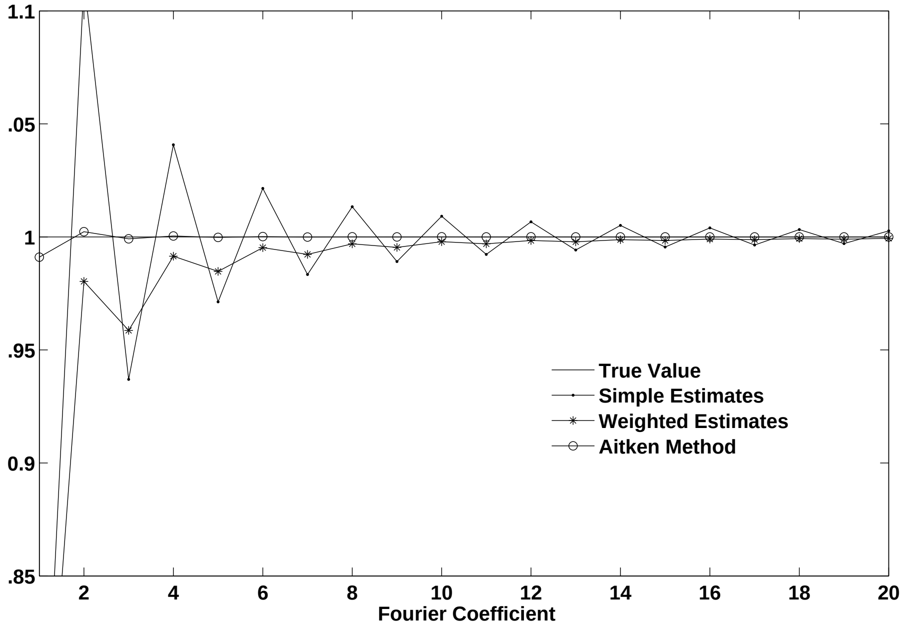

Figure 1. Performances of three type of estimates: Simple, Weighted Averages and Aitken's $\triangle^2$ method

4. Closed-form Exact Estimators

In regular models, the estimator is consistent in the large sample or small noise limit; neither of these limits can be evaluated exactly from any actual observations. In singular models, there often exists an estimator that is consistent in the limit that can potentially be evaluated exactly from the available observations. Still, no expression can be evaluated on a computer unless the expression involves only finitely many operations of addition, subtraction, multiplication, and division.

Definition 4.1. *An estimator is called* `closed-form exact` *if it produces the exact value of the unknown parameter after a finite number of additions, subtractions, multiplications, and divisions performed on the elementary functions of the observations.*

Closed-form exact estimators exist for the model (3.1) *if we assume that the observations are* $u_k(t),\ k\geq 1,\ t\in[0,T]$.

As an illustration, consider the simple example

$$du-\theta u_{xx}dt=(u/2)dt+udW(t),$$

where $x\in(0,\pi)$ and zero boundary conditions are assumed.

With $h_k=\sqrt{2/\pi}\,\sin(kx)$, we find

$$du_k(t)=-k^2\theta u_k(t)dt+(u_k/2)dt+u_k(t)dW(t).$$

Set $v_k(t)=\ln(u_k(t)/u_k(0))$. Then

$$dv_k(t)=-k^2\theta dt+dW(t).$$

In particular,

$$v_1(T)=-\theta T+W(T),\ v_2(T)=-4\theta T+W(t)$$

so that

$$\theta=\frac{v_1(T)-v_2(T)}{3T}$$

or

$$\theta=\frac{1}{3T}\ln\frac{u_1(T)u_2(0)}{u_1(0)u_2(T)}. \tag{4.1}$$

Notice that given $u_1,\dots,u_N$, we have $N(N-1)/2$ exact estimators of this type.

If there are two Wiener processes driving the equation, then we will need three different u_k to construct an estimator of the type (4.1). The general result is as follows.

Theorem 4.2. *In addition to conditions of Theorem 3.1 assume that there exist two finite sets of indices* $(k^i_1,k^i_2,\dots,k^i_n),\ i=1,2,$ *and a positive integer* p *such that*

$$\sum_{\ell=1}^{n}\nu_{k^1_\ell}\mu_{pk^2_\ell}\neq\sum_{\ell=1}^{n}\nu_{k^2_\ell}\mu_{pk^1_\ell}.$$

