Acta Applicandae Mathematicae **84:** 279–314, 2004.

Robust Parameter Estimation for Stochastic Differential Equations

THOMAS DECK[1] and THOMAS GORM THETING[2]
[1]*Lehrstuhl für Mathematik V, Fakultät für Mathematik und Informatik, Universität Mannheim, D-68131 Mannheim, Germany. e-mail: deck@math.uni-mannheim.de*
[2]*Department of Mathematics, Norwegian University of Science and Technology, N-7491 Trondheim, Norway. e-mail: tgt@math.ntnu.no*

(Received: 26 January 2004; in final form: 25 October 2004)

Abstract. We consider the estimation of parameters in stochastic differential equations (SDEs). The problem is treated in the setting of nonlinear filtering theory with a degenerate diffusion matrix. A robust stochastic Feynman–Kac representation for solutions of SDEs of Zakai-type is derived. It is verified that these solutions are conditional densities for the conditional measures defined by degenerate filtering problems. We show that the corresponding estimator for the parameters is robust in the following sense: It depends continuously on both the measurement path and on the intensity of the measurement noise. An algorithm based on a Monte-Carlo approach is given for the practical application of the estimator, and numerical results are reported.

Mathematics Subject Classifications (2000): Primary: 62M05, 62M20; secondary: 62F15.

Key words: Zakai equation, robust nonlinear filtering, Feynman–Kac formula, numerics.

1. Introduction

Consider a stochastic process X_t in $\mathbb{R}^d$, which satisfies a stochastic differential equation (SDE) of Itô type with respect to a d-dimensional Brownian motion $W_t^{(1)}$:

$$dX_t = b_\theta(t, X_t)\, dt + \sigma_\theta(t, X_t)\, dW_t^{(1)}. \tag{1.1}$$

The successful application of such equations to real-world phenomena often depends on the ability to identify unknown parameters θ in (1.1), based on observations of the modelled phenomena. Depending on the kind of observations several rather distinct approaches to this problem exist, see, e.g., the recent monograph [40], and also [1, 29, 31]. Here we suppose that X_t can only partially be observed, and moreover the observation errors introduce some additional stochastic noise. Our work consists of two main parts: The *first part* investigates the estimation problem for θ from the mathematical viewpoint. Let $(W_t^{(1)}, W_t^{(2)})_{t\geqslant 0}$ be an $\mathbb{R}^d \times \mathbb{R}^m$-valued Brownian motion (BM) on an probability space $(\Omega_1, \mathcal{F}_1, P_1)$. We assume X is the unique strong solution of (1.1), and X_0 is independent of the Brownian filtration. The drift vector b_θ and the diffusion matrix σ_θ are assumed to be known functions (cf. conditions 5.3), up to an unknown parameter θ from some set $\Theta \subset \mathbb{R}^k$. Moreover, we assume that X is partially observed through the $\mathbb{R}^m$-valued process

$$Y_t = \int_0^t h(X_s)\,\mathrm{d}s + \alpha W_t^{(2)}, \tag{1.2}$$

where h denotes an $\mathbb{R}^m$-valued bounded function. The parameter $\alpha > 0$ reflects the size of the noise in the observations and is related to how the measurements are taken (details are given below). Our estimation problem reads: *Find an optimal estimate* (*in a sense to be specified*) *for θ, given a single observation path* $(Y_t)_{t\in[0,T]}$. To solve this problem we adopt a Bayesian viewpoint, and consider θ as a random variable, denoted θ_0. We augment (1.1) by the additional 'state' equation

$$\mathrm{d}\theta_t = 0, \tag{1.3}$$

which has the trivial solution $\theta_t = \theta_0$ for all $t \geqslant 0$. (This procedure is sometimes called 'state augmentation'.) Then we estimate θ by the expectation value of θ_0, conditioned on an observation path. This estimator is optimal in the mean square sense. We will investigate some of its mathematical properties in detail. In the *second part of this work* we provide a numerical method for the explicit calculation of the estimator, and we present some simulation examples which show that it performs quite well. Our main practical obstacle for the parameter estimation based on nonlinear filtering is to solve the Zakai equation. The numerical solution of this equation has been the subject of intensive research during the last two decades. Many different methods were proposed, cf. [7, 11, 14, 18, 19, 32, 37, 42]. In the present paper we introduce a solution method for the Zakai equation based on the Monte-Carlo simulation of a robust, recursive Feynman–Kac formula. Before we can describe our main results we need to supply some further details about the measurement model and the Bayesian estimator to be used in this work.

The Measurement Model. To measure a continuous process X without errors is practically impossible, and quite often one measures only a real function h of X_t, called the *measurement function* henceforth. The measurement result is then a quantity of the form $h(X_t)+\delta_t$, where δ_t denotes some random error. Let us assume for a moment that measurements are taken at times $t_n = n \cdot T/N$ for $n = 1, \ldots, N$ and that the errors δ_{t_n} are i.i.d. with zero mean (unbiased measurements) and finite variance σ_M^2. Then

$$Y_{t_n} := \sum_{k=1}^{n} h(X_{t_k})\Delta t_k + \sum_{k=1}^{n} \delta_{t_k}\Delta t_k \quad (n = 0, \ldots, N) \tag{1.4}$$

approximates our measurement model (1.2). (For sufficiently large n the second term in (1.4) behaves approximately like a scalar multiple of a Wiener process, by the central limit theorem.) The parameter α is determined by the requirement of equal variances in (1.2) (for $t = t_n$) and in (1.4). We find $\alpha^2 = \sigma_M^2 T/N$. So α becomes smaller as the number of observations N increases. In real applications there is an upper bound on N because for too large N the independence of measurement errors δ_{t_n} breaks down. Thus, for continuous measurements some maximal N has to be chosen to fix α. This choice depends on the characteristics of the measurement device and allows to fix an *approximate* value for α. Consequently, if an estimate is based on (1.2) then *it should depend continuously on* $\alpha > 0$, to make practical sense. We note that the previous arguments remain valid for the case of m measurements $h = (h_1, \dots, h_m)$, and for i.i.d. errors δ_t in $\mathbb{R}^m$ with nondegenerate covariance matrix C_M: One only has to replace $\alpha W_t^{(2)}$ in (1.2) by $\alpha \cdot C \cdot W_t^{(2)}$, with an m-dimensional BM $W_t^{(2)}$ and a suitable $m \times m$-matrix C. The process $Y_t^C := C^{-1} Y_t$ is then of the (m-dimensional) form (1.2) with h replaced by $C^{-1}h$.

The Bayesian Estimator. Subsequently we choose the random vector θ_0 to be independent of the $(W_t^{(1)}, W_t^{(2)})$-filtration, and independent of the initial value X_0. In view of the extension (1.3) we consider the coefficients in (1.1) as functions of three variables: $b(t, x, \theta) := b_\theta(t, x)$ and $\sigma(t, x, \theta) := \sigma_\theta(t, x)$. We estimate θ by $\hat{\theta}_t^\alpha := E[\theta_0 | \mathcal{Y}_t^\alpha]$, where $\mathcal{Y}_t^\alpha = \sigma(Y_s, s \in [0, t])$ denotes the σ-algebra generated by the observation (1.2). Since $Y = (Y_s)_{0 \leqslant s \leqslant t}$ is a continuous process the factorization lemma [38, Proposition 44.1] provides us with a measurable mapping S_t^α from $C([0, t])^m$ to $\mathbb{R}^d$, such that

$$\hat{\theta}_t^\alpha = E[\theta_0 | \mathcal{Y}_t^\alpha] = S_t^\alpha(Y), \quad P_1\text{-a.s.} \tag{1.5}$$

Note that since the estimate $\hat{\theta}_t^\alpha$ is only defined modulo sets of P_1-measure zero, the function S_t^α is not path-wise unique. Therefore, one has to choose a suitable version of S_t^α in order to evaluate (1.5) for a concrete measurement. A satisfying situation arises when one can choose S_t^α to be *continuous*, i.e. for any fixed $\bar{Y} \in C([0, t])^m$ it should hold

$$S_t^\alpha(Y) \to S_t^\alpha(\bar{Y}), \quad \text{as } \|Y - \bar{Y}\|_\infty \to 0. \tag{1.6}$$

In this case small variations in the observations $(Y_s)_{0 \leqslant s \leqslant t}$ result in nearly the same estimate $\hat{\theta}_t^\alpha$. (An estimator without this robustness property is practically useless.) For numerical evaluation of (1.5) it is even better if (1.6) holds for all $Y, \bar{Y} \in B[0, t]^m$ (the space of bounded, measurable, $\mathbb{R}^m$-valued functions on $[0, t]$) because then one can approximate Y uniformly by step functions. In addition to (1.6) our discussion of the measurement model (1.2) shows that the following “modelling robustness” should be satisfied: For any $\alpha_0 \in (0, \infty)$ it should hold

$$S_t^\alpha(Y) \to S_t^{\alpha_0}(Y), \quad \text{as } \alpha \to \alpha_0. \tag{1.7}$$

The main theoretical result of this paper, Theorem 5.2, gives a concrete formula to evaluate $S_t^\alpha(Y)$, and establishes the continuity of $(\alpha, Y) \mapsto S_t^\alpha(Y)$ on $(0, \infty) \times B[0, t]^m$. This covers (1.6) and (1.7).

Remarks. (1) The notion of robustness is used in different meanings in nonlinear filtering theory, cf. [4], and references given there.

(2) To obtain our results we assume sufficiently smooth and bounded coefficients b, σ, h, and also compact support for the law of X_0. But we need not impose nondegeneracy assumptions on the diffusion matrix σ. This is a crucial point because the system (1.1), (1.3) is highly degenerate.

(3) It is to be expected that our results remain valid for certain unbounded cases as well, e.g., for $h(x) = x$. In fact the case of unbounded h is covered by our formula, provided the process X takes its values only in some interval $[a, b]$: In that case we can simply replace h by a bounded function h_0 which coincides with h on $[a, b]$. Example 3 in Section 7 is of this type.

Comparison to Related Approaches. The filtering approach to parameter estimation is a known concept: For linear filtering see [22, 36]. Nonlinear filtering in discrete time based on a stochastic representation formula is discussed in [15] (with regular diffusion matrix). In continuous time drift estimates are discussed in [10, 20], and in [30] a parameter in the measurement equation is estimated. A standard estimation tool in the engineering literature is to combine state augmentation with the extended Kalman filter. But only few theoretical results are available, and practically this approach only works if the parameter values are known in advance with some precision, otherwise the filter estimates usually diverge in time.

The paper is organized as follows: Section 2 recalls the required background on nonlinear filtering theory and on a stochastic FK-formula. In Section 3 a robust version of the FK-formula is given, and in the following section it is shown that this formula defines a density for the conditional measure of our estimation problem. In Section 5 the continuity results (1.6) and (1.7) are verified. Section 6 discusses the numerical approximation of (1.5), and the final section provides simulation studies for specific estimation problems: a linear problem, and two nonlinear ones.

2. Preparations and Notation

Throughout this paper all probability spaces $(\Omega, \mathcal{F}, P)$ are assumed to be complete, and random variables are real-valued. We denote by P_X the distribution of a random variable X, that is $P_X(B) = P(X \in B)$, and by E_P the expectation with respect to P. In addition we let $C^k(U)$ denote the space of k-times continuously differentiable real-valued functions on $U \subset \mathbb{R}^n$, we use $C_c^k(U)$ to denote functions in $C^k(U)$ with compact support in U, and we let $C_b^k(U)$ denote functions in $C^k(U)$ with bounded derivatives up to order k. Additional notation will be defined as needed.

Let us first provide some background and notations from filtering theory. Let $(\Omega_1, \mathcal{F}_1, P_1)$, $(W^{(1)}, W^{(2)})$ and X_0 be given as in Section 1, and assume that P_{X_0} has Lebesgue density p_0. Moreover, let X be a stochastic process in $\mathbb{R}^d$ satisfying

$$\mathrm{d}X_t = b(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}W_t^{(1)}, \tag{2.1}$$

with initial condition X_0, and with b and σ denoting vector and diffusion matrix, respectively. Later we will impose conditions securing the existence and uniqueness of strong solutions to (2.1). Let $h \in C_b(\mathbb{R}^d)^m$ and assume that X is observed through the stochastic process

$$Y_t = \int_0^t h(X_s)\,\mathrm{d}s + W_t^{(2)}, \quad t \in [0, T]. \tag{2.2}$$

The nonlinear filtering problem for (2.1)–(2.2) is to find $\pi_t(f) := E_{P_1}[f(X_t)|\mathcal{Y}_t]$ for all bounded measurable functions f, where $\mathcal{Y}_t = \sigma(Y_s, 0 \leqslant s \leqslant t)$. That is, $\pi_t(f)$ denotes the best, in mean square sense, estimate for $f(X_t)$, given $\mathcal{Y}_t$. We follow here the so-called reference measure approach to this problem, which leads to the Zakai equation for the unnormalized conditional density [21, 26, 28, 31, 37, 43, 44]: Let $\mathcal{F}_t$ denote the filtration generated by $(W^{(1)}, W^{(2)})$, and consider the exponential martingale

$$\Lambda_t := \exp\left\{\int_0^t h(X_s)\,\mathrm{d}Y_s - \frac{1}{2}\int_0^t h(X_s)^2\,\mathrm{d}s\right\}, \quad \text{for } t \in [0, T].$$

By Girsanov's Theorem $(Y_t)_{t\in[0,T]}$ is a Wiener process on $(\Omega_1, \mathcal{F}_1, \overline{P}_1)$, where $\mathrm{d}\overline{P}_1 = \Lambda_T\,\mathrm{d}P_1$. Note that the measures P_1 and $\overline{P}_1$ are equivalent, so P_1-a.s. and $\overline{P}_1$-a.s. mean the same. For $t \in [0, T]$ and for each bounded measurable function f on $\mathbb{R}^d$ define $\rho_t(f) := E_{\overline{P}_1}[f(X_t)\Lambda_t|\mathcal{Y}_t]$. Then

$$\rho_t(f) = \rho_t(1)\pi_t(f), \tag{2.3}$$

and there is a stochastic process $(\rho_t)_{0\leqslant t\leqslant T}$ on $(\Omega_1, \mathcal{F}_1, P_1)$ with values in the space of finite measures over $\mathbb{R}^d$ such that for every $f \in C_c^\infty(\mathbb{R}^d)$ and $t \in [0, T]$ we have

$$\rho_t(f)(\omega_1) = \int_{\mathbb{R}^d} f(x)\rho_t(\omega_1, \mathrm{d}x), \quad P_1\text{-a.s.},$$

cf. [28]. Moreover, the following Zakai-equation holds under fairly general conditions

$$\rho_t(f) = \rho_0(f) + \int_0^t \rho_s(A_s f)\,\mathrm{d}s + \int_0^t \rho_s(hf)\,\mathrm{d}Y_s, \quad P_1\text{-a.s.}, \tag{2.4}$$

cf. [26, 37, 43, 44]. Here $\rho_0 = P_{X_0}$, and A_s is the generator of X, given on $C_c^\infty(\mathbb{R}^d)$ by

$$A_s = \frac{1}{2}\sum(\sigma\sigma^{\mathrm{T}})_{ij}(s, x)\frac{\partial^2}{\partial x_i \partial x_j} + \sum b_i(s, x)\frac{\partial}{\partial x_i}.$$

Remark 2.1. For (2.4) to hold one may (beside conditions on h and X_0) only assume that b and σ are Lipschitz continuous and of linear growth. It is not necessary to assume that the matrix $\sigma\sigma^{\mathrm{T}}$ in A_s is nondegenerate. Condition 2.3 given below is such that (2.4) holds for (2.1), (2.2).

