Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Multivariate Analysis 97 (2006) 563–573
Journal of Multivariate Analysis
www.elsevier.com/locate/jmva

Asymptotic properties of Bayes estimators for Gaussian Itô-processes with noisy observations

T. Deck*, [1]

Fakultät für Mathematik und Informatik, Universität Mannheim, D-68131 Mannheim, A5, Germany

Received 9 March 2004
Available online 23 may 2005

Abstract

The estimation of a real parameter θ in a linear stochastic differential equation of the simple type $dX_t = \theta\beta(t)\,dt + \sigma(t)\,dB_t$ is investigated, based on noisy, time continuous observations of X_t. Sufficient conditions on the continuous functions β and σ are given such that the (conditionally normal) Bayes estimators of θ satisfy certain error bounds and are strongly consistent.

AMS 1991 subject classification: 62M20; 62F15

Keywords: Parameter estimation; Linear filtering theory

1. Introduction

A *Gaussian Itô-process* $(X_t)_{t\geqslant 0}$ is a continuous, real stochastic process of the form

$$X_t = X_0 + \int_0^t \beta(s)\,ds + \int_0^t \sigma(s)\,dB_s, \tag{1.1}$$

where β, σ are deterministic functions of time t, and X_0 is a Gaussian random variable, independent of the Brownian motion (BM) $(B_t)_{t\geqslant 0}$. We will interpret (1.1) as a very simple

* Fax: +49 621 181 2504.
E-mail address: deck@math.uni-mannheim.de.
[1] Financial support from FCP, Portugal, is gratefully acknowledged.

0047-259X/$ - see front matter
doi:10.1016/j.jmva.2005.04.003

stochastic differential equation (SDE) for X_t, and we assume that the drift coefficient β contains an unknown parameter $\theta \in \mathbb{R}$ as follows:

$$dX_t = \theta\beta(t)\,dt + \sigma(t)\,dB_t\,, \quad t \geqslant 0. \tag{1.2}$$

We suppose that the *continuous functions* β, σ are known, and that we observe the process X_t corrupted by another BM W_t (independent of the first one) as follows:

$$Y_t = \int_0^t X_s\,ds + W_t. \tag{1.3}$$

The problem to be treated in this work is to estimate θ in (1.2), based on one observation path of (1.3) up to time *t*. We study this problem from the Bayesian viewpoint, i.e. we model θ as a random variable (r.v.), denoted θ_0. We choose θ_0 normally distributed and independent of $\sigma(B_t, W_t, t \geqslant 0)$. Thus we can consider θ_0 also as a stochastic process satisfying the trivial dynamical equation $d\theta_t = 0$. We write the resulting two component system for (X_t, θ_t) in matrix form as follows:

$$\begin{pmatrix} dX_t \\ d\theta_t \end{pmatrix} = \begin{pmatrix} 0 & \beta(t) \\ 0 & 0 \end{pmatrix} \begin{pmatrix} X_t \\ \theta_t \end{pmatrix} dt + \begin{pmatrix} \sigma(t) \\ 0 \end{pmatrix} dB_t. \tag{1.4}$$

Similarly, observations (1.3) can be expressed in matrix form as

$$dY_t = (1 \;\; 0) \begin{pmatrix} X_t \\ \theta_t \end{pmatrix} dt + 1\,dW_t. \tag{1.5}$$

This reformulation of (1.2) and (1.3) is useful because the extended system (1.4) and (1.5) has the standard (state space) form investigated in linear filtering theory, cf. [3,10]. As usual in that theory we estimate θ_0 by the conditional expectation

$$\hat{\theta}_t := E[\theta_0 | Y_s, 0 \leqslant s \leqslant t], \tag{1.6}$$

i.e. the Bayes estimator w.r.t. quadratic loss; we also put $\hat{X}_t := E[X_t | Y_s, 0 \leqslant s \leqslant t]$.