Then there exists a closed-form exact estimator for θ*.*

Proof. Let $v_k(t)=\ln(u_k(t)/u_k(0))$. From (3.3), by Itô's formula, we get

$$dv_k=-\Big(\rho_k+\theta\nu_k+\frac{1}{2}M_k\Big)dt+\sum_{j\geq 1}\mu_{jk}dW_j(t), \tag{4.2}$$

and then

$$\theta \sum_{\ell=1}^{n} \Big(\nu_{k^2_\ell} \mu_{pk^1_\ell} - \nu_{k^1_\ell} \mu_{pk^2_\ell} \Big) = \frac{1}{T} \sum_{\ell=1}^{n} \Big(\mu_{pk^2_\ell} v_{k^1_\ell}(T) - \mu_{pk^1_\ell} v_{k^2_\ell}(T) + \mu_{pk^2_\ell} (\rho_{k^1_\ell} + \frac{1}{2} M_{k^1_\ell}) - \mu_{pk^1_\ell} (\rho_{k^2_\ell} + \frac{1}{2} M_{k^2_\ell}) \Big), \tag{4.3}$$

which completes the proof. □

If there are n_0 Wiener processes driving the equation, then the extra condition of the theorem can always be ensured with $n = n_0 + 1$, because every collection of n vectors in an $n-1$-dimensional space is linearly dependent. While relation (4.3) gives a closed-form exact estimator, the resulting formulas can be rather complicated when the number of Wiener processes in the equation is large; if this number is infinite, then the estimator might not exist at all. For comparison, the complexity of the maximum likelihood estimator (3.7) does not depend on the number of Wiener processes in the equation. As a result, when it comes to actual computations, the closed-form exact estimator is not necessarily the best choice. On the other hand, the very existence of such an estimator is rather remarkable.

We conclude this section with three examples of closed-form exact estimators. The first example shows that such estimators can exist for equations that are not diagonalizable in the sense of Definition 2.1.

Example 3. Consider the equation

$$du(t,x) = \theta u_{xx}(t,x)dt + u(t,x)dW(t),\ 0 < t \le T,\ x \in \mathbb{R}.$$

By the Itô formula,

$$u(t,x) = v(t,x) \exp(W(t) - (t/2)),$$

where v solves the heat equation $v_t = \theta v_{xx}$, $v(0,x) = u(0,x)$. Assume that $u(0,x)$ is a smooth compactly supported function. Then $u(t,x)$ is a smooth bounded function for all $t > 0, x \in \mathbb{R}$ and $\mathbb{E} \int_{\mathbb{R}} |u(t,x)|^p dx < \infty$ for all $p > 0$, $t \ge 0$. In particular, the Fourier transform $U(t,y)$ of u is defined and satisfies

$$dU(t,y) = -\theta y^2 U(t,y)dt + U(t,y)dW(t).$$

Let $V(t) = \ln(U(t)/U(0))$. Then

$$V(T,y) = -y^2\theta T - (T/2) + W(T),$$

and

$$\theta = \frac{V(T, y_1) - V(T, y_2)}{T(y_2^2 - y_1^2)}.$$

The next example shows that conditions (2.3) and (3.10) are not related to the existence of a closed-form exact estimator.

Example 4. Consider the equation

$$du - (\mathbf{\Delta} u + \theta u)dt = (I - \mathbf{\Delta})^{3/4} u dW(t)$$

on $(0, \pi)$ with zero boundary conditions. Clearly both (2.3) and (3.10) are not satisfied. While the equation is not parabolic, there exists a unique solution in weighted Wiener chaos spaces, and we can therefore consider

$$du_k = (-k^2 u_k + \theta u_k)dt - (1 + k^2)^{3/4} u_k dW(t).$$

For $v_k(t) = \ln(u_k(t)/u_k(0))$ we find

$$v_k(T) = (-k^2 - \frac{(1 + k^2)^{3/2}}{2})T + \theta T + (1 + k^2)^{3/4} W(t).$$

In particular,

$$v_1(T) = a_1 T + b_1 W(t) + \theta T,\ v_2(T) = a_2 T + b_2 W(T) + \theta T,$$

and so

$$\theta = \frac{b_1 v_2(T) - b_2 v_1(T) - (a_2 b_1 - a_1 b_2)T}{T(b_1 - b_2)}.$$

The last example shows that, as long as there is no spacial structure in the noise, multiplicativity of the noise is not necessary to have a closed-form exact estimator.

Example 5. Consider the equation

$$du(t, x) = \theta u_{xx}(t, x)dt + dW(t), t > 0,\ x \in (0, \pi),$$

with Neumann boundary conditions, so that $h_1 = 1/\sqrt{\pi}$ and $h_k = \sqrt{2/\pi}\cos((k - 1)x)$, $k \geq 2$. Then $du_2(t) = -\theta u_2(t)dt$, and, as long as $u_2(0) \neq 0$, we have