In general, the conditional probability $P[X_t \in B|\mathcal{Y}_t]$ does not have a Lebesgue density, because σ in (2.1) may be degenerate. If one imposes conditions such as uniform ellipticity, hypoellipticity, or conditions related to the Malliavin calculus [5, 26, 34, 37], one proves that the conditional probability does possess a density $p_t(x, \omega_1)$. In view of (2.3) $\rho_t(\omega_1)$ then has the density

$$u_t(x, \omega_1) := \rho_t(1)(\omega_1) p_t(x, \omega_1). \tag{2.5}$$

This relation gives $p_t(x, \omega_1) = u_t(x, \omega_1)/\int_{\mathbb{R}^d} u_t(y, \omega_1)\,\mathrm{d}y$, and explains the name *unnormalized conditional density* for u_t. Under the assumption of sufficient smoothness of u_t and of the data, partial integrations in (2.4) imply that u_t satisfies

$$u_t(x, \omega_1) = p_0(x) + \int_0^t A_s^* u_s(x, \omega_1)\,\mathrm{d}s + \int_0^t h(x) u_s(x, \omega_1)\,\mathrm{d}Y_s(\omega_1), \tag{2.6}$$

with A_s^* denoting the formal $L^2(\mathbb{R}^d, \mathrm{d}x)$-adjoint of A_s. To distinguish the more general equation (2.4) from (2.6) we call (2.4) the *Zakai measure equation* and (2.6) the *Zakai density equation*.

Remark 2.2. The transition from (2.4) to (2.6) using partial integration is only justified if one knows *in advance* that ρ_t is endowed with a density u_t with specific analytic properties, cf. [21, p. 275]. Nondegeneracy assumptions on the generator A_s are often crucial to obtain such properties. In Section 4 we will replace this kind of assumption by smoothness assumptions on coefficients and data, and still obtain that (2.6) gives rise to the correct density of degenerate filtering problems.

We next present a stochastic Feynman–Kac (FK) formula for the solution of (2.6). We first give a formal argument suggesting this formula and give conditions securing that it defines a strong solution of (2.6). Details can be found in [2]. First, note that the adjoint operator A_t^* in (2.6) is of the general type $A_t^* = L_t + c_t$ with differential operator

$$L_t = \frac{1}{2}\sum_{i,j=1}^{d} (\sigma\sigma^{\mathrm{T}})_{ij}(t, x)\frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{d} \beta_i(t, x)\frac{\partial}{\partial x_i},$$

and multiplication operator c_t, connected to the scalar potential $c(t, x)$. In particular, we find

$$\begin{aligned} \beta_i(t, x) &= \sum_{j=1}^{d} \frac{\partial}{\partial x_j}(\sigma\sigma^{\mathrm{T}})_{ij}(t, x) - b_i(t, x), \\ c(t, x) &= \frac{1}{2}\sum_{i,j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j}(\sigma\sigma^{\mathrm{T}})_{ij}(t, x) - \sum_{i=1}^{d} \frac{\partial}{\partial x_i} b_i(t, x). \end{aligned} \tag{2.7}$$

Using this notation, and expressing (2.6) in differential form, leads to the Itô equation

$$\mathrm{d}u_t(x,\omega_1) = L_t u_t(x,\omega_1)\,\mathrm{d}t + c(t,x)u_t(x,\omega_1)\,\mathrm{d}t + h(x)u_t(x,\omega_1)\,\mathrm{d}Y_t(\omega_1).$$

Formally, we may rewrite this equation using Stratonovic integrals, and for *fixed* ω_1 apply the FK-formula [17] to the resulting ("nonstochastic") equation. This means that we have to consider a new probability space $(\Omega_2, \mathcal{F}_2, P_2)$, and express the solution $u_t(x,\omega_1)$ as a particular expectation value w.r.t. P_2. This procedure leads to the following candidate for the solution:

$$u_t(x,\omega_1) = E_{P_2}\biggl[p_0(\xi_t^{t,x})\exp\biggl\{\int_0^t C(s,\xi_{t-s}^{t,x})\,\mathrm{d}s + \int_0^t h(\xi_{t-s}^{t,x})\,\mathrm{d}Y_s(\omega_1)\biggr\}\biggr]. \tag{2.8}$$

Here we abbreviated

$$C(s,\xi_{t-s}^{t,x}) := c(s,\xi_{t-s}^{t,x}) - \frac{1}{2}|h(\xi_{t-s}^{t,x})|^2, \tag{2.9}$$

and the $\mathbb{R}^d$-valued process $\xi^{t,x} = (\xi_s^{t,x})_{0\leqslant s\leqslant t}$ on $(\Omega_2, \mathcal{F}_2, P_2)$ solves the *associated reversed Itô equation*

$$\mathrm{d}\xi_s^{t,x} = \beta(t-s,\xi_s^{t,x})\,\mathrm{d}s + \sigma(t-s,\xi_s^{t,x})\,\mathrm{d}B_s, \quad \xi_0^{t,x} = x,\ s\in[0,t], \tag{2.10}$$

with an $\mathbb{R}^d$-valued Wiener process $B := (B_s)_{s\geqslant 0}$ on $(\Omega_2, \mathcal{F}_2, P_2)$. We now impose conditions which guarantee that (2.8) in fact *defines* a strong solution to (2.6), if one interprets all functions in (2.8) as random variables in $(\Omega_1\times\Omega_2, \mathcal{F}_1\otimes\mathcal{F}_2, \overline{P}_1\otimes P_2)$. The space of functions $f\in C_b^1([0,T]\times\mathbb{R}^d)$ having continuous and bounded space derivatives up to the kth order is denoted $C_b^{1,k}([0,T]\times\mathbb{R}^d)$.

CONDITION 2.3. Assume that $c, \beta_i, \sigma_{ij} \in C_b^{1,3}([0,T]\times\mathbb{R}^d)$ for all $i,j\in\{1,\dots,d\}$, and that $p_0, h_i\in C_b^3(\mathbb{R}^d)$ for all $i\in\{1,\dots,m\}$.

It follows from [2, Proposition 2.6 and Theorem 2.7] that, under Condition 2.3, the *stochastic FK-formula* (2.8) is a strong solution to the Zakai density equation (2.6). We remark that condition (D) in Theorem 2.7 is satisfied because the $\mathcal{S}$-transformed equation (2.6) is solved by the classical (i.e., nonstochastic) FK-formula. This follows with the technique of stochastic flows, as given, e.g., in [27], but can also be checked using the calculus developed in Sections 3 and 4 in [2] (cf. [2, Lemma 5.1]). We also mention that the differentiability assumptions in Condition 2.3 can be relaxed by one degree if one imposes a uniform Hölder condition on the highest remaining derivatives [2, Theorem 2.7].

Remarks. (1) Formula (2.8) defines a strong solution also in the case of positive semi-definite differential operators L_t. This is crucial in order to apply (2.8) to parameter estimation.

(2) Stochastic representation formulas such as (2.8) have played an important role since the beginning of nonlinear filtering theory [9, 23, 26, 33, 37]. The fact

that these formulas represent conditional measures or densities allows one to derive properties, for example, about the moments of $p_t(\omega, \mathrm{d}x)$ or about the positivity of $u_t(\omega, \mathrm{d}x)$ (in an almost sure sense). Note that (2.8) defines a strong solution to (2.6) also for the case when Equation (2.6) is *not* induced by a filtering problem. Therefore, some of the results in the following sections, such as Theorem 3.5 and Lemma 4.1, also apply for SPDEs of the Zakai type when arguments based on conditional densities are not at hand.

3. A Robust Feynman–Kac Representation for SPDEs

In this section we give a version of (2.8), which depends continuously on the observation path Y. We first define the random variable $\omega_1 \mapsto u_t(x, \omega_1)$ for all $\omega_1 \in \Omega_1$. By (2.8) $u_t(x)$ is well-defined modulo sets of P_1-measure zero, but it is not clear from (2.8) how to define $u_t(x, \omega_1)$ for all $\omega_1 \in \Omega_1$, because the term $\int_0^t h(\xi_{t-s}^{t,x})\, \mathrm{d}Y_s$ is, a priori, only defined modulo sets of $\overline{P}_1 \otimes P_2$-measure zero. A partial integration in this term will solve this problem and simultaneously give us a robust version. But notice that, since the reversed process $\xi_{t-s}^{t,x}$ is not adapted to the (Y, B)-filtration, we cannot simply apply Itô calculus for this partial integration. This technicality is solved as follows:

LEMMA 3.1. *If Condition* 2.3 *holds, if* $0 \leqslant t_0 < t$, $x \in \mathbb{R}^d$, *and if* $\xi^{t,x}$ *is defined by* (2.10)*, then the following identity holds* $\overline{P_1} \otimes P_2$*-almost surely:*

$$\int_{t_0}^{t} h(\xi_{t-s}^{t,x})\, \mathrm{d}Y_s = h(x)Y_t - h(\xi_{t-t_0}^{t,x})Y_{t_0} + \int_0^{t-t_0} Y_{t-s}\, \mathrm{d}h(\xi_s^{t,x}). \tag{3.1}$$

Proof. Define the Wiener process $\tilde{Y}_s := Y_t - Y_{t-s}$ for $s \in [0, t]$. Let us first verify that

$$\int_{t_0}^{t} h(\xi_{t-s})\, \mathrm{d}Y_s = \int_0^{t-t_0} h(\xi_u)\, \mathrm{d}\tilde{Y}_u, \quad \overline{P}_1 \otimes P_2\text{-a.s.}, \tag{3.2}$$

where ξ_u abbreviates $\xi_u^{t,x}$. Choose partitions $t_0 = s_0^n < s_1^n < \cdots < s_{N_n}^n = t$ whose mesh goes to zero. Since the paths $s \mapsto h(\xi_{t-s})$ are continuous the Riemann sums

$$I_n := \sum_{k=0}^{N_n-1} h(\xi_{t-s_k^n})(Y_{s_{k+1}^n} - Y_{s_k^n})$$

converge to the left side in (3.2) in probability. For convenience we suppress n from the notation. Setting $u_k := t - s_k$ and writing $h(\xi_{u_k}) = h(\xi_{u_{k+1}}) + [h(\xi_{u_k}) - h(\xi_{u_{k+1}})]$, it follows that

$$I_n = \sum_{k=0}^{N_n-1} h(\xi_{u_{k+1}})(\tilde{Y}_{u_k} - \tilde{Y}_{u_{k+1}}) + \sum_{k=0}^{N_n-1} [h(\xi_{u_k}) - h(\xi_{u_{k+1}})](\tilde{Y}_{u_k} - \tilde{Y}_{u_{k+1}}).$$

By independence the second term vanishes as $n \to \infty$, so I_n converges in probability to the right side in (3.2), i.e. (3.2) holds. Next, for $h \in C_b(\mathbb{R}^d)^m \cap C^2(\mathbb{R}^d)^m$ the right side in (3.2) can be computed by Itô calculus with respect to the Wiener process $(\tilde{Y}, B) := (\tilde{Y}_s, B_s)_{0 \leqslant s \leqslant t-t_0}$:

$$\int_0^{t-t_0} h(\xi_s)\,\mathrm{d}\tilde{Y}_s = h(\xi_{t-t_0})\tilde{Y}_{t-t_0} - h(\xi_0)\tilde{Y}_0 - \int_0^{t-t_0} \tilde{Y}_s\,\mathrm{d}h(\xi_s).$$

Using $\xi_0 = x$, $\tilde{Y}_{t-t_0} = Y_t - Y_{t_0}$, $\tilde{Y}_0 = 0$ and $\tilde{Y}_s = Y_t - Y_{t-s}$ we obtain (3.1). □

As a consequence of Lemma 3.1 we may replace the stochastic integral in (2.8) by the right side in (3.1) (with $t_0 = 0$) without changing $u_t(x)$ except on a set of P_1-measure zero. This gives

$$u_t(x, \omega_1) = \mathrm{e}^{h(x)Y_t(\omega_1)} U_t(x, \omega_1), \quad P_1\text{-a.s.}, \tag{3.3}$$

where the random variable $U_t(x)$ is defined (P_1-a.s.) by

$$U_t(x) := E_{P_2}\left[p_0(\xi_t^{t,x}) \exp\left\{\int_0^t (C(s, \xi_{t-s}^{t,x})\,\mathrm{d}s + \int_0^t Y_{t-s}\,\mathrm{d}h(\xi_s^{t,x})\right\}\right]. \tag{3.4}$$

Remark 3.2. The decomposition of u into the two factors given in (3.3) is well–known in nonlinear filtering theory, see for example [12]. Through this splitting one usually *defines* a function U and one shows that this function satisfies the so-called *robust Zakai-equation*, provided u satisfies the Zakai density equation (2.6). This fits together with our approach.

Observe that $\int_0^t Y_{t-s}\,\mathrm{d}h(\xi_s^{t,x})$ in (3.4) is defined modulo sets of $\overline{P}_1 \otimes P_2$-measure zero. Thus $U_t(x)$ is – similar to $u_t(x)$ – only defined modulo sets of P_1-measure zero. But $U_t(x)$ has an ω_1-wise defined version, obtained as follows. We start with choosing a version of the Wiener process Y with continuous sample paths for all $\omega_1 \in \Omega_1$. We fix such a version for the rest of this paper. Because $t \mapsto Y_t(\omega_1)$ is continuous we can define $\int_0^t Y_{t-s}(\omega_1)\,\mathrm{d}h(\xi_s)$ as a P_2-a.s. defined limit *for each fixed* ω_1, separately. We distinguish the two possible definitions by writing $\int_0^t Y_{t-s}(\omega_1)\,\mathrm{d}h(\xi_s)(\omega_2)$, for ω_1 fixed, and $\int_0^t Y_{t-s}\,\mathrm{d}h(\xi_s)(\omega_1, \omega_2)$, for the $\overline{P}_1 \otimes P_2$-a.s. definition. If the random variable

$$\begin{aligned} F(t, x, Y)(\omega_2) \ := \ & p_0(\xi_t^{t,x}(\omega_2)) \exp\left\{\int_0^t C(s, \xi_{t-s}^{t,x}(\omega_2))\,\mathrm{d}s + \right. \\ & \left. + \int_0^t Y_{t-s}\,\mathrm{d}h(\xi_s^{t,x})(\omega_2)\right\}, \end{aligned} \tag{3.5}$$

is in $L^1(P_2)$ for every function $Y \in C([0, T])^m$, we expect that the random variable

$$\tilde{U}_t(x, \omega_1) := E_{P_2}[F(t, x, Y_\cdot(\omega_1))] \tag{3.6}$$

coincides P_1-a.s. with $U_t(x, \omega_1)$, given by (3.4). If this holds then an ω_1-wise everywhere defined version of $u_t(x, \omega_1)$ follows from (3.3). The next lemma establishes the required equality.

LEMMA 3.3. *Assume Condition* 2.3 *holds and let* $(t, x) \in [0, T] \times \mathbb{R}^d$ *be fixed. Then the random variables defined in* (3.4) *and* (3.6) *satisfy* $U_t(x) = \tilde{U}_t(x)$, P_1*-a.s. In particular, the FK-representation* (2.8) *for the solution to* (2.6) *can be expressed as*

$$u_t(x, \omega_1) = \mathrm{e}^{h(x)Y_t(\omega_1)} E_{P_2}[F(t, x, Y_\cdot(\omega_1))], \quad P_1\text{-a.s.} \tag{3.7}$$

Proof. We show that there exists $A \subset \Omega_1$ with $P_1(A) = 1$, such that for all $\omega_1 \in A$

$$\int_0^t Y_{t-s}(\omega_1)\, \mathrm{d}h(\xi_s)(\omega_2) = \int_0^t Y_{t-s}\, \mathrm{d}h(\xi_s)(\omega_1, \omega_2), \quad P_2\text{-a.s.}, \tag{3.8}$$

for suitable, pointwise everywhere defined versions of the integrals. Fubini's theorem then implies $E_{P_1}[|U_t(x) - \tilde{U}_t(x)|] = 0$, so the claim will follow. To prove (3.8) choose a continuous version of the process $(\xi_s^{t,x})_{0\leqslant s\leqslant t}$, and let $0 = s_0^n < \cdots < s_{N_n}^n = t$ be partitions whose mesh goes to zero.

Then the Riemann-approximations to the right side in (3.8),

$$I_n(\omega_1, \omega_2) := \sum_{k=0}^{N_n-1} Y_{t-s_k^n}(\omega_1)\big[h(\xi_{s_{k+1}^n}(\omega_2)) - h(\xi_{s_k^n}(\omega_2))\big] \tag{3.9}$$

are well defined for all (ω_1, ω_2), and jointly measurable. A suitable subsequence of $(I_n)_{n\in\mathbb{N}}$ converges $\overline{P}_1 \otimes P_2$-a.s. to the right side in (3.8). Denote by $N \subset \Omega_1 \times \Omega_2$ the exceptional set, and put $I(\omega_1, \omega_2) = 0$ on N, in order to obtain a definition of the right side in (3.8) for all (ω_1, ω_2).