Remark. (1) For general linear SDEs two kinds of drift estimators are intensively studied in the literature [1,2,9,10,13]: The Maximum Likelihood (cf. [6] for noisy observations) and Bayes estimators. But the latter differ from those considered here: Instead of (1.3) it is usually assumed that X_t is observed without measurement error, cf. [9, p. 85]. Also linear filtering theory has been applied to the estimation of drift coefficients θ in the measurement equation; two distinct cases are given in [9, p. 102; 12, Section 6]. It is somewhat surprising that (1.2) and (1.3) has apparently not been investigated from the Bayesian viewpoint before.

(2) The problem (1.2) and (1.3) can be viewed as a special "signal + noise" problem, because $X_t = X_0 + \theta \int_0^t \beta(s)\,ds + \int_0^t \sigma(s)\,dB_s$, so Y_t in (1.3) can be written as

$$Y_t = \theta \cdot \psi(t) + N_t,$$

where ψ is a known deterministic function, and N_t is a continuous, centered Gaussian process with known covariance. To estimate θ based on observations $(Y_s)_{0 \leqslant s \leqslant t}$ is a well-studied problem (see, e.g. [9, Chapter 2]), but the methods employed there are quite different from those in the present work.

In linear filtering theory it is well known that for given Gaussian initial conditions X_0 and θ_0 for (1.4) the error covariance matrix $S(t)$, defined by

$$\begin{pmatrix} S_{xx}(t) & S_{x\theta}(t) \\ S_{\theta x}(t) & S_{\theta\theta}(t) \end{pmatrix} := \begin{pmatrix} E[(X_t - \hat{X}_t)^2] & E[(X_t - \hat{X}_t)(\theta_0 - \hat{\theta}_t)] \\ E[(X_t - \hat{X}_t)(\theta_0 - \hat{\theta}_t)] & E[(\theta_0 - \hat{\theta}_t)^2] \end{pmatrix},$$

satisfies the matrix Riccati equation ($\dot{S}$ abbreviates dS/dt)

$$\dot{S}(t) = FS + SF^T - SG^T(DD^T)^{-1}GS + CC^T, \tag{1.7}$$

where F, G, D and C are the coefficient matrices determined by (1.4) and (1.5):

$$F(t) = \begin{pmatrix} 0 & \beta(t) \\ 0 & 0 \end{pmatrix}, \quad C(t) = \begin{pmatrix} \sigma(t) \\ 0 \end{pmatrix}, \quad G(t) = (1 \ \ 0), \quad D(t) = 1.$$

In this work we first study the estimation error $\theta_0 - \hat{\theta}_t$ as time goes to infinity, based on an asymptotic analysis of (1.7). We give sufficient conditions on the coefficient functions β and σ in (1.2) such that this error satisfies certain L^2-bounds. We finally show that strong consistency holds for the Bayes estimator (1.6), whenever the error converges to 0 in L^2-sense. Other properties, such as (asymptotic) normality and asymptotic unbiasedness, are straightforward consequences.

Remark. (1) It is well known that sequences of Bayes estimators are consistent under relatively weak conditions, but much less is known for continuous time Bayes estimators such as (1.6), cf. [11, p. 76] and references given there.

(2) The asymptotics of *state* estimators $\hat{X}_t$ is studied in [8]; the methods used there do not apply to our case, because our coefficient matrix F is degenerated, and we do not impose any ergodicity assumptions.

(3) This work is motivated by Deck and Theting [4], where Bayes estimators (including diffusion coefficients) for nonlinear SDEs are investigated. Asymptotic properties (as $t \to \infty$) are not studied in [4]. For the nonlinear case this is known to be a hard problem, cf. [7]. The present work investigates the most elementary case, i.e. (1.2) and (1.3). Already this case leads to a system of nonlinear ordinary differential equations (ODEs) whose asymptotic analysis is not quite simple.