$$\theta = \frac{1}{T} \ln \frac{u_2(0)}{u_2(T)}.$$

References

1. S. I. Aihara (1992) *Regularized maximum likelihood estimate for an infinite-dimensional parameter in stochastic parabolic systems*, SIAM J. Control Optim. **30**(4):745–764.
2. A. Bagchi and V. Borkar (1984) *Parameter identification in infinite-dimensional linear systems*, Stochastics **12**(3-4):201–213.
3. R. Cont (2005) *Modeling term structure dynamics: an infinite dimensional approach*, Int. J. Theor. Appl. Finance **8**(3):357–380.
4. C. Frankignoul (2000) *Sst anomalies, planetary waves and rc in the middle rectitudes*, Reviews of Geophysics **30**(7):1776–1789.
5. J. Gáll, G. Pap, M. C. A. van Zuijlen (2006) *Forward interest rate curves in discrete time settings driven by random fields*, Comput. Math. Appl. **51**(3-4):387–396.
6. R. S. Goldstein (2000) *The term structure of interest rates as random field*, Review of Financial Studies (13):365–384.
7. M. Huebner, S. Lototsky, B. L. Rozovskii (1997) *Asymptotic properties of an approximate maximum likelihood estimator for stochastic PDEs*, Statistics and control of stochastic processes (Moscow, 1995/1996), World Sci. Publishing, pp139–155.
8. M. Huebner, B. Rozovskii, R. Khasminskii (1992) *Two examples of parameter estimation*, In:Stochastic Processes, ed. Cambanis, Chos, Karandikar, Berlin, Springer.
9. M. Huebner, B. L. Rozovskiĭ (1995) *On asymptotic properties of maximum likelihood estimators for parabolic stochastic PDE's*, Probab. Theory Related Fields **103**(2):143–163.
10. I. A. Ibragimov, R. Z. Khas′minskiĭ (1981) *Statistical estimation*, Applications of Mathematics, vol. 16, Springer-Verlag, New York.
11. I. A. Ibragimov, R. Z. Khas′minskiĭ (1998) *Problems of estimating the coefficients of stochastic partial differential equations. I*, Teor. Veroyatnost. i Primenen. **43**(3):417–438.
12. I. A. Ibragimov, R. Z. Khas′minskiĭ (1999) *Problems of estimating the coefficients of stochastic partial differential equations. II*, Teor. Veroyatnost. i Primenen. **44**(3):526–554.
13. I. A. Ibragimov, R. Z. Khas′minskiĭ (2000) *Problems of estimating the coefficients of stochastic partial differential equations. III*, Teor. Veroyatnost. i Primenen. **45**(2):209–235.
14. R. Khasminskii, N. Krylov, N. Moshchuk (1999) *On the estimation of parameters for linear stochastic differential equations*, Probab. Theory Related Fields **113**(3):443–472.
15. S. G. Kreĭn, Yu. Ī. Petunīn, E. M. Semënov (1982) *Interpolation of linear operators*, Translations of Mathematical Monographs, vol. 54, American Mathematical Society, Providence, R.I.
16. R. S. Liptser, A. N. Shiryayev (2000) *Statistics of random processes I. General theory*, 2nd ed., Springer-Verlag, New York.
17. S. V. Lototsky (2003) *Parameter estimation for stochastic parabolic equations: asymptotic properties of a two-dimensional projection-based estimator*, Stat. Inference Stoch. Process. **6**(1):65–87.

18. S. V. Lototsky, B. L. Rozovskii (1999) *Spectral asymptotics of some functionals arising in statistical inference for SPDEs*, Stochastic Process. Appl. **79**(1):69–94.

19. S. V. Lototsky, B. L. Rozovskii (2000) *Parameter estimation for stochastic evolution equations with non-commuting operators*, In:Skorohod's Ideas in Probability Theory, V.Korolyuk, N.Portenko and H.Syta (ed), Institute of Mathematics of National Academy of Sciences of Ukraine, Kiev, Ukraine, 2000, pp271–280.

20. S. V. Lototsky, B. L. Rozovskii (2006) *Wiener chaos solutions of linear stochastic evolution equations*, Ann. Probab. **34**(2):638–662.

21. L. I. Piterbarg (2001/02) *The top Lyapunov exponent for a stochastic flow modeling the upper ocean turbulence*, SIAM J. Appl. Math. **62**(3):777–800 (electronic).

22. L. I. Piterbarg (2005) *Relative dispersion in 2D stochastic flows*, J. Turbul. **6** (2005), Paper 4 (electronic).

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

24. B. L. Rozovskii (1990) *Stochastic evolution systems*, Mathematics and its Applications (Soviet Series), vol. 35, Kluwer Academic Publishers Group, Dordrecht, Linear theory and applications to nonlinear filtering.

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

26. M. A. Shubin (2001) *Pseudodifferential operators and spectral theory*, second ed., Springer-Verlag, Berlin.

27. J. B. Walsh (1986) *An introduction to stochastic partial differential equations*, École d'été de probabilités de Saint-Flour, XIV—1984, Lecture Notes in Math., vol. 1180, Springer, Berlin, 1986, pp265–439.