Next, consider the left side in (3.8). By Fubini's theorem $\omega_1 \mapsto g(\omega_1) := \int 1_N(\omega_1, \omega_2)\, \mathrm{d}P_2(\omega_2)$ is $\mathcal{F}_1$-measurable and satisfies $g = 0$ $\overline{P}_1$-a.s. Thus $A := g^{-1}(\{0\}) \in \mathcal{F}_1$ and $\overline{P}_1(A) = 1$. We now fix $\omega_1 \in A$, and notice, again by Fubini, that $N_{\omega_1} := \{\omega_2 \in \Omega_2 : (\omega_1, \omega_2) \in N\} \in \mathcal{F}_2$, and $P_2(N_{\omega_1}) = 0$. By the definition of N the sub-sequence from (3.9) converges to the right side in (3.8) for all $\omega_2 \in N_{\omega_1}^c$. Thus the limit defines (P_2-a.s.) a version of the left side in (3.8), and this establishes (3.8). The lemma now follows from (3.4) and Fubini's theorem. □

Remark 3.4. Equality (3.7) appears obvious at first sight. It would follow immediately from

$$\int_0^t Y_{t-s}(\omega_1)\, \mathrm{d}h(\xi_s)(\omega_2) = \int_0^t Y_{t-s}\, \mathrm{d}h(\xi_s)(\omega_1, \omega_2), \quad \overline{P}_1 \otimes P_2\text{-a.s.}, \tag{3.10}$$

but there is a technical problem to prove this equation. The point is that for fixed ω_1 the map $\omega_2 \mapsto \int_0^t Y_{t-s}(\omega_1)\, \mathrm{d}h(\xi_s)(\omega_2)$ on the left side in (3.10) is only well-defined modulo P_2-zero sets, i.e. one has to *choose* for every ω_1 a representative to get a function on $\Omega_1 \times \Omega_2$. These choices may result in a map $(\omega_1, \omega_2) \mapsto \int_0^t Y_{t-s}(\omega_1)\, \mathrm{d}h(\xi_s)(\omega_2)$ which is not jointly measurable. In the proof of Lemma 3.3

we avoided this problem by establishing (3.8) instead of (3.10). After this construction we can now define, for every $\omega_1 \in A$, a representative of $\int_0^t Y_{t-s}(\omega_1)\,\mathrm{d}h(\xi_s)$ by the right side of (3.8). For $\omega_1 \notin A$ one simply picks any representative. With *this* choice the map $(\omega_1, \omega_2) \mapsto \int_0^t Y_{t-s}(\omega_1)\,\mathrm{d}h(\xi_s)(\omega_2)$ is jointly measurable, and (3.10) holds.

Observe that the right side in (3.7) gives an ω_1-wise defined representation of the random field $(t, x) \mapsto u_t(x)$. This representation is robust in the sense that $u_t(x)$ depends on ω_1 only through the observation path $Y_\cdot(\omega_1)$, and considered as a function of Y it is in fact continuous. This is the main result of the present section:

THEOREM 3.5. *Assume Condition* 2.3 *holds and fix* $(t, x) \in [0, T] \times \mathbb{R}^d$. *Then* $F(t, x, Y)$ *defined by* (3.5) *is in* $L^1(P_2)$, *for all* $Y \in B[0, T]^m$. *Moreover, the map*

$$Y \mapsto \mathrm{e}^{h(x)Y_t} E_{P_2}[F(t, x, Y)]$$

is a continuous functional on $(B[0, T]^m, \|\cdot\|_\infty)$.

Proof. We abbreviate $\xi_s^{t,x}$ by ξ_s, and let $Z = (Z^1, \dots, Z^m) \in B[0, T]^m$. Then, using Itô's rule and the summation convention over repeated indices, we find

$$\begin{aligned}\int_0^t Z_s\,\mathrm{d}h(\xi_s) &= \int_0^t Z_s^k\,\mathrm{d}h_k(\xi_s)\\ &= \int_0^t Z_s^k\{\gamma_k(s, \xi_s)\,\mathrm{d}s + g_{kj}(s, \xi_s)\,\mathrm{d}B_s^j\},\end{aligned} \tag{3.11}$$

where we have used the following abbreviations:

$$\begin{aligned}\gamma_k(s, \xi_s) &:= \frac{\partial h_k}{\partial x_i}(\xi_s)\beta_i(s, \xi_s) + \frac{1}{2}\frac{\partial^2 h_k}{\partial x_i \partial x_j}(\xi_s)(\sigma\sigma^{\mathrm{T}})_{ij}(s, \xi_s),\\ g_{kj}(s, \xi_s) &:= \frac{\partial h_k}{\partial x_i}(\xi_s)\sigma_{ij}(s, \xi_s).\end{aligned} \tag{3.12}$$

Notice that γ_k and g_{kj} are bounded. For bounded functions $g(s, x) = (g_1(s, x), \dots, g_d(s, x))$ we put $g_s := g(s, \xi_s)$ and estimate, using the exponential martingale

$$\begin{aligned}&E_{P_2}\left[\exp\left(\int_0^t g_s\,\mathrm{d}B_s\right)\right]\\ &\quad\leqslant E_{P_2}\left[\exp\left(\int_0^t g_s\,\mathrm{d}B_s - \frac{1}{2}\int_0^t g_s^2\,\mathrm{d}s\right)\right]\mathrm{e}^{t\|g^2\|_\infty/2} = \mathrm{e}^{t\|g\|_\infty^2/2}.\end{aligned} \tag{3.13}$$

With this and (3.11), applied to $g_j(s, x) = g_{kj}(s, x)$ for fixed k, we find

$$E_{P_2}\left[\exp\left\{\int_0^t Z_s\,\mathrm{d}h(\xi_s)\right\}\right] \leqslant \exp\{c_1\|Z\|_\infty\}\exp\{c_2\|Z\|_\infty^2\}, \tag{3.14}$$

where c_1, c_2 denote suitable constants. This implies $F(t,x,Y) \in L^1(P_2)$, for all $Y \in B[0,T]^m$, because the terms in (3.5), beside the stochastic integral, are uniformly bounded. To prove continuity we use $|e^a - e^b| \leqslant (e^a + e^b)|a-b|$ which holds for all $a, b \in \mathbb{R}$. For $Y, \bar{Y} \in B[0,t]^m$

$$\begin{aligned}
&\left|E_{P_2}[F(t,x,Y)] - E_{P_2}[F(t,x,\bar{Y})]\right|^2 \\
&\quad \leqslant c_3 E_{P_2}\Big[\big|e^{\int_0^t Y_{t-s}\,\mathrm{d}h(\xi_s)} - e^{\int_0^t \bar{Y}_{t-s}\,\mathrm{d}h(\xi_s)}\big|\Big]^2 \\
&\quad \leqslant 2c_3 E_{P_2}\Big[e^{2\int_0^t Y_{t-s}\,\mathrm{d}h(\xi_s)} + e^{2\int_0^t \bar{Y}_{t-s}\,\mathrm{d}h(\xi_s)}\Big] E_{P_2}\left[\left|\int_0^t \Delta Y_s\,\mathrm{d}h(\xi_s)\right|^2\right],
\end{aligned} \tag{3.15}$$

where $\Delta Y_s := Y_{t-s} - \bar{Y}_{t-s}$. The first factor is bounded as in (3.14). To estimate the second factor in (3.15) we use (3.11) with $Z_s := \Delta Y_s$, the Itô isometry, and the boundedness of γ_k, g_{kj}:

$$\begin{aligned}
E_{P_2}\left[\left|\int_0^t \Delta Y_s\,\mathrm{d}h(\xi_s)\right|^2\right] &\leqslant 2E_{P_2}\left[\left|\int_0^t \Delta Y_s^k \gamma_k(\xi_s)\,\mathrm{d}s\right|^2\right] + \\
&\quad + 2E_{P_2}\left[\left|\int_0^t \Delta Y_s^k g_{kj}(s,\xi_s)\,\mathrm{d}B_s^j\right|^2\right] \\
&\leqslant c_4\|\Delta Y\|_\infty^2 + c_5\|\Delta Y\|_\infty^2.
\end{aligned}$$

This and (3.14) allows to derive the following estimate from (3.15):

$$\begin{aligned}
&E_{P_2}[|F(t,x,Y) - F(t,x,\bar{Y})|] \\
&\quad \leqslant c_6 e^{2c_1(\|Y\|_\infty + \|\bar{Y}\|_\infty) + 2c_2(\|Y\|_\infty^2 + \|\bar{Y}\|_\infty^2)} \|Y - \bar{Y}\|_\infty.
\end{aligned} \tag{3.16}$$

This clearly implies the asserted (local Lipschitz) continuity. □

4. The Conditional Measure for Degenerate Filtering Problems

In this section we study degenerate filtering problems. We prove that the FK-formula (3.7) defines a density $u_t(x)$ of the unnormalized conditional measure given by (2.3). Our proof does not presuppose assumptions on the analytic properties of $u_t(x)$ besides those implied by (3.7) and Condition 2.3 (cf. Remark 2.2). We finally refine the general results for parameter estimation.

Our approach is based on a result by Kurtz and Ocone [28]. The idea goes as follows: The FK-solution (3.7) to Equation (2.6) defines a measure ρ_t via

$$\rho_t(\omega_1, \mathrm{d}x) := u_t(x, \omega_1)\,\mathrm{d}x. \tag{4.1}$$

In view of [28, Theorem 4.2] it suffices to verify that this measure is finite, and it satisfies (2.4) for all $f \in C_c^\infty(\mathbb{R}^d)$ and for $f = 1$, to conclude that ρ coincides with the measure (2.3). (Related uniqueness results are given in [3, 37].) The finiteness of the measure ρ_t follows immediately from the next result which is also of independent interest.

LEMMA 4.1. *Suppose Condition* 2.3 *holds, but only assume that* p_0 *is measurable and has compact support in* $\mathbb{R}^d$. *Let* u *be the robust functional* (3.7). *Then there is a constant* $c > 0$ *such that for each* $\omega_1 \in \Omega_1$ *and a suitable constant* $M(\omega_1)$ *the following estimate holds:*

$$|u_t(x, \omega_1)| \leqslant M(\omega_1)\mathrm{e}^{-x^2/2c}, \quad \forall (t, x) \in [0, T] \times \mathbb{R}^d. \tag{4.2}$$

Proof. Fix ω_1 and put $Y_t = Y_t(\omega_1)$. The same estimation as in the proof of Theorem 3.5 yields

$$|u_t(x, \omega_1)|^2 \leqslant \mathrm{e}^{2h(x)Y_t} E_{P_2}[p_0(\xi_t^{t,x})^2] E_{P_2}[\mathrm{e}^{2(\cdots)}] \leqslant M_1(\omega_1) E_{P_2}[f(\xi_t^{t,x})]. \tag{4.3}$$

Here $f := p_0^2$, and the term $\mathrm{e}^{2(\cdots)}$ denotes the squared exponential term in (3.5), whose expectation is bounded as the proof of the previous theorem shows. By Condition 2.3 there exist positive constants $a, b < \infty$ such that β and σ from (2.10) satisfy $|\beta(t, x)| := (\sum_i |\beta_i(t, x)|^2)^{1/2} < b$ and $\sum_{ij} |\sigma_{ij}(t, x)|^2 \leqslant a^2$ for all $(t, x) \in [0, T] \times \mathbb{R}^d$. From [41, Lemma 4.7] we get the uniform estimate

$$P\big(|\xi_t^{t,x} - x| \geqslant R\big) \leqslant 2d\mathrm{e}^{-(R-\sqrt{d}bT)^2/(2a^2dT)}, \quad \forall (t, x) \in [0, T] \times \mathbb{R}^d, \tag{4.4}$$

which holds for all $R > \sqrt{d}bT$. Now let $r > 0$ be such that $p_0(x) = 0$ for all $|x| \geqslant r$, then for $|x| - r > \sqrt{d}bT$ we get the estimate

$$\begin{aligned} E_{P_2}\big[f(\xi_t^{t,x})\big] &= \int_{\mathbb{R}^d} f(y)\,\mathrm{d}P_{\xi_t^{t,x}}(y) \leqslant \|f\|_\infty P(|\xi_t^{t,x}| < r) \\ &\leqslant \|f\|_\infty P(|\xi_t^{t,x} - x| \geqslant |x| - r) \\ &\leqslant \|f\|_\infty 2d\mathrm{e}^{-(|x|-r-\sqrt{d}bT)^2/(2a^2T)}. \end{aligned}$$

Together with (4.3) this implies (4.2) for a suitable choice of $M(\omega_1)$ and c. □

THEOREM 4.2. *Suppose Condition* 2.3 *holds with* $p_0 \in C_c^3(\mathbb{R}^d)$. *Let* $\xi_s^{t,x}$ *be the solution to* (2.10)*, and let the solution* u *to* (2.6) *be given in the robust form* (3.7). *Then the random measure* ρ_t *defined by* (4.1) *solves the filtering problem* (2.3).

Proof. From [2, Theorem 2.7] we know that u given in (2.8) is a strong solution to (2.6). In particular, this implies that $u(\cdot, \cdot, \omega_1)$ is (P_1-a.s.) continuous on $[0, T] \times \mathbb{R}^d$, and C^2 in the space variables. Now multiply (2.6) by $f(x)$, with $f \in C_c^\infty(\mathbb{R}^d)$, and integrate over $\mathbb{R}^d$. This gives (with obvious abbreviations)

$$\int_{\mathbb{R}^d} f(x) u_t(x, \omega_1)\,\mathrm{d}x = \int_{\mathbb{R}^d} f(x) p_0(x)\,\mathrm{d}x + I_1 + I_2. \tag{4.5}$$

Since $A_\cdot^* u(\cdot, \cdot, \omega_1) \in C([0, T] \times \mathbb{R}^d)$ and $f \in C_c^\infty(\mathbb{R}^d)$ we can first apply Fubini's Theorem to I_1 and then do partial integrations. This gives

$$I_1 = \int_0^t \left[\int_{\mathbb{R}^d} (A_s f(x)) u_s(x, \omega_1)\,\mathrm{d}x\right] \mathrm{d}s. \tag{4.6}$$

The second term I_2 requires a stochastic version of Fubini's Theorem, that is, we want to have

$$\begin{aligned} I_2 &= \int_{\mathbb{R}^d} f(x)\left[\int_0^t h(x)u_s(x,\omega_1)\,\mathrm{d}Y_s(\omega_1)\right]\mathrm{d}x \\ &= \int_0^t \left[\int_{\mathbb{R}^d} f(x)h(x)u_s(x,\omega_1)\,\mathrm{d}x\right]\mathrm{d}Y_s(\omega_1). \end{aligned} \tag{4.7}$$

In view of [24, Lemma 3.2] this is in fact justified because of the uniform L^p-estimate

$$\begin{aligned} &E_{\overline{P}_1}\left[|u_t(x)|^p\right] \\ &\quad\leqslant \|p_0\|_\infty^p \exp\left\{pt\left(\|c\|_\infty + \frac{1}{2}(p-1)\|h\|_\infty^2\right)\right\}, \quad \forall p \geqslant 1, \end{aligned} \tag{4.8}$$

which follows easily from (3.13). These arguments show that (4.5) can be written as (2.4). Finally, let $f_n \in C_c^\infty(\mathbb{R}^d)$ be such that $f_n(x) = 1$ for all $|x| \leqslant n$, and f_n, together with all derivatives $\partial_i f_n$, $\partial_i\partial_j f_n$ are uniformly bounded for all $n \in \mathbb{N}$, and all $|x| > n$. Writing the Zakai measure equation (2.4) for f_n and letting n go to infinity, yields

$$\rho_t(1) = 1 + \int_0^t \rho_s(h)\,\mathrm{d}Y_s, \tag{4.9}$$

where we have used the estimate (4.2). With (2.4) and (4.9) we have verified the conditions for [28, Theorem 4.2], thus ρ_t is the unnormalized conditional measure of the filtering problem. □

Remark 4.3. The uniqueness theorem [28, Theorem 4.2] only states the assertion for diffusion processes, i.e., for time independent b, σ. On page 90 the authors write: "However, there is no loss of generality because the time dependent case can always be recast into the autonomous form by the standard trick of including time as a component of ξ. Hence, all results that follow may be interpreted for the time dependent case." It is this case that we adopt for our Theorem 4.2.