2. Solution of the Riccati equation

Denote the components of the error matrix S by $a := S_{xx}$, $b := S_{x\theta} = S_{\theta x}$, and $c := S_{\theta\theta}$. Then (1.7) leads to the following system of equations for a, b and c:

$$\begin{aligned} \dot{a} &= 2\beta b - a^2 + \sigma^2, \\ \dot{b} &= \beta c - ab, \\ \dot{c} &= -b^2. \end{aligned} \tag{2.1}$$

Notice that (2.1) is a non-trivial system of nonlinear ODEs, with time-dependent coefficients $\sigma^2(t)$ and $\beta(t)$. It is known that such Riccati equations have unique solutions for all $t \in$

$\mathbb{R}_+ = [0, \infty)$. (For (2.1) this also follows from the proof of Proposition 3.1 given below.) The equation $\dot{c} = -b^2$ shows that the error variance $c(t) = E[(\theta_0 - \hat{\theta}_t)^2]$ is monotonically decreasing. This is quite clear because when t increases the conditional expectation $\hat{\theta}_t = E[\theta_0 | Y_s, 0 \leqslant s \leqslant t]$ projects onto increasing subspaces in L^2, so that the L^2-distance $\sqrt{c(t)}$ decreases. The point in the L^2-asymptotics of the error is, so to say, to find conditions such that this function in fact decreases to zero. This is not always the case, as the counter example given after Theorem 3.1 shows.

In the following we assume that the Gaussian initial conditions X_0 and θ_0 for (1.4) are independent and have non-vanishing variances. For system (2.1) this means that $a(0) > 0$, $c(0) > 0$, and by independence $b(0) = 0$. Thus $S(0)$ is a regular matrix, and by continuity the inverse matrix $S^{-1}(t)$ exists at least for small times. Our asymptotic analysis of (2.1) is based on the asymptotic behavior of $S^{-1}(t)$. Therefore we need to know that this matrix is regular for all $t \geqslant 0$. But this is always satisfied, without specific restrictions on β and σ:

Proposition 2.1. *Let $\beta, \sigma \in C(\mathbb{R}_+)$ and assume the initial conditions for* (2.1) *satisfy $a(0) > 0$, $b(0) = 0$ and $c(0) > 0$. Then the solution to* (2.1) *satisfies* $\det(S(t)) > 0$ *for all $t \geqslant 0$, and moreover*

$$S_{xx}(t) > 0, \quad S_{\theta\theta}(t) > 0 \quad \forall t \geqslant 0. \tag{2.2}$$

Proof. *Step* 1: As long as $\det(S(t)) > 0$ define

$$R(t) := S^{-1}(t) := \begin{pmatrix} u(t) & v(t) \\ v(t) & w(t) \end{pmatrix}. \tag{2.3}$$

$RS = I$ implies $\dot{R}S + R\dot{S} = 0$, thus $\dot{R} = -R\dot{S}R$. Multiplication of (1.7) with R from both sides shows that R satisfies $\dot{R} = -RF - F^T R + G^T(DD^T)^{-1}G - RCC^T R$. A simple computation now shows that the components of R satisfy

$$\begin{aligned} \dot{u} &= 1 - \sigma^2 u^2, \\ \dot{v} &= -\beta u - \sigma^2 uv, \\ \dot{w} &= -2\beta v - \sigma^2 v^2, \end{aligned} \tag{2.4}$$

subject to initial conditions $u(0) > 0$, $v(0) = 0$ and $w(0) > 0$. These equations hold as long as $\det(S(t)) > 0$. The first equation $\dot{u} = 1 - \sigma^2 u^2$ with initial condition $u(0) > 0$ has a unique solution on a maximal time interval $[0, T)$, with $T \in (0, \infty]$. Assume there is a $\bar{t} \in (0, T)$ such that $u(\bar{t}) = 0$. Then there is also a smallest time $t_0 > 0$ such that $u(t_0) = 0$. For $0 \leqslant t < t_0$ we then have $u(t) > 0$, and thus

$$\dot{u}(t_0) = \lim_{h \downarrow 0} \frac{u(t_0) - u(t_0 - h)}{h} \leqslant 0.$$

This contradicts $\dot{u}(t_0) = 1 - \sigma^2(t_0)u^2(t_0) = 1$. Thus $u(t) > 0$ on $[0, T)$. On the other hand $\dot{u}(t) = 1 - \sigma^2(t)u^2(t) \leqslant 1$, for all $t \in [0, T)$. So $u(t)$ is bounded from below by 0 and from above by $u(0) + t$. This shows that $u(t)$ cannot explode in finite time, thus $T = \infty$. This implies that system (2.4) has a unique solution on $\mathbb{R}_+$, because the second equation is (given u) a linear equation for v, which can be solved analytically on $\mathbb{R}_+$, and finally w simply follows by integration.