With these preparations we now investigate the estimation problem for the parameterized equation (1.1). Let us first discuss (path-wise) smooth parametric dependence of the solution X^θ to (1.1). This can be derived from smooth dependence on initial data (see, e.g., [26, 27]) as the proof of the following lemma shows. Subsequently $\Theta \subset \mathbb{R}^d$ is an open set, and open balls in $\mathbb{R}^k$ with radius r and center x are denoted $B_r(x)$. Depending on the context we will view $\theta \in \Theta$ as a fixed parameter or as a variable, and for convenience, we adopt the notation $\sigma_\theta(t,x)$, $\sigma(t,x,\theta)$ and so forth, when emphasizing the parameter and variable dependency, respectively.

LEMMA 4.4. *Let $m \in \mathbb{N}$, and assume that the coefficients b and σ in* (1.1) *satisfy $b_i, \sigma_{ij} \in C_b^{1,m+1}([0,T] \times \mathbb{R}^d \times B)$ for all i, j, and all bounded open sets $B \subset \Theta$. Then there is a jointly measurable mapping X: $[0,T] \times \mathbb{R}^d \times \Theta \times \Omega_1 \to \mathbb{R}^d$ such that*:

(α) *For fixed (x,θ) the process $X_t^{x,\theta}(\omega_1) := X(t,x,\theta,\omega_1)$ solves* (1.1) *with $X_0^{x,\theta}(\omega_1) = x$.*
(β) *For fixed ω_1 the function $(t,x,\theta) \mapsto X_t^{x,\theta}(\omega_1)$ is continuous.*
(γ) *For fixed (t,ω_1) the function $(x,\theta) \mapsto X_t^{x,\theta}(\omega_1)$ is in $C^m(\mathbb{R}^d \times \Theta)$.*

Proof. Choose $(\theta_n, r_n)_{n\in\mathbb{N}}$ such that $B_{r_n}(\theta_n) \subset \Theta$, and $\Theta_n := B_{r_n/3}(\theta_n)$ satisfies $\Theta = \bigcup_{n\in\mathbb{N}} \Theta_n$. Let $\psi_n \in C^\infty(\mathbb{R}^k)$ be such that $\psi_n(\theta) = 1$ for $\theta \in \Theta_n$ and $\psi_n(\theta) = 0$ for $\theta \notin B_{r_n/2}(\theta_n)$. Define

$$b_n(t,x,\theta) := \begin{cases} \psi_n(\theta) b(t,x,\theta) & \text{for } \theta \in B_{r_n}(\theta_n), \\ 0 & \text{for } \theta \in \mathbb{R}^k \backslash B_{r_n}(\theta_n), \end{cases}$$

and correspondingly define σ_n. Clearly b_n, σ_n are in $C_b^{1,m+1}([0,T] \times \mathbb{R}^d \times \mathbb{R}^k)$, which implies that the (extended) SDE

$$dX_t = b_n(t, X_t, \theta_t)\, dt + \sigma_n(t, X_t, \theta_t)\, dW_t \tag{4.10}$$

$$d\theta_t = 0, \qquad (X_0, \theta_0) = (x,\theta) \in \mathbb{R}^d \times \mathbb{R}^k, \tag{4.11}$$

has a unique solution, denoted $X_{n,t}^{x,\theta}$. Moreover, from [27, Theorem II.3.3] it follows that the map X_n: $(t,x,\theta,\omega_1) \mapsto X_{n,t}^{x,\theta}(\omega_1)$ has a modification which satisfies (α), (β), (γ), with b, σ replaced by b_n, σ_n. For θ in Θ_n the solution $\theta_t = \theta$ inserted in (4.10) shows that $X_{n,t}^{x,\theta}$ solves (1.1). Thus, X_n satisfies (α), (β), (γ) for each $\theta \in \Theta_n$. By uniqueness and continuity, two functions $X_n(\cdot,\omega_1)$ and $X_m(\cdot,\omega_1)$ coincide (P_1-a.s.) on $[0,T] \times \mathbb{R}^d \times (\Theta_n \cap \Theta_m)$. Thus the map $\hat{X}(t,x,\theta,\omega_1) := X_n(t,x,\theta,\omega_1)$ (with n chosen such that $\theta \in \Theta_n$) is well-defined P_1-a.s., and it satisfies (α), β), (γ) P_1-a.s. Thus, with a suitable P_1-zero set N the map $X := \hat{X} \cdot 1_{N^c}$ satisfies (α), (β), (γ). □

Remark 4.5. The localization procedure in the previous proof is not needed if one assumes conditions uniformly with respect to $\theta \in \Theta$. But this assumption is usually not satisfied since θ often appears in an analytic expression in $b_\theta(t,x)$, $\sigma_\theta(t,x)$ and has unbounded domain Θ.

We are now prepared for the parameter estimation of (1.1) based on the observations (1.2). We follow the idea described in Section 1. It is no loss of generality to assume $\alpha = 1$ because the case $\alpha > 0$ follows by a simple rescaling, see (5.1) below. Consider once more the extended system (1.1), (1.3) subject to initial conditions $(X(0), \theta(0)) = (X_0, \theta_0)$, with X_0 and θ_0 independent of the $(W^{(1)}, W^{(2)})$-filtration. We impose the following conditions on (1.1), (1.3), and on the measurements (2.2):

CONDITION 4.6. For every bounded open $B \subset \Theta$ it holds that $b_i \in C_b^{1,4}([0,T] \times \mathbb{R}^d \times B)$, $\sigma_{ij} \in C_b^{1,5}([0,T] \times \mathbb{R}^d \times B)$, and $h_i \in C_b^3(\mathbb{R}^d)$ for all i, j. X_0 has density $p_0 \in C_c^3(\mathbb{R}^d)$, and θ_0 has a distribution $\mu := P_{\theta_0}$ with compact support in Θ.

Observe that this condition implies that the associated reversed equation (cf. (2.10))

$$d\xi_s = \beta_\theta(t-s, \xi_s)\, ds + \sigma_\theta(t-s, \xi_s)\, dB_s, \quad s \in [0,t], \tag{4.12}$$

with initial condition $\xi_0 = x$ has a solution $\xi^{t,x,\theta}$ for which Lemma 4.4 applies (here t is fixed and s plays the role of t in Lemma 4.4). This follows since by Condition 4.6 the coefficients $\beta_i(t,x,\theta)$ and $\sigma_{ij}(t,x,\theta)$ are in $C_b^{1,3}([0,T] \times \mathbb{R}^d \times B)$, for every bounded open $B \subset \Theta$. Notice also that the differential operator A_t, associated to the system (1.1) and (1.3), is given by

$$\begin{aligned} A_t f(x,\theta) \;=\; & \frac{1}{2} \sum_{i,j=1}^d (\sigma\sigma^{\mathrm{T}})_{ij}(t,x,\theta) \frac{\partial^2}{\partial x_i \partial x_j} f(x,\theta) + \\ & + \sum_{i=1}^d b_i(t,x,\theta) \frac{\partial}{\partial x_i} f(x,\theta). \end{aligned}$$

It thus essentially coincides with the family of operators A_t^θ, associated to equations (1.1) indexed by $\theta \in \Theta$. We can now extend Theorem 4.2 to the state and parameter filtering problem defined by (1.1), (1.3), and by the measurement equation (2.2).

THEOREM 4.7. *Suppose Condition* 4.6 *holds and define* $u_t(x,\theta,\omega_1)$ *by the right-hand side of* (3.7)*, but with* $\xi_s^{t,x}$ *replaced by* $\xi_s^{t,x,\theta}$. *Define the random measure* ρ_t *on* $\mathbb{R}^d \times \Theta$ *by*

$$d\rho_t(x,\theta,\omega_1) := u_t(x,\theta,\omega_1)\, d\mu(\theta)\, dx. \tag{4.13}$$

Then ρ_t *is an unnormalized conditional distribution of* (X_t, θ_0) *based on observations* (2.2).

Proof. As in the proof of Theorem 4.2 it suffices to show that

$$\rho_t f = \rho_0 f + \int_0^t \rho_s(A_s f)\, ds + \int_0^t \rho_s(hf)\, dY_s, \quad P_1\text{-a.s.}, \tag{4.14}$$

holds for all $f \in C_c^\infty(\mathbb{R}^d \times \Theta)$ and for $f = 1$. For fixed $\theta \in \Theta$ and $f \in C_c^\infty(\mathbb{R}^d \times \Theta)$ we have $f^\theta := f(\cdot,\theta) \in C_c^\infty(\mathbb{R}^d)$. With $\rho_t^\theta f^\theta(\omega_1) := \int_{\mathbb{R}^d} u_t(x,\theta,\omega_1) f^\theta(x)\, dx$ Theorem 4.2 gives

$$\rho_t^\theta f^\theta = \rho_0^\theta f^\theta + \int_0^t \rho_s^\theta(A_s^\theta f^\theta)\, ds + \int_0^t \rho_s^\theta(hf^\theta)\, dY_s, \quad P_1\text{-a.s.} \tag{4.15}$$

Since $(t,x,\theta) \mapsto u_t(x,\theta,\cdot)$ is a well-defined random field the random variable

$$I^\theta := \rho_t^\theta f^\theta - \rho_0^\theta f^\theta - \int_0^t \rho_s^\theta(A_s^\theta f^\theta)\, ds \tag{4.16}$$

is well-defined for all $\omega_1 \in \Omega_1$. A version of the stochastic integral in (4.15) is thus defined by I^θ, and with this version (4.15) holds for all ω_1. (So we got rid of the θ-dependent P_1-zero set in (4.15).) Moreover, since $f \in C_c^\infty$ each term in (4.16) is continuous in θ. So we can integrate (4.15) term by term over $\mathrm{d}\mu(\theta)$. This gives

$$\begin{aligned}\int_\Theta \int_{\mathbb{R}^d} & u_t(x,\theta,\omega_1) f(x,\theta)\,\mathrm{d}x\,\mathrm{d}\mu(\theta) \\ &= \int_\Theta \int_{\mathbb{R}^d} p_0(x) f(x,\theta)\,\mathrm{d}x\,\mathrm{d}\mu(\theta) + \\ &\quad + \int_\Theta \int_0^t \int_{\mathbb{R}^d} u_s(x,\theta,\omega_1) A_s f(x,\theta)\,\mathrm{d}x\,\mathrm{d}s\,\mathrm{d}\mu(\theta) + \\ &\quad + \int_\Theta \int_0^t \int_{\mathbb{R}^d} u_s(x,\theta,\omega_1) h(x) f(x,\theta)\,\mathrm{d}x\mathrm{d}Y_s\,\mathrm{d}\mu(\theta). \qquad (4.17)\end{aligned}$$

Since $(x,\theta) \mapsto u_t(x,\theta,\omega_1)$ and $x \mapsto p_0(x)$ are continuous, and $f \in C_c^\infty$, the first two terms in (4.17) coincide with the ones in (4.14). In view of the L^p-bound (4.8) for $u_t(x,\theta,\cdot)$, which is uniform in (x,θ), we can apply [24, Lemma 3.2] to interchange the integrals in the third term in (4.17). This gives the third term in (4.14). Applying again the uniform L^p-bound together with Itô's isometry to the stochastic integral in (4.15) shows that the integrability conditions of [24, Lemma 3.3] are satisfied. This allows to interchange the order of integration in the last term of (4.17), which finally gives the last term in (4.14). The assertion (4.14) for $f = 1$ follows by a limit argument similar to the one in the proof of Theorem 4.2. □

Remarks. (1) In general ρ_t does not have a Lebesgue density, but a density with respect to $\mathrm{d}\mu(\theta)\,\mathrm{d}x$. For numerical purposes it is convenient that $\mathrm{d}\mu(\theta)$ only needs to have compact support: One can choose for $\mathrm{d}\mu$, e.g., a discrete, uniform measure on a sufficiently fine grid in a bounded subset of Θ. The resulting integrals over $\mathrm{d}\mu(\theta)$ are then finite sums, so there is no numerical error due to an approximation of $\mathrm{d}\mu$-integrals.

(2) Theorem 4.7 in fact gives the solution to the problem of *combined state and parameter estimation*. This problem has been considered for discrete time processes with regular diffusion matrices σ depending only on θ in [15]. In that work a discrete version of the Bayes formula (2.3) and a discrete time Zakai equation has been exploited.

5. Robust Parameter Estimation

In this section we construct an estimator for the parameter θ in (1.1), based on the robust FK-formula (3.7). Moreover, we establish continuity of the resulting estimator with respect to observations Y and with respect to the size of the error in the measurement model.

Let $\alpha > 0$ be fixed. We scale the observation $Y_t = \int_0^t h(X_s)\,\mathrm{d}s + \alpha W_t^{(2)}$ as follows: $Y_t^\alpha := Y_t/\alpha$. Since $\sigma(Y_s^\alpha, 0 \leqslant s \leqslant t) = \sigma(Y_s, 0 \leqslant s \leqslant t) = \mathcal{Y}_t^\alpha$, we can determine $E[\theta_0|\mathcal{Y}_t^\alpha]$ by applying Theorem 4.7 with respect to the measurement process

$$Y_t^\alpha := \int_0^t h_\alpha(X_s)\,\mathrm{d}s + W_t^{(2)}, \tag{5.1}$$

with $h_\alpha := h/\alpha$. In particular, instead of (3.4) we obtain

$$U_t^\alpha(x,\theta,Y) := E_{P_2}\biggl[p_0(\xi_t^{x,\theta})\exp\biggl(\int_0^t C_\alpha(s,\xi_{t-s}^{x,\theta})\,\mathrm{d}s + \\ + \frac{1}{\alpha^2}\int_0^t Y_{t-s}\,\mathrm{d}h(\xi_s^{x,\theta})\biggr)\biggr], \tag{5.2}$$

with $\xi_u^{x,\theta} := \xi_u^{t,x,\theta}$ and $C_\alpha(s,x) := c(s,x) - \frac{1}{2\alpha^2}h(x)^2$. Observe that Y_t (not Y_t^α) appears in (5.2). Recall from (3.11) that $\mathrm{d}h_k(\xi_s) = \gamma_k(s,\xi_s)\,\mathrm{d}s + g_{kj}(s,\xi_s)\,\mathrm{d}B_s^j$. This gives

$$U_t^\alpha(x,\theta,Y) = E_{P_2}\biggl[p_0(\xi_t^{x,\theta})\exp\biggl(\int_0^t C_\alpha(s,\xi_{t-s}^{x,\theta})\,\mathrm{d}s + \\ + \int_0^t Y_{t-s}\,\gamma^\alpha(\xi_s^{x,\theta})\,\mathrm{d}s + \int_0^t Y_{t-s}\,g^\alpha(\xi_s^{x,\theta})\,\mathrm{d}B_s\biggr)\biggr], \tag{5.3}$$

with vector $\gamma^\alpha := (\gamma_k^\alpha) = (\gamma_k/\alpha^2)$ and matrix $g^\alpha := (g_{kj}^\alpha) = (g_{kj}/\alpha^2)$. From (3.7) it is clear that the unnormalized conditional $\mathrm{d}\mu(\theta)\,\mathrm{d}x$-density for the system (X_t,θ_t) reads

$$u_t^\alpha(x,\theta,Y) := \mathrm{e}^{h(x)Y_t/\alpha^2}U_t^\alpha(x,\theta,Y). \tag{5.4}$$

This yields the normalized joint $\mathrm{d}\mu(\theta)\,\mathrm{d}x$-density

$$p_t^\alpha(x,\theta,Y) = \frac{u_t^\alpha(x,\theta,Y)}{\int u_t^\alpha(x,\theta,Y)\,\mathrm{d}\mu(\theta)\,\mathrm{d}x}, \tag{5.5}$$

provided $0 < \int u_t^\alpha(x,\theta,Y)\,\mathrm{d}\mu(\theta)\,\mathrm{d}x < \infty$. From this the $\mathrm{d}\mu(\theta)$-density of $\hat{\theta}_t^\alpha$ follows:

$$p_t^\alpha(\theta,Y) = \int_{\mathbb{R}^d} p_t^\alpha(x,\theta,Y)\,\mathrm{d}x. \tag{5.6}$$

Remark 5.1. In view of $p_t^\alpha(x,\theta,Y(\omega_1)) = E_{P_1}[\Lambda_t|\mathcal{Y}_t]u_t^\alpha(x,\theta,Y(\omega_1))$ and $\int p_t^\alpha(x,\theta,Y(\omega_1))\,\mathrm{d}x\,\mathrm{d}\mu(\theta) = 1$ (P_1-a.s.) we see that $\int u_t^\alpha(x,\theta,Y(\omega_1))\,\mathrm{d}x\,\mathrm{d}\mu(\theta) > 0$ P_1-a.s. But since we have chosen a specific version of u_t^α it may happen that the denominator in (5.5) vanishes for certain values of Y. In fact this is not the case, as the following main theorem of this section shows.

THEOREM 5.2. *Assume Condition* 4.6 *is satisfied, define* u_t^α *by* (5.4)*, and let* $Y \in B[0,T]^m$. *Then*

$$0 < \int u_t^\alpha(x,\theta,Y)\,\mathrm{d}\mu(\theta)\,\mathrm{d}x < \infty, \tag{5.7}$$

and a version of the conditional expectation $\hat{\theta}_t^\alpha(\omega_1) = E[\theta_0|\mathcal{Y}_t^\alpha](\omega_1)$ *is defined by* $S_t^\alpha(Y(\omega_1))$ *with*

$$S_t^\alpha(Y) = \int_\Theta \theta p_t^\alpha(\theta,Y)\,\mathrm{d}\mu(\theta) = \frac{\int \theta u_t^\alpha(x,\theta,Y)\,\mathrm{d}x\,\mathrm{d}\mu(\theta)}{\int u_t^\alpha(x,\theta,Y)\,\mathrm{d}x\,\mathrm{d}\mu(\theta)}. \tag{5.8}$$

Moreover, the map $(\alpha, Y) \mapsto S_t^\alpha(Y)$ *is continuous on* $(0,\infty) \times B[0,T]^m$.

As an immediate conclusion note that $E[\theta_0|\mathcal{Y}_t^\alpha] = S_t^\alpha(Y(\cdot))$ P_1-a.s., and moreover $\alpha \to \alpha_0$ implies $S_t^\alpha(Y(\omega_1)) \to S_t^{\alpha_0}(Y(\omega_1))$, for all $\omega_1 \in \Omega_1$. Since the family $\{E[\theta_0|\mathcal{Y}_t^\alpha], \alpha > 0\}$ is uniformly integrable Vitali's characterization of L^1-convergence implies:

COROLLARY 5.3. *Suppose* α *converges to* $\alpha_0 > 0$. *Then* $E[\theta_0|\mathcal{Y}_t^\alpha]$ *converges to* $E[\theta_0|\mathcal{Y}_t^{\alpha_0}]$ *in* $L^1(P)$.

Remarks. (1) Theorem 5.2 makes no statement about the (singular) case $\alpha_0 = 0$, which is of interest in state estimation. This case has been investigated in a number of works, see for example [6, 39], and references given there.

(2) Corollary 5.3 is a continuity statement for conditional expectations $E[\theta|\mathcal{Y}]$ with respect to "variable" $\mathcal{Y}$. Besides martingale convergence it seems that only few results are known for this kind of continuity, see [8, 16, 35].

Before we prove Theorem 5.2 we give two preparations:

LEMMA 5.4 (Robustness of the conditional density). *Assume Condition* 4.6 *holds, define* u_t^α *by* (5.4)*, and fix* $t > 0$. *Then the map*

$$(\alpha, x, \theta, Y) \mapsto u_t^\alpha(x,\theta,Y)$$
$$\text{is continuous on } (0,\infty) \times \mathbb{R}^d \times \Theta \times B[0,T]^m. \tag{5.9}$$

Proof. In view of (5.4) it suffices to check the continuity of $(\alpha, x, \theta, Y) \mapsto U_t^\alpha(x,\theta,Y)$. We first show that for fixed Y the map

$$(\alpha, x, \theta) \mapsto U_t^\alpha(x,\theta,Y) \text{ is continuous at } (\bar{\alpha}, \bar{\theta}, \bar{x}). \tag{5.10}$$

By Lemma 4.4, for fixed $\omega_2 \in \Omega_2$, the map $(x,\theta) \mapsto \xi_t^{x,\theta}(\omega_2)$ is in $C^2(\mathbb{R}^d \times \Theta)$. Therefore the terms $p_0(\xi_t^{x,\theta}(\omega_2))$, $\int_0^t C_\alpha(s, \xi_{t-s}^{x,\theta})\,\mathrm{d}s$ and $\int_0^t Y_{t-s}\gamma^\alpha(\xi_s^{x,\theta})\,\mathrm{d}s$ in (5.3) are continuous. We check that a suitable version of $\int_0^t Y_{t-s} g^\alpha(\xi_s^{x,\theta})\,\mathrm{d}B_s$ is continuous in (x,θ). Define

$$Z_u(x,\theta) := \int_0^u Y_{t-s} g^\alpha(\xi_s^{x,\theta})\,\mathrm{d}B_s, \quad u \in [0,t].$$

Choose $\delta > 0$ such that the closed ball $B_\delta[\bar{x}, \bar{\theta}]$ is contained in $\mathbb{R}^d \times \Theta$. Then the process $f: s \mapsto Y_{t-s} g^\alpha(X_s^{x,\theta})$ takes values $f_s(\omega_2) \in C_b^1(B_\delta[\bar{x}, \bar{\theta}])$. The continuity of $(s, x, \theta) \mapsto f_s(\omega_2)(x, \theta)$ on $[0, t] \times B_\delta[\bar{x}, \bar{\theta}]$ implies that also

$$s \mapsto \|f_s(\omega_2)\|_{\infty,1} := \sum_{k=0}^{1} \sup\{|D^k f_s(\omega_2)(x, \theta)|, (x, \theta) \in B_\delta[\bar{x}, \bar{\theta}]\}$$

is continuous. In particular $\int_0^u \|f_s(\omega_2)\|_{\infty,1}^p \, \mathrm{d}s < \infty$, for all $\omega_2 \in \Omega_2$, $p \geqslant 2$. So the conditions of [27, Theorem 7.6, p. 180] are satisfied and $(u, x, \theta) \mapsto Z_u(x, \theta)$ has a continuous modification. In particular $(x, \theta) \mapsto Z_t(x, \theta)$ has a continuous modification. So we conclude that the integrand in (5.3) has a modification which is continuous in (α, x, θ). To infer (5.10) we first note that the integrand in (5.3) is L^2-bounded, uniformly on a small ball $B_r(\bar{\alpha}, \bar{x}, \bar{\theta})$. Up to simple modifications this is verified by the same estimation as given in the proof of Theorem 3.5. This uniform L^2-boundedness implies uniform integrability, and thus the (parametric) continuity of the integrand in (5.3) turns over to the continuity of the integral, by Vitali's theorem. Thus (5.10) holds. To infer (5.9) fix $(\bar{\alpha}, \bar{x}, \bar{\theta}, \bar{Y})$ and proceed as follows:

$$\begin{aligned} |U_t^\alpha(x, \theta, Y) - U_t^{\bar{\alpha}}(\bar{x}, \bar{\theta}, \bar{Y})| \leqslant\ & |U_t^\alpha(x, \theta, Y) - U_t^\alpha(x, \theta, \bar{Y})| + \\ & + |U_t^\alpha(x, \theta, \bar{Y}) - U_t^{\bar{\alpha}}(\bar{x}, \bar{\theta}, \bar{Y})|. \end{aligned} \tag{5.11}$$

A simple modification of the estimate (3.16) shows that the first term on the right side of (5.11) is dominated by $\|Y - \bar{Y}\|_\infty g(\alpha, \|Y\|_\infty, \|\bar{Y}\|_\infty)$ (uniformly in x, θ) where g denotes a continuous function. Thus the first term in (5.11) goes to 0 as $Y \to \bar{Y}$. In view of (5.10) also the second term goes to zero as (α, x, θ) goes to $(\bar{\alpha}, \bar{x}, \bar{\theta})$. So (5.11) yields (5.9). □

LEMMA 5.5. *Assume Condition* 4.6 *holds, and define* u_t^α *by* (5.4)*. Then*

$$0 < \int_{\mathbb{R}^d} u_t^\alpha(x, \theta, Y) \, \mathrm{d}x < \infty \tag{5.12}$$

holds for all $(t, \alpha, \theta, Y) \in [0, T] \times (0, \infty) \times \Theta \times B[0, T]^m$. *Moreover, for fixed* $t \in [0, T]$ *the map*

$$(\alpha, \theta, Y) \mapsto u_t^\alpha(\theta, Y) := \int_{\mathbb{R}^d} u_t^\alpha(x, \theta, Y) \, \mathrm{d}x \text{ is continuous.} \tag{5.13}$$

Proof. Fix (α, t, θ, Y) and put $u(x) := u_t^\alpha(x, \theta, Y)$. Then $u := \int_{\mathbb{R}^d} u(x) \, \mathrm{d}x < \infty$ by (4.2). This is the upper bound in (5.12). To verify the lower bound assume there exists (t, α, θ, Y) such that $\int_{\mathbb{R}^d} u_t^\alpha(x, \theta, Y) \, \mathrm{d}x = 0$. By continuity and since $u_t^\alpha \geqslant 0$ it follows that $u_t^\alpha(x, \theta, Y) = 0$ for all $x \in \mathbb{R}^d$, so $E_{P_2}[p_0(\xi_t^{x,\theta}) \mathrm{e}^{\cdots}] = 0$ for all x. Thus $E_{P_2}[p_0(\xi_t^{x,\theta})] = 0$ for all x. Integration over dx and Fubini's Theorem yield $E_{P_2}[\int_{\mathbb{R}^d} p_0(\xi_t^{x,\theta}) \, \mathrm{d}x] = 0$, that is

$$\int_{\mathbb{R}^d} p_0(\xi_t^{x,\theta}) \, \mathrm{d}x = 0 \quad P_2\text{-a.s.} \tag{5.14}$$

In the theory of stochastic flows one shows (see, for example [26]) that $x \mapsto \xi_t^{x,\theta}(\omega_2)$ is a homeomorphism of $\mathbb{R}^d$ for all $\omega_2 \in A$, with $P_2(A) = 1$. In particular, to $\omega_2 \in A$ and $y \in \mathbb{R}^d$ satisfying $p_0(y) > 0$ there is an $x \in \mathbb{R}^d$ such that $y = \xi_t^{x,\theta}(\omega_2)$. By continuity of $x \mapsto p_0(\xi_t^{x,\theta})(\omega_2)$ and since $p_0 \geqslant 0$, it follows that $\int_{\mathbb{R}^d} p_0(\xi_t^{x,\theta})(\omega_2)\,\mathrm{d}x > 0$ for all $\omega_2 \in A$. This contradicts (5.14), so the lower bound in (5.12) must in fact hold for all (t, α, θ, Y).

Next, fix $(t, \bar{\alpha}, \bar{\theta}, \bar{Y})$ and choose $\varepsilon > 0$. Let $(\alpha_n, \theta_n, Y_n)$ converge to $(\bar{\alpha}, \bar{\theta}, \bar{Y})$. Abbreviate $u_n(x) := u_t^{\alpha_n}(x, \theta_n, Y_n)$ and $u_n := \int_{\mathbb{R}^d} u_n(x)\,\mathrm{d}x$. Then

$$\begin{aligned} |u_n - u| \;\leqslant\; & \int_{|x|\leqslant N} |u_n(x) - u(x)|\,\mathrm{d}x + \\ & + \int_{|x|>N} |u_n(x) - u(x)|\,\mathrm{d}x, \quad \forall n, N \in \mathbb{N}. \end{aligned} \tag{5.15}$$

The proof of Lemma 4.1 shows that the estimate (4.2) holds uniformly in a small neighborhood of $(\bar{\alpha}, \bar{\theta}, \bar{Y})$. Thus, for sufficiently big N the last term in (5.15) is less that $\varepsilon/2$, for all $n \geqslant n_0$. For this N the first term, T_n, on the right side of (5.15) is estimated by

$$\begin{aligned} T_n \;\leqslant\; & \int_{|x|\leqslant N} |u_t^{\alpha_n}(x, \theta_n, Y_n) - u_t^{\alpha_n}(x, \theta_n, Y)|\,\mathrm{d}x + \\ & + \int_{|x|\leqslant N} |u_t^{\alpha_n}(x, \theta_n, Y) - u_t^{\alpha}(x, \theta, Y)|\,\mathrm{d}x \\ \leqslant\; & c \cdot \|Y_n - Y\|_\infty g(\alpha_n, \|Y_n\|_\infty, \|Y\|_\infty) + \\ & + \int_{|x|\leqslant N} |u_t^{\alpha_n}(x, \theta_n, Y) - u_t^{\alpha}(x, \theta, Y)|\,\mathrm{d}x, \end{aligned} \tag{5.16}$$

where $c = (2N)^d$ and g denotes a continuous function (cf. (3.16)). Clearly the first term goes to zero but also the second term does because, with Lemma 5.4, the map $(\alpha, x, \theta) \mapsto u_t^{\alpha}(x, \theta, Y)$ is uniformly continuous on the compact $B_r[(\bar{\alpha}, \bar{\theta})] \times [-N, N]^d$ (with suitable $r > 0$). Thus, the first term in (5.15) is less than $\varepsilon/2$, for all $n \geqslant n_1$. This proves (5.13). □

Proof of Theorem 5.2. The bounds (5.7) follow from (5.12), from the continuity of the function $\theta \mapsto u_t^{\alpha}(\theta, Y) := \int_{\mathbb{R}^d} u_t^{\alpha}(x, \theta, Y)\,\mathrm{d}x$, and because μ has compact support. In particular (5.5), (5.6) and (5.8) are well-defined. From Theorem 4.7 we know that $u_t^{\alpha}(x, \theta, Y(\omega_1))$ is a version of the unnormalized conditional density and therefore (5.5) and (5.6) define versions of the corresponding conditional densities. This implies the second assertion in Theorem 5.2. Finally, to prove the last assertion it suffices to show that for any continuous function f the map

$$(\alpha, Y) \mapsto \int_{\Theta} f(\theta) u_t^{\alpha}(\theta, Y)\,\mathrm{d}\mu(\theta) \tag{5.17}$$

is continuous (choose $f(\theta) = \theta$ respectively $f(\theta) = 1$ in (5.8)). So let (α_n, Y_n) converge to (α, Y). Since the support C of μ is compact the continuity of (5.17) follows if $\sup_{\theta\in C} |u_t^{\alpha_n}(\theta, Y_n) - u_t^{\alpha}(\theta, Y)| \to 0$ as $n \to \infty$. But this can be inferred from a slight modification in the proof of Lemma 5.5 as follows: First put $\theta_n = \theta$ in that proof. Then observe that the second term on the right side of (5.15) can be estimated by $\varepsilon/2$, uniformly in θ, as follows from (4.2) together with the compactness of C. The first term in (5.15) is estimated as in (5.16). Now the first term in (5.16) does not depend on θ while the second term converges uniformly for $\theta \in C$ because $(\alpha, \theta) \mapsto u_t^{\alpha}(\theta, Y)$ is uniformly continuous on the compact $[\alpha - r, \alpha + r] \times C$ (with a fixed $r \in (0, \alpha)$). □

6. An Algorithm for the Estimation

In this section we study the practical implementation of the estimator (5.8). We first express the unnormalized density $u_t^{\alpha}(x, \theta, Y)$ from (5.4) in a recursive form which is more suitable for numerical computations. We then combine this with a Monte-Carlo approach, to construct a numerical approximation of the estimator $S_t^{\alpha}(Y)$ in (5.8). Finally we present the resulting estimation procedure in algorithmic form, and discuss some of its general properties.