With $z := uw - v^2$ $(= \det(R))$ let us verify the estimate

$$0 < u(t) \leqslant \frac{u(0)}{z(0)} z(t) \quad \forall t \geqslant 0. \tag{2.5}$$

A simple computation based on (2.4) shows that u/z satisfies

$$\frac{d}{dt}\left(\frac{u}{z}\right) = \frac{\dot{u}z - u\dot{z}}{z^2} = \cdots = -\frac{v^2}{z^2} \leqslant 0.$$

Thus $\dot{u}/u \leqslant \dot{z}/z$, as long as $z(t) > 0$. This shows that $u(t) \leqslant \frac{u(0)}{z(0)} z(t)$, and $0 < u(t)$ for all t gives $z(t) > 0$ for all t. Thus (2.5) holds.

Step 2: Let $R(t)$ be the matrix solution of (2.4) constructed in Step 1. Since $\det(R(t)) > 0$ for all $t \geqslant 0$ we see that $S(t) := R^{-1}(t)$ is the unique solution to the initial value problem for (2.1), and our first claim $\det(S(t)) = 1/\det(R(t)) > 0$ for all t follows. Moreover, $S_{\theta\theta}(t) = u(t)/z(t)$ implies $S_{\theta\theta}(t) > 0$ for all t, by (2.5). Finally assume there exists t_0 such that $S_{xx}(t_0) = 0$. Then we also have $w(t_0) = S_{xx}(t_0)z(t_0) = 0$, so $z(t_0) = u(t_0)w(t_0) - v^2(t_0) \leqslant 0$. But this contradicts $z(t_0) > 0$, and thus $S_{xx}(t) > 0$ for all $t \geqslant 0$. □

Remark. (1) Estimates (2.2) show that there is no exact estimation for X_t and for θ_0 in finite time t. Of course this is not a surprise.

(2) Notice that the main point in the proof is the miraculous simplification which has occurred in (2.4), as compared with (2.1): The first equation in (2.4) already determines u uniquely, the second equation is then a linear equation for v, and finally w follows by integration! However, the equation for u has in general no analytic solution, which makes the asymptotic analysis still non-trivial. □

3. Asymptotic error analysis

For preparation let us first solve the equation for $u(t)$ in (2.4) when $\sigma(t)$ is equal to a constant $\sigma > 0$. In case $u(0) \neq 1/\sigma$ one finds

$$u(t) = \frac{1}{\sigma} \frac{Me^{2\sigma t} - 1}{Me^{2\sigma t} + 1} \qquad \text{with} \quad M = \frac{1 + \sigma u(0)}{1 - \sigma u(0)}.$$

In the other case, $u(0) = 1/\sigma$, the solution reads $u(t) = 1/\sigma$, for all $t \geqslant 0$. For each $\sigma > 0$ the solution u obviously satisfies

$$u(t) \to \frac{1}{\sigma} \quad \text{as} \quad t \to \infty. \tag{3.1}$$

The following result gives sufficient conditions such that $L^2 - \lim_{t\to\infty} \hat{\theta}_t = \theta_0$:

Theorem 3.1 (*Error bounds*). *Assume $\beta, \sigma \in C(\mathbb{R}_+)$, and there are constants $\beta_1, \beta_2, \sigma_1, \sigma_2, t_0$ such that the following estimates are satisfied*:

(a) $0 < \beta_1 \leqslant |\beta(t)| \leqslant \beta_2$ *for all* $t \geqslant t_0$.