LEMMA 6.1 (Recursive formulations of FK). *Suppose Condition* 2.3 *holds, and let* $0 \leqslant t_0 < t$. *Then* $u_t(x, \omega_1)$*, given in* (2.8)*, can be written as*

$$u_t(x, \omega_1) = E_{P_2}\left[u_{t_0}(\xi^{t,x}_{t-t_0}, \omega_1) \exp\left\{ \int_{t_0}^{t} C(s, \xi^{t,x}_{t-s})\, \mathrm{d}s + \right.\right.$$
$$\left.\left. + \int_{t_0}^{t} h(\xi^{t,x}_{t-s})\, \mathrm{d}Y_s(\omega_1) \right\}\right] \tag{6.1}$$
$$= E_{P_2}\big[u_{t_0}(\xi^{t,x}_{t-t_0}, \omega_1) \exp\{(*)\}\big], \quad P_1\text{-}a.s., \tag{6.2}$$

where the exponent $(*)$ *is given by*

$$(*) = h(x)Y_t - h(\xi^{t,x}_{t-t_0})Y_{t_0} + \int_{t_0}^{t} C(s, \xi^{t,x}_{t-s})\, \mathrm{d}s + \int_0^{t-t_0} Y_{t-s}\, \mathrm{d}h(\xi^{t,x}_s).$$

Proof. A decomposition of $\int_0^t$ into $\int_0^{t_0} + \int_{t_0}^t$ in the Zakai density equation (2.6) leads to

$$u_t(x, \omega_1) = u_{t_0}(x, \omega_1) + \int_{t_0}^{t} (L_s + c(s, x))u_s(x, \omega_1)\, \mathrm{d}s +$$
$$+ \int_{t_0}^{t} h(x)u_s(x, \omega_1)\, \mathrm{d}Y_s(\omega_1), \tag{6.3}$$

where $u_{t_0}(x, \omega_1)$ is given in (2.8). Now we do a time substitution $s \mapsto s + t_0$, write $\tilde{u}_t := u_{t+t_0}$, $\tilde{Y}_s := Y_{s+t_0} - Y_{t_0}$, and replace c by $\tilde{c}(u, x) := c(u + t_0, x)$ (similarly for σ, β in L). Then (6.3) gives

$$\begin{aligned}\tilde{u}_{t'}(x, \omega_1) &= \tilde{u}_0(x, \omega_1) + \int_0^{t'} (\tilde{L}_s + \tilde{c}(s, x))\tilde{u}_s(x, \omega_1)\,\mathrm{d}s + \\ &\quad + \int_0^{t'} h(x)\tilde{u}_s(x, \omega_1)\,\mathrm{d}\tilde{Y}_s(\omega_1), \end{aligned}\tag{6.4}$$

with $t' := t - t_0$. This is again the Zakai density equation (2.6) with the given substitutions. The FK-formula therefore applies, where now we have to use the (transformed) reverse equation

$$\begin{aligned}\mathrm{d}\tilde{\xi}_s^{t',x} &= \tilde{\beta}(t' - s, \tilde{\xi}_s^{t',x})\,\mathrm{d}s + \tilde{\sigma}(t' - s, \tilde{\xi}_s^{t',x})\,\mathrm{d}B_s \\ &= \beta(t - s, \tilde{\xi}_s^{t',x})\,\mathrm{d}s + \sigma(t - s, \tilde{\xi}_s^{t',x})\,\mathrm{d}B_s, \quad \xi_0^{t',x} = 0.\end{aligned}$$

Since this is identical with the original reversed equation we have $\tilde{\xi}_s^{t',x} = \xi_s^{t,x}$. The reverse time substitution, $s \mapsto s - t_0$ performed in the resulting FK-formula, finally gives (6.1). (6.2) is now a direct consequence of Lemma 3.1 and the discussion leading to the robust formulation (3.7). □

Remark 6.2. Formula (6.1) is useful for two reasons. When estimates are to be updated as new observations become available, the form (6.1) allows to base the next estimate on the previously calculated $u_{t_0}(x, \omega_1)$ and the values of Y on $[t_0, t]$. (This is why we call (6.1) "recursive".) Moreover, for computations the formulas for $u_t(x, \omega_1)$ in (2.8), (3.7) and (5.4), have a serious drawback: The exponent in these formulas can easely become so big that it leads to representation problems (i.e., floating point overflow) in the computer. This problem is avoided if we normalize the calculated $u_t(x, \omega_1)$ at suitable time-intervals, and restart our calculation using (6.1).

Numerical Estimation. By applying the same argument as in Section 5, we find that the unnormalized conditional density for (X_t, θ_t) in (4.10), (4.11) can be written as

$$u^\alpha(t, x, \theta, Y) = E_{P_2}\big[u^\alpha(t_0, \xi_{t-t_0}^{x,\theta}, \theta, Y)\exp\{(*)_\alpha\}\big],\tag{6.5}$$

where the exponent $(*)_\alpha$ is given by

$$\begin{aligned}(*)_\alpha &= \frac{1}{\alpha^2}(h(x)Y_t - h(\xi_{t-t_0}^{x,\theta})Y_{t_0}) + \int_{t_0}^{t} C_\alpha(s, \xi_{t-s}^{x,\theta})\,\mathrm{d}s + \\ &\quad + \frac{1}{\alpha^2}\int_0^{t-t_0} Y_{t-s}\,\mathrm{d}h(\xi_s^{x,\theta}),\end{aligned}$$

with C_α introduced in (5.2). Using (6.5) we construct a numerical approximation for the estimator $S_t^\alpha(Y)$ as follows. We assume the measurements are given at times $0 = t_0 < t_1 < \cdots < t_{N_T} = T$, and we choose a set of grid-points $\{(x_i, \theta_j) : i = 1, \ldots, M_X, j = 1, \ldots, M_\theta\}$ in $\mathbb{R}^d \times \Theta$. Then we approximate $u^\alpha(t_n, \cdot, \cdot, Y)$ on this grid using a Monte-Carlo based approach, that is, in each (x_i, θ_j) we estimate $u^\alpha(t_n, x_i, \theta_j, Y)$ by averaging over a number of simulations of the associated reversed process ξ given in (4.12), which starts in (x_i, θ_j). Once we have estimated $u^\alpha(t_n, x_i, \theta_j, Y)$, we approximate $S_{t_n}^\alpha(Y)$ using numerical integration in (5.8). In order to stabilize the method, we normalize and reset the estimate $\hat{u}^\alpha(t_n, \cdot, \cdot, Y)$ using (6.5) at suitable time-intervals.

The *numerical details* are as follows: Consider the nth time step in the above described procedure. Let $t_{\hat{n}}$ denote the last time-instant when we normalized $\hat{u}^\alpha$, and assume that we have constructed R independent paths of ξ solving (4.12) on the time-grid $s_k := t_n - t_{n-k}$ $(k = 0, \ldots, n - \hat{n})$. Then we approximate the exponent in (6.5) using the scheme

$$(*)^{(r)}_{\alpha,\hat{n},n} = \sum_{k=0}^{n-\hat{n}-1} C_\alpha\big(t_n - s_k, \xi_k^{(r)}\big)\Delta s_k + \\ + \frac{1}{\alpha^2}\sum_{k=1}^{n-\hat{n}}\big(h_k^{(r)}\Delta Y_{n-k} - \Delta Y_{n-k}\Delta h_{k-1}^{(r)}\big). \tag{6.6}$$

Here $\xi_k^{(r)}$ denotes the rth simulation of ξ^{t_n, x_i, θ_j} evaluated in s_k, and we abbreviated $h_k^{(r)} = h(\xi_k^{(r)})$, $\Delta h_k^{(r)} = h_{k+1}^{(r)} - h_k^{(r)}$, $\Delta Y_k = Y_{t_{k+1}} - Y_{t_k}$, and $\Delta s_k = s_{k+1} - s_k$. The expectation value in (6.5) can now be approximated by averaging over the simulated paths

$$\hat{u}^\alpha(t_n, x_i, \theta_j, Y) := \frac{1}{R}\sum_{r=1}^{R} u^\alpha(t_{\hat{n}}, \xi^{(r)}_{t_n - t_{\hat{n}}}, \theta_j, Y)\exp\{(*)^{(r)}_{\alpha,\hat{n},n}\}. \tag{6.7}$$

Our *numerical estimator* is finally given by the following discrete version of (5.8):

$$\hat{S}_{t_n}^\alpha(Y) = \frac{\sum_j^{M_\theta}\sum_i^{M_X}\theta_j\hat{u}^\alpha(t_n, x_i, \theta_j, Y)}{\sum_j^{M_\theta}\sum_i^{M_X}\hat{u}^\alpha(t_n, x_i, \theta_j, Y)}. \tag{6.8}$$

Remarks. (1) The scheme in (6.6) is the result of approximating the integrals in (6.5) using piece-wise approximations of the form

$$\eta_s(\omega_1) = \sum_{k=0}^{N-1}\eta_{s_k}(\omega_1)1_{(s_k, s_{k+1}]}(s)$$

for $\eta = \xi$ and $\eta = Y$. Note that this approximation of the exponent differs from the corresponding one for the exponent of the FK-formula (2.8) by the 'discrete joint quadratic variation terms' $\sum_{k=1}^{n-\hat{n}}\Delta Y_{n-k}\Delta h_{k-1}/\alpha^2$.

(2) Formula (6.8) assumes a uniform density for θ_0, which holds in all our examples. If θ_0 is not uniformly distributed one has to modify this formula accordingly.

Our approach can be summarized as follows:

ALGORITHM 6.3.

(1) Fix a normalization time $\delta > 0$ and set $\hat{n} = 0$.
(2) For each time-step t_n
  (2.1) For each grid-point (x_i, θ_j)
    (2.1.1) Simulate R independent paths $\xi^{t_n, x_i, \theta_j, (r)}$ $(r = 1, \dots, R)$.
    (2.1.2) Form the sum in (6.7) to find $\hat{u}^\alpha(t_n, x_i, \theta_j, Y)$.
  (2.2) Approximate $S^\alpha_{t_n}(Y)$ in (5.8) by the numerical estimator (6.8).
  (2.3) If $t_n - t_{\hat{n}} > \delta$
    (2.3.1) Approximate the $L^1(\mathbb{R}^d \times \Theta, \mathrm{d}x\, \mathrm{d}\mu)$-norm of $\hat{u}^\alpha(t_n, \cdot, \cdot, Y)$.
    (2.3.2) Set $\hat{n} = n$ and
      $\hat{u}^\alpha(t_n, x_i, \theta_j, Y) := \hat{u}^\alpha(t_n, x_i, \theta_j, Y)/\|\hat{u}^\alpha(t_n, \cdot, \cdot, Y)\|_1$.
(3) Plot and visualize the results.

Comments on the algorithm. The accuracy of the numerical estimator clearly depends on the choice of grid and the number R of simulations for ξ in each grid point. If the grid is too coarse, the accuracy in the numerical integration in Step 2.2 might suffer, and if R is too small, the estimate for u in each grid-point has less accuracy. On the other hand, each time-step requires $M_X M_\theta R$ simulations of the associated reversed system (4.12), so choosing too fine grid or high R can make the estimator impractical. The time interval δ for normalization should be chosen with some care. If it is too short (causing normalization at each time-step) the computational effort increases unnecessarily. If it is too long we risk numerical instabilities in the estimator. An adaptive approach could be considered here. The most computational parts of the algorithm are Step 2.1.1 and Step 2.1.2. In our implementations we simulated ξ from (4.12) using a strongly convergent scheme (most of the examples can actually be solved exactly). It would suffice for the evaluation of (6.7) to use a weak approximation of ξ, which is faster to generate (see, e.g., [25]). Note the inherit parallelism in the algorithm, in the sense that for a given time-step t_n, the value of u on a grid-point (x_i, θ_j) can be approximated independently of the value of u on the other grid-points. Also note that the simulations of ξ may be precomputed ('off-line') and stored, to be readily available. Thus, there is the potential for a considerable gain in speed if implementing the algorithm on a parallel computer, in particular for higher-dimensional problems. Finally, in our examples, we have assumed a rectangular grid combined with the midpoint method for the numerical integration in Step 2.2 and Step 2.3.1. Such an integration is not optimal with regards to numerical error, but is suitable for the purpose of illustrating Algorithm 6.3. For higher-dimensional problems, other griding techniques could be considered.

7. Simulation Examples

In this section we study three specific estimation problems for simulated system processes X, based on simulated measurement data Y. In all examples X is a one-dimensional process, and θ is a single real parameter. For convenience we picked the system processes (1.1) such that they can be represented as explicit functions of Brownian motion. We plot and discuss some of our simulation results together with (sample) statistical properties for each example.

Let us first describe the common features of the three numerical experiments. In each example we fix one or two values θ_{real} for the parameter to be estimated, together with the finishing time $T > 0$. We assume observations are given by (5.1) with $\alpha = 1$. There are then two different types of simulations to be done. The *first type* is the simulation of M observation paths $Y^{(1)}, \dots, Y^{(M)}$. [With real world data we would typically have only one observation path to estimate θ_{real}. But here we use M simulations in order to make an empirical statistical analysis to see how good our estimator performs.] In view of (5.1) the kth simulation $Y^{(k)}$ first requires a simulation of a sample path of $(X_t)_{t\in[0,T]}$ (for which we use its explicit representation by BM $W_t^{(1)}$), a simulation of a sample path of $(W_t^{(2)})_{t\in[0,T]}$, and finally a numerical integration in (5.1). The *second type* of simulation is due to an application of Algorithm 6.3: For each measurement $Y^{(k)}$ we apply this algorithm to estimate θ_{real}, which requires for each grid point (x_i, θ_j) a simulation of the reversed process ξ, defined by (4.12). We estimate $u(t_n, x_i, \theta_j, Y)$ using (6.7), and our approximated estimator is given by (6.8). The result of this procedure are time dependent θ-estimates $\hat{S}^1_{t_n}(Y^{(1)}), \dots, \hat{S}^1_{t_n}(Y^{(M)})$ for $n = 1, \dots, N_T$, with N_T depending on the example. We investigate how Algorithm 6.3 performs, by varying numerical parameters: The Monte-Carlo parameter R (i.e. the number of simulations for the reversed process ξ per grid point), the grid parameters $(x_i, \theta_j) \in [A, B] \times [C, D]$, and different values for A, B, C, D. We assume that observations are made at times $t_n = n\Delta t$ $(n = 0, \dots, N_T)$, with time-steps $\Delta t = T/N_T$. For each experiment we present a table which gives the sample mean and standard deviation of the independent estimates $\hat{S}^1_T(Y^{(1)}), \dots, \hat{S}^1_T(Y^{(M)})$, for different choices of numerical parameters. This table shows how good the estimator performed in the underlying experiment, and also indicates a potential bias. Besides this statistical analysis we plot the time development of some typical samples $(\hat{S}^1_{t_n}(Y^{(k)}))_{n=1,\dots,N_T}$ for each experiment, together with the corresponding state estimates. As a rule we observe that the estimates stabilize after a certain time, and afterwards exhibits only small fluctuations. We also plot a typical time development of the numerical estimator for the conditional $\hat{\theta}_t$-density, i.e. for four times t_n we plot the following approximation of (5.6):

$$\hat{p}(t_n, \theta_j, Y) = \frac{\sum_{i=1}^{M_X} \hat{u}(t_n, x_i, \theta_j, Y)}{\sum_{i=1}^{M_X} \sum_{k=1}^{M_\theta} \hat{u}(t_n, x_i, \theta_k, Y)}, \quad j = 1, \dots, M_\theta. \tag{7.1}$$

This conditional density in all examples develops a peak close to θ_{real}. (Notice that the mean of this peak depends on the sample, and the statistical properties of this mean are given in the corresponding table.) After each experiment we discuss the results and give some additional details. At the end we give some additional remarks. We start with a simple linear example, and proceed to more difficult nonlinear ones.

EXAMPLE 1. We consider the process $X_t = X_0 + \theta t + W_t^{(1)}$ which satisfies the linear Itô equation

$$\mathrm{d}X_t = \theta\,\mathrm{d}t + \mathrm{d}W_t^{(1)}, \quad t \in [0, T], \tag{7.2}$$

with initial condition X_0 independent of W. We assume that h in (5.1) is smooth, bounded, and satisfies $h(x) = x$ for $|x| \leqslant C$ with the constant C chosen big enough (cf. the following remark). The potential $c(t, x, \theta)$ in (2.7) vanishes, and the associated reversed system (4.12) becomes

$$\mathrm{d}\xi_s^{t,x,\theta} = -\theta\,\mathrm{d}s + \mathrm{d}B_s, \qquad \xi_0^{t,x,\theta} = x, \quad s \in [0, t],\ x, \theta \in \mathbb{R}. \tag{7.3}$$

Remarks. (1) The constant C used in the definition of h should be so big that most realizations of X and ξ remain inside the region $\{x \in \mathbb{R} : |x| \leqslant C\}$. We assume C to be roughly 10^{38}. The choice of C is important in the following sense: Since $h(\xi_s^{t,x,\theta})$ and Y appear in the exponent in (6.5) it is possible that very rare events (very big values of $\xi_s^{t,x,\theta}$ or X) can significantly affect the estimator.

(2) This filtering problem is not exactly linear, and we do not assume a Gaussian initial distribution for θ_0. Therefore the well-known Kalman–Bucy filter does not apply to compute $\hat{\theta}_t$.

EXPERIMENT 7.1. We investigate two cases, $\theta_{\text{real}} = 0$ and $\theta_{\text{real}} = 0.5$. We simulate $M = 50$ independent sample paths of (X, Y), to obtain estimates $\hat{S}_T^1(Y^{(1)}), \dots, \hat{S}_T^1(Y^{(50)})$ of θ_{real}, using different choices for the Monte-Carlo parameter R, as given in Table I. We assume observation time-steps $\Delta t = T/N_T$, with $T = 100$, and $N_T = 4096$. The parameters (x_i, θ_j), are chosen on the following rectangular grid:

$$\begin{aligned} x_i &:= A + (i - 1/2)\Delta x, \quad \Delta x := (B - A)/M_X,\ i = 1, \dots, M_X, \\ \theta_j &:= C + (j - 1/2)\Delta\theta, \quad \Delta\theta := (D - C)/M_\theta,\ j = 1, \dots, M_\theta. \end{aligned} \tag{7.4}$$

We make different choices for A, B, C, D, M_X, M_θ, as given in Table I. Moreover, we use the normalization interval $\delta = 2\Delta t$ and assume that the initial joint density is given by $u(0, x, \theta) = p_0(x)\mu_0(\theta)$, where p_0 denotes a centered Gaussian density with variance 0.25 restricted to $[A, B]$ and normalized to get a probability measure, and where μ_0 represents a uniform distribution on $[C, D]$. The simulations of the associated system (7.3) are performed using the exact solution

$$\xi_s^{t_n - t_{\hat{n}}, x_i, \theta_j} = x_i - \theta_j s + B_s, \quad \text{for } s \in [0, t_n - t_{\hat{n}}].$$

Results of Experiment 7.1. They are given in Table I (statistics) and Figures 1–2 (sample paths and densities). Table I shows that an increased number R of ξ-simulations and a finer grid improves the estimate and reduces the variance. The data indicate convergence in our approximation procedure. The choice of grid affects the θ-estimate: A too coarse grid leads, in this example, to an overestimation of θ_{real}. We find that the spatial grid should be fine enough to capture the movement in the state variable. If not, the estimates for θ_{real} suffer inaccuracies. Figure 1 shows satisfying estimates for X_t and θ_{real}. The irregular density plot in Figure 2 is due to the variance in the Monte-Carlo approximation of $\hat{u}(t_n, x_i, \theta_j, Y)$, which is proportional to $1/R$. When we increase R and the number of evaluation points (t_n, x_i, θ_j), we observe a smoothing of the estimated density, as expected. Note, however, that the estimated values in Table I seem quite good even for small R. We

Table I. Example 1: The (sample) mean and standard deviation of the estimator $\hat{S}_T^1$ for the linear case (7.2). (e−n denotes 10^{-n}).

θ_{real}	Mean	Std.dev.	$[A, B] \times [C, D]$	M_X	M_θ	R
0.0	-8.21e-2	4.51e-1	[-30,30]×[-4,4]	60	60	5
	-1.60e-2	8.77e-2		120	100	50
	-1.35e-2	8.44e-2		120	100	100
0.5	6.99e-1	1.61e-1	[-10,80]×[-2,2]	180	30	10
	7.03e-1	1.50e-1		180	100	15
	7.10e-1	1.48e-1		180	100	50
	5.67e-1	1.47e-1		250	60	25

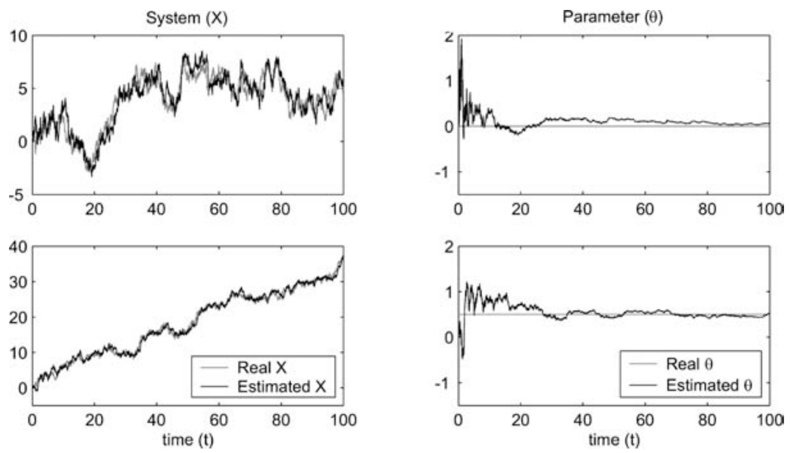

Figure 1. Example 1: The plots show two typical sample paths for the estimator of the system (X) and the parameter (θ). The first row is for the case $\theta_{\text{real}} = 0$ with $(M_X, M_\theta, R) = (120, 100, 100)$, and the second row for the case $\theta_{\text{real}} = 0.5$ with $(M_X, M_\theta, R) = (180, 100, 100)$.

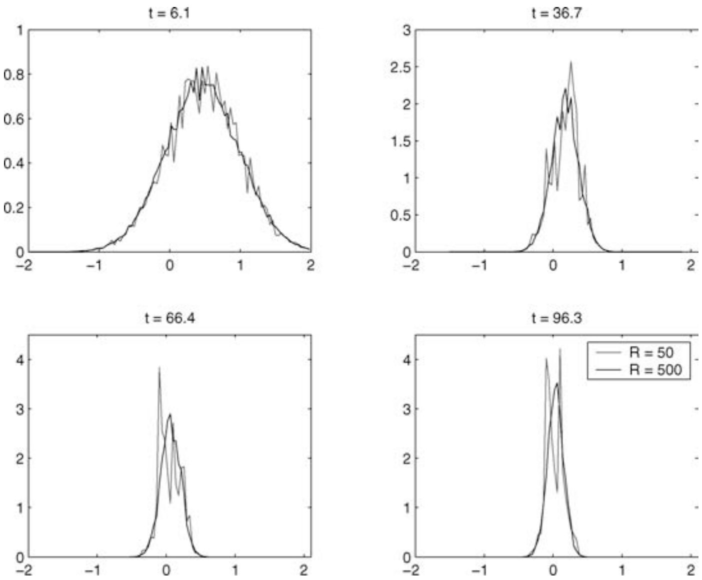

Figure 2. Example 1, case A: The approximated density $\hat{p}(t_n, \theta_j, Y)$ given in (7.1) is plotted as a function of θ_j at four different times, with two different choices of R. This plot is based on a simulation for the case $\theta_{\text{real}} = 0$ with $M_X = 120$ and $M_\theta = 200$.

believe the reason for this is the smoothing effect the integration in (6.8) has on the resulting estimator, making the local variations in the density less important for the estimate.

EXAMPLE 2. We consider geometric BM $X_t = X_0 \exp\{(\nu - \theta^2/2)t + \theta W_t^{(1)}\}$ which satisfies

$$\mathrm{d}X_t = \nu X_t\,\mathrm{d}t + \theta X_t \mathrm{d}W_t^{(1)}, \quad t \in [0, T], \tag{7.5}$$

with X_0 independent of W. We investigate two cases. In Case A we assume ν is *unknown* but equal to $\theta^2/2$. In Case B we assume ν is *known* in advance. In both cases we estimate θ on the basis of (5.1), where h is smooth, bounded and satisfies $h(x) = x$ for $|x| \leqslant C$ with the constant C chosen big enough. From (2.7) we find the potential $c(t, x, \theta, \nu) = \theta^2 - \nu$, and the associated reversed system (4.12) becomes

$$\begin{aligned} &\mathrm{d}\xi_s = (2\theta^2 - \nu)\xi_s\,\mathrm{d}s + \theta\xi_s\,\mathrm{d}B_s, \\ &\xi_0 = x, \quad s \in [0, t],\ x \in \mathbb{R},\ \theta > 0. \end{aligned} \tag{7.6}$$

Note that in Case A we have a combined drift and diffusion estimation problem, and in Case B a pure diffusion estimation problem. We would expect that the estimator performs somewhat better for Case B since in this case we have more knowledge of the process.

EXPERIMENT 7.2. For both cases we fix $\theta_{\text{real}} = 0.25$ and simulate $M = 10$ paths of (X, Y). We apply Algorithm 6.3 to estimate θ_{real}, using different choices for the grid and for the Monte-Carlo parameter R, as given in Table II. For Case B we set $\nu = \theta_{\text{real}}^2/2 = 0.03125$. Thus, when applying Algorithm 6.3 the value of ν remains fixed in this case. Our choice of ν allows to use the same simulations of Y in both cases, and we can more clearly observe the effect these assumptions have on the estimator. Observations are given with time-steps $\Delta t = T/N_T$, with $T = 100$, and $N_T = 4096$. Moreover, we assume $[A, B]\times[C, D] = [0, 60]\times[0, 2]$ and the parameters (x_i, θ_j) are chosen as in (7.4). We use the normalization interval $\delta = 10\Delta t$, and suppose the initial joint distribution is constructed as in Example 1, with normal distribution $N(1, 1/4)$ for the construction of p_0. For the simulations of (7.3) we use the exact solution given by

$$\xi_s^{t_n - t_{\hat{n}}, x_i, \theta_j} = x_i \exp\{bs + \theta B_s\}, \quad \text{where } b = \begin{cases} \theta_j^2, & \text{for Case A,} \\ 3\theta_j^2/2 - \nu, & \text{for Case B.} \end{cases}$$

Results of Experiment 7.2. They are displayed in Table II and Figures 3–4. The numbers in Table II show that the estimates in Case B are quite close to θ_{real}, and in Case A we see a slight underestimation of θ_{real}. Also the variances in Case A are a bit larger than in Case B, but they are still of the same order. So the knowledge of the value of ν does not improve the estimation of θ_{real} drastically. We interpret this result as follows: The information about the diffusion coefficient θ_{real} is completely encoded in the short time behavior of the paths of X, i.e. the quadratic variation

$$\langle X\rangle_t = \lim_{n\to\infty} \sum_{j=1}^{2^n} (X_{jt2^{-n}} - X_{(j-1)t2^{-n}})^2 = \theta_{\text{real}}^2 \int_0^t X_s^2 \, ds \tag{7.7}$$

allows to determine $\theta_{\text{real}} \geqslant 0$ uniquely from a single exact observation. The drift ν is irrelevant for $\langle X\rangle_t$. From this point of view we expect that a knowledge of

Table II. Example 2: The (sample) mean and standard deviation for the estimator $\hat{S}_T^1$ for the diffusion estimation (7.5). Here $\theta_{\text{real}} = 0.25$, and $\nu = \theta_{\text{real}}^2/2$. The drift ν is unknown in Case A and known in Case B.

	Mean	Std.dev.	M_X	M_θ	R
Case A	2,05e-1	1,30e-1	60	40	15
	2,04e-1	1,38e-1	90	60	15
	1,97e-1	1,27e-1	90	60	50
	1,94e-1	1,20e-1	120	90	50
Case B	2,69e-1	1,19e-1	60	40	15
	2,46e-1	1,03e-1	90	60	15
	2,42e-1	9,66e-2	90	60	50
	2,41e-1	9,91e-2	120	90	50

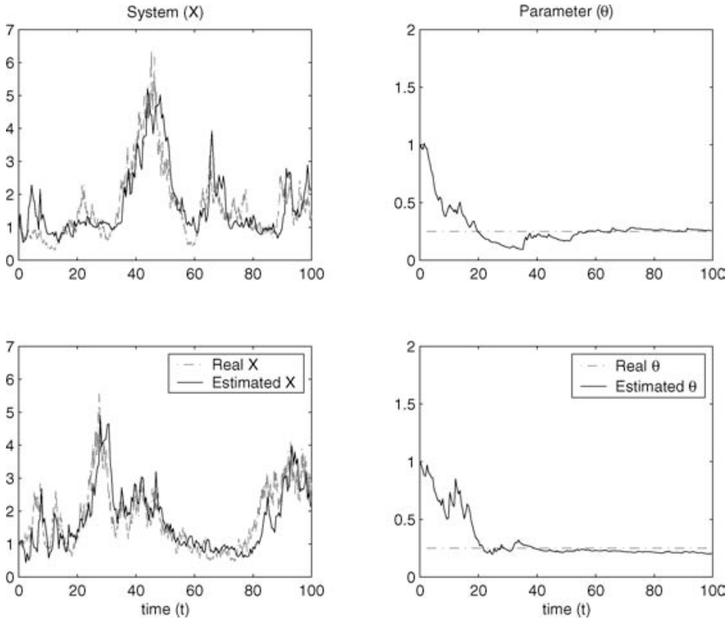

Figure 3. Example 2: The plots show two typical sample paths for the estimator of the system (X) and the parameter (θ). The first row corresponds to Case A and the second to Case B. In these plots we used $M_X = 120$, $M_\theta = 90$ and $R = 50$.

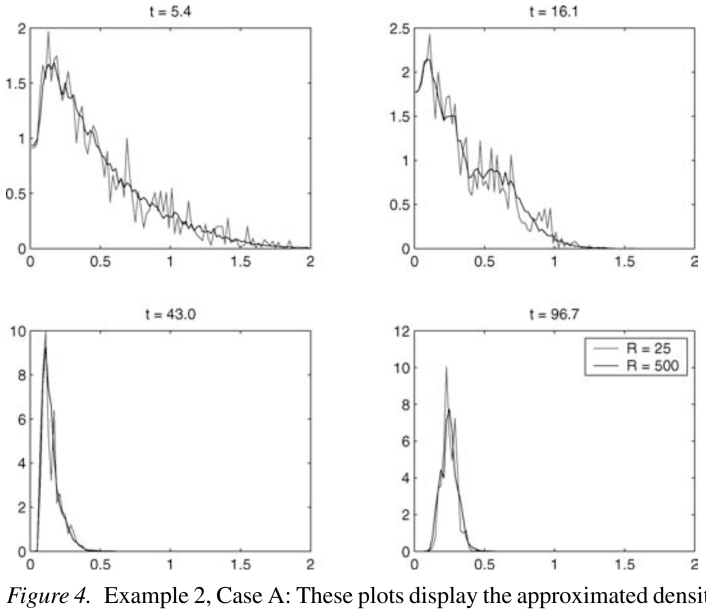

Figure 4. Example 2, Case A: These plots display the approximated density $\hat{p}(t_n, \theta_j, Y)$ given in (7.1) as a function of θ_j, at four different times. In each plot we use two different choices of R. These plots are based on a simulation for the case $\theta_{\text{real}} = 0.25$ with $M_X = 80$ and $M_\theta = 200$.

the drift ν adds only little information about the diffusion coefficient θ, if the measurement errors are sufficiently small.

EXAMPLE 3. We consider the process $X_t = X_0 + \sin(\theta W_t^{(1)})$ which satisfies the nonlinear equation

$$\mathrm{d}X_t = -\frac{1}{2}\theta^2 X_t \,\mathrm{d}t + \theta\sqrt{1 - X_t^2}\,\mathrm{d}W_t^{(1)}, \quad t \in [0, T], \tag{7.8}$$

with X_0 uniformly distributed in $[-1, 1]$, and independent of W. This process remains in the bounded region $[-2, 2]$ for all $t \geqslant 0$. We investigate two different choices for h in (5.1), denoted Case A and Case B. In both cases we assume h smooth and bounded, and we set $h(x) = x$ in Case A, and $h(x) = 10x$ in Case B, with $x \in [-2, 2]$. Using (2.7) we obtain the potential $c(t, x, \theta) = -\theta^2/2$, and the associated reversed system (4.12) becomes

$$\mathrm{d}\xi_s = -\frac{3}{2}\theta^2 \xi_s \,\mathrm{d}s + \theta\sqrt{1 - \xi_s^2}\,\mathrm{d}B_s, \quad s \in [0, t]. \tag{7.9}$$

This example is more difficult than Examples 1 and 2: We have to solve (7.9) numerically, and since X takes values only in $[-2, 2]$, the noise remains relatively big compared to $h(X_s)$, in particular for Case A. Thus, we do not expect to be able to estimate θ as good as in the two previous examples. Moreover, we expect that the estimator performs better in Case B, since for this case, the signal/noise ratio is better.

EXPERIMENT 7.3. For $\theta_{\text{real}} = 0.25$ and for Cases A and B, we simulate $M = 10$ independent sample paths of (X, Y), using $X_t = X_0 + \sin(\theta W_t^{(1)})$. For each simulation $Y^{(k)}$ we apply Algorithm 6.3 to estimate θ_{real} using different choices for the grid and for the Monte-Carlo parameter R, as given in Table III. We assume observation time-steps $\Delta t = T/N_T$, with $T = 200$, and $N_T = 8192$. Moreover, we choose $[A, B] \times [C, D] = [-2, 2] \times [0, 1]$ and the parameters (x_i, θ_j) are chosen from the rectangular grid in (7.4). We use the normalization interval $\delta = 2\Delta t$, and choose the initial joint distribution to be uniform on $[-1, 1] \times [0, 1]$. For the simulation of (7.9) we apply the Milstein scheme [25]:

$$y_{k+1} = y_k - \frac{3}{2}\theta_j^2 y_k \Delta s_k + \theta_j\sqrt{1 - y_k^2}\Delta B_k - \frac{1}{2}\theta_j^2 y_k(\Delta B_k^2 - \Delta s_k),$$

where y_k approximates $\xi_{s_k}^{t_n - t_{\hat{n}}, x_i, \theta_j}$, $\Delta s_k := s_{k+1} - s_k$, and where ΔB_k is the increment of a Wiener process on $[s_k, s_{k+1}]$.

Results of Experiment 7.3. They are presented in Table III and Figures 5–6. From Table III we see that Case B performs better than Case A, as expected. We also see that the method seems to under-estimate the actual value of θ_{real} in both cases, and this can hardly by explained by the sample variances of the estimator,

Table III. Example 3: The (sample) mean and standard deviation for the estimator $\hat{S}_T^1$ for the nonlinear case (7.8). Here $\theta_{\text{real}} = 0.25$, $h(x) = x$ in Case A, and $h(x) = 10x$ in Case B.

	Mean	Std.dev.	M_X	M_θ	R
Case A	1,53e-1	2,55e-2	30	30	10
	1,41e-1	1,80e-2	50	50	10
	1,38e-1	1,81e-2	50	50	50
	1,30e-1	1,69e-2	90	90	50
Case B	1,95e-1	1,92e-2	30	30	10
	1,97e-1	1,26e-2	50	50	10
	1,99e-1	7,52e-3	50	50	50
	2,03e-1	1,00e-2	90	90	50

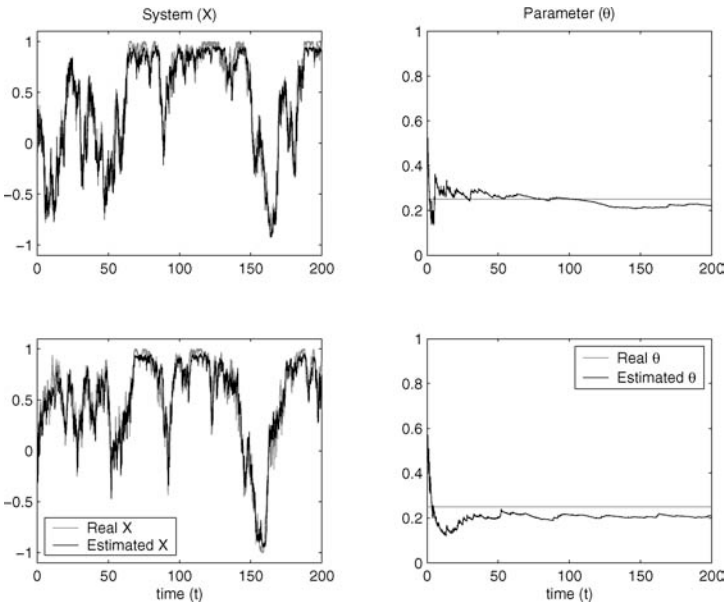

Figure 5. Example 3: The plot shows two typical sample paths for the estimator of the system (X) and the parameter (θ). The first row corresponds to Case B and the second to Case A. In all plots we used $M_X = 50$, $M_\theta = 50$ and $R = 50$.

which are relatively small. A typical time development of $\hat{S}_{t_n}^1$ is displayed together with the state estimator in Figure 5. Note that also the state estimator performs better in Case B compared to Case A. This is no surprise, since $h(x) = 10x$ (Case B) allows to "see" changes in X much easier than $h(x) = x$ (Case A).

Final remarks. (1) When θ_{real} is outside $[C, D]$ we observed that the $\hat{\theta}_t$-density systematically accumulates on the boundary next to θ_{real}. So an erroneous choice for $[C, D]$ can easily be detected from the density plot.

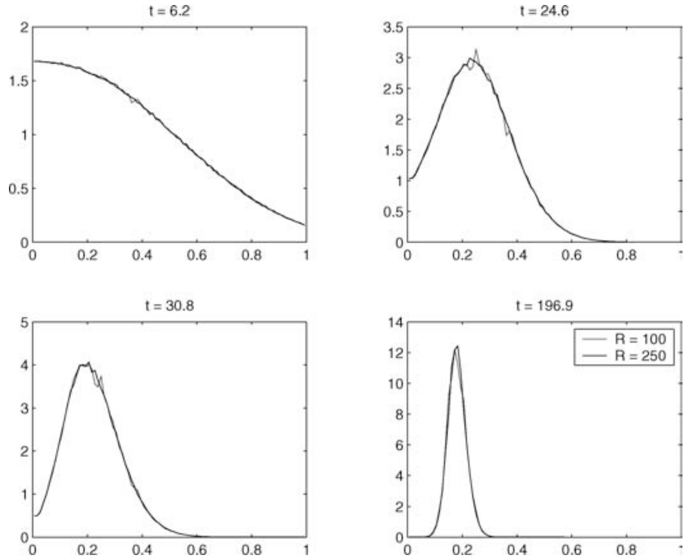

Figure 6. Example 3, case A: The approximated density $\hat{p}(t_n, \theta_j, Y)$ given in (7.1) is plotted as a function of θ_j at four different times, with two different choices of R. This plot is based on a simulation for the case $\theta_{\text{real}} = 0.25$ with $M_X = 50$ and $M_\theta = 200$.

(2) In each of our density plots we observe that θ_{real} is inside the density peaks, even in Example 3 where we observed some bias in the numerical estimator. In other words, the density plot gives in each case a reliable range for θ_{real}.

(3) As a conclusion from Examples 1 and 3 we can say that the estimates are close to θ_{real}, as long as the state estimates are close to the state. In other words, whenever the measurement noise is small we obtain good estimates, both for drift and for diffusion parameters.

(4) Example 3 shows that the estimator not always works as desired. More work has to be done to elaborate, under what conditions the Bayesian estimator gives satisfying results. From the mathematical viewpoint the basic statistical properties of an estimator (such as consistency, cf. [13]) have to be investigated. From the numerical viewpoint the algorithm has to be tested further and it needs to be improved, in particular for higher-dimensional parameter estimation problems. This is left for future research.

Acknowledgements

The authors acknowledge financial support from FCP, Portugal. The second author acknowledges support from the University of Mannheim, Germany, and from the Research Council of Norway under grant 131908/410. Discussions with Professor Jürgen Potthoff and Professor Ludwig Streit are gratefully acknowledged.

References

1. Basawa, I. V. and Prakasa Rao, B. L. S.: *Statistical Inference for Stochastic Processes*, Probability and Mathematical Statistics, Academic Press, London, 1980.
2. Benth, F. E., Deck, T., Potthoff, J. and Våge, G.: Explicit strong solutions of SPDE's with applications to non-linear filtering, *Acta Appl. Math.* **51**(2) (1998), 215–242.
3. Bhatt, A. G., Kallianpur, G. and Karandikar, R. L.: Uniqueness and robustness of solution of measure-valued equations of nonlinear filtering, *Ann. Probab.* **23** (1995), 1895–1938.
4. Bhatt, A. G., Kallianpur, G. and Karandikar, R. L.: Robustness of the nonlinear filter, *Stochastic Process. Appl.* **81**(2) (1999), 247–254.
5. Bismut, J. M. and Michel, D.: Diffusions conditionnelles I, Hypoellipticité partielle, *J. Funct. Anal.* **44** (1981), 174–211.
6. Bobrovsky, B.-Z. and Zakai, M.: Asymptotic a priori estimates for the error in the nonlinear filtering problem, *IEEE Trans. Inform. Theory* **28** (1982), 371–376.
7. Brigo, D., Hanzon, B. and Le Gland, F.: A differential geometric approach to nonlinear filtering: The projection filter, *IEEE Trans. Automat. Control* **43**(2) (1998), 247–252.
8. Bryc, W. and Smolenski, W.: On the stability problem for conditional expectations, *Statist. Probab. Lett.* **15** (1992), 41–46.
9. Bucy, R. C.: Nonlinear filtering theory, *IEEE Trans. Automat. Control* **10** (1965), 198–199.
10. Campillo, F. and Le Gland, F.: MLE for partially observed diffusions: Direct maximization vs. the EM algorithm, *Stochastic Process. Appl.* **33** (1989), 245–274.
11. Crisan, D., Gaines, J. and Lyons, T.: Convergence of a branching particle method to the solution of the Zakai equation, *SIAM J. Appl. Math.* **58**(5) (1998), 1568–1590.
12. Davis, M. H. A.: A pathwise solution of the equations of nonlinear filtering, *Teor. Veroyatnost. i Primenen.* **27** (1982), 160–167.
13. Deck, T.: Asymptotic properties of Bayes estimators for non-stationary linear SDEs with noisy observations, Preprint, 2004.
14. Di Masi, G. B. and Runggaldier, W. J.: Continuous-time approximations for the nonlinear filtering problem, *Appl. Math. Optim.* **7** (1981), 233–245.
15. Di Masi, G. B. and Runggaldier, W. J.: On measure transformations for combined filtering and parameter estimation in discrete time, *Systems Control Lett.* **2**(1) (1982), 57–62.
16. Fetter, H.: On the continuity of conditional expectations, *J. Math. Anal. Appl.* **61** (1997), 227–231.
17. Freidlin, M.: *Functional Integration and Partial Differential Equations*, Princeton University Press, 1985.
18. Germani, A., Jetto, L. and Piccioni, M.: Galerkin approximation for optimal linear filtering of infinite-dimensional linar systems, *SIAM J. Control Optim.* **26**(6) (1998), 1287–1305.
19. Ito, K.: Approximation of the Zakai equation for nonlinear filtering, *SIAM J. Control Optim.* **34**(2) (1996), 620–634.
20. James, M. R. and Le Gland, F.: Consistent parameter estimation for partially observed diffusions with small noise, *Appl. Math. Optim.* **32** (1995), 47–72.
21. Kallianpur, G.: *Stochastic Filtering Theory*, Springer, New York, 1980.
22. Kallianpur, G. and Selukar, R. S.: Parameter estimation in linear filtering, *J. Multivariate Anal.* **39**(2) (1991), 284–304.
23. Kallianpur, G. and Striebel, C.: Estimation of stochastic systems: Arbitrary system process with additive white noise observation errors, *Ann. Math. Statist.* **39** (1968), 785–801.
24. Kallianpur, G. and Striebel, C.: Stochastic differential equations occurring in the estimation of continuous parameter stochastic processes, *Theory Probab. Appl.* **14** (1969), 567–594.
25. Kloeden, P. E. and Platen, E.: *Numerical Solution of Stochastic Differential Equations*, Applications of Mathematics 23, Springer, Berlin, 1992.

26. Kunita, H.: Stochastic partial differential equations connected with nonlinear filtering, In: *Nonlinear Filtering and Stochastic Control (Cortona, 1981)*, Lecture Notes in Math. 972, Springer, 1982, pp. 100–169.
27. Kunita, H.: Stochastic differential equations and stochastic flows of diffeomorphisms, In: *École d'Éte de Probabilités de Saint-Flour XII–1982*, Lecture Notes in Math. 1097, Springer, 1984, pp. 143–303.
28. Kurtz, T. G. and Ocone, D. L.: Unique characterization of condtitional distributions in nonlinear filtering, *Ann. Probab.* **16** (1998), 80–107.
29. Kutoyants, Yu. A.: *Parameter Estimation for Stochastic Processes*, Heldermann Verlag, Berlin, 1984.
30. Le Gland, F.: Nonlinear filtering equation and a problem of parametric estimation, In: M. Hazewinkel and J. C. Williams (eds), *Stochastic Systems: The Mathematics of Filtering and Identification and Applications (Les Arcs, 1980)*, NATO Advanced Study Institute Series C: Math. Phys. Sci. 78, Reidel, 1981, pp. 613–620.
31. Lipster, R. S. and Shiryayev, N.: *Statistics of Random Processes I*, Springer, New York, 1997.
32. Lototsky, S., Mikulevicius, R. and Rozovskii, B. L.: Nonlinear filtering revisited: A spectral approach, *SIAM J. Control Optim.* **35**(2) (1997), 435–461.
33. Maybank, S. J.: Path integrals and finite dimensional filters, In: *Stochastic Partial Differential Equations (Edinburgh, 1994)*, London Math. Soc. Lecture Note Ser. 216, Cambridge University Press, 1995, pp. 209–229.
34. Michel, D.: Régularité des lois conditionnelles en théorie du filtrage non-linéaire et calcul des variations stochastique, *J. Funct. Anal.* **41** (1981), 8–36.
35. Millet, A. and Smolenski, W.: Small pertubations of Gaussian regressors, *Statist. Probab. Lett.* **24** (1995), 21–31.
36. Øksendal, B.: *Stochastic Differential Equations*, 3rd edn, Springer, 1992.
37. Pardoux, E.: Filtrage non lineaire et equations aux derivees partielles stochastiques associees, In: *École d'Éte de Probabilités de Saint-Flour XIX–1989*, Lecture Notes in Math. 1464, Springer, 1991, pp. 69–163.
38. Parthasarathy, K. R.: *Introduction to Probability and Measure*, Springer, New York, 1978.
39. Picard, J.: Asymptotic study of estimation problems with small observation noise, In: *Stochastic Modelling and Filtering (Rome, 1984)*, Lecture Notes in Control and Inference Sci. 91, Springer, 1987, pp. 131–146.
40. Prakasa Rao, B. L. S.: *Statistical Inference for Diffusion Type Processes*, Kendall's Library of Statistics 8, Oxford University Press, New York, 1999.
41. Stroock, D. W.: *An Introduction to the Theory of Large Deviations*, Springer, New York, 1984.
42. Theting, T. G.: Solving parabolic Wick-stochastic boundary value problems using a finite element method, *Stochastics Stochastics Rep.* **75** (2003), 49–77.
43. Wong, E.: *Stochastic Processes in Information and Dynamical Systems*, McGraw-Hill, New York, 1971.
44. Zakai, M.: On the optimal filtering of diffusion processes, *Z. Wahrsch. Verw. Gebiete* **11** (1969), 230–243.