(b) $0 < \sigma_1 \leqslant |\sigma(t)| \leqslant \sigma_2$ *for all* $t \geqslant t_0$.
(c) $2\beta_1\sigma_1^3 > \beta_2\sigma_2^3$.
Then there are constants $p, q > 0$ *such that* $\hat{\theta}_t$ *defined in* (1.6) *satisfies*

$$E[(\theta_0 - \hat{\theta}_t)^2] \leqslant \frac{p}{q+t} \quad \forall t \geqslant 0. \tag{3.2}$$

In particular $\hat{\theta}_t$ *converges in* L^2*-sense to* θ_0, *as* $t \to \infty$.

Proof. In view of $S_{\theta\theta} = u/z$ it suffices to show that u is bounded from above, and that $z \to \infty$ sufficiently fast, as $t \to \infty$. For $i = 1, 2$ denote by u_i the solution to $\dot{u}_i = 1 - \sigma_i^2 u_i^2$ subject to $u_i(t_0) = u(t_0)$. Since the estimates

$$1 - \sigma_2^2 u^2 \leqslant 1 - \sigma^2 u^2 \leqslant 1 - \sigma_1^2 u^2$$

hold for all $t \geqslant t_0$ the comparison theorem for ODEs [15] gives

$$u_2(t) \leqslant u(t) \leqslant u_1(t) \qquad \forall t \geqslant t_0.$$

In view of (3.1) this firstly implies the boundedness of u, and secondly for given $\delta \in (0, 1)$ allows to choose $t_1 \geqslant t_0$ such that

$$0 < \frac{1-\delta}{\sigma_2} \leqslant u(r) \leqslant \frac{1+\delta}{\sigma_1} \quad \forall r \geqslant t_1. \tag{3.3}$$

In view of (2.4) and $v(0) = 0$ the function $v(t)$ is given for $t \geqslant t_1$ by

$$\begin{aligned} v(t) &= -\int_0^t e^{-\int_s^t \sigma^2(r)u(r)\,dr} \beta(s)u(s)\,ds \\ &= -e^{-\int_0^t \sigma^2(r)u(r)\,dr} \int_0^{t_1} e^{\int_0^s \sigma^2(r)u(r)\,dr} \beta(s)u(s)\,ds \\ &\quad - \int_{t_1}^t e^{-\int_s^t \sigma^2(r)u(r)\,dr} \beta(s)u(s)\,ds. \end{aligned} \tag{3.4}$$

The second term in (3.4) can now be estimated by

$$\begin{aligned} \left| \int_{t_1}^t e^{-\int_s^t \sigma^2(r)u(r)\,dr} \beta(s)u(s)\,ds \right| &\leqslant \int_0^t e^{-(t-s)\sigma_1^2(1-\delta)/\sigma_2} \beta_2(1+\delta)/\sigma_1\,ds \\ &\leqslant \beta_2 \frac{\sigma_2(1+\delta)}{\sigma_1^3(1-\delta)}. \end{aligned} \tag{3.5}$$

The first term in (3.4) goes to 0 as $t \to \infty$, because $\sigma^2(r)u(r) \geqslant \sigma_1^2(1-\delta)/\sigma_2 > 0$ for all $t \geqslant t_1$. This combined with (3.5) shows that for each $\varepsilon > 0$ there exists $t(\varepsilon) > 0$ such that

$$\frac{\beta_1(\sigma_1 - \varepsilon)}{\sigma_2^3} \leqslant |v(t)| \leqslant \frac{\beta_2(\sigma_2 + \varepsilon)}{\sigma_1^3} \qquad \forall t \geqslant t(\varepsilon), \tag{3.6}$$

where the first estimate follows by similar arguments. Invoking now the equation $\dot{w} = -2\beta v - \sigma^2 v^2$ from (2.4) yields, for $t \geqslant t(\varepsilon)$:

$$\dot{w} = (2|\beta| - \sigma^2|v|)|v| \geqslant \left(2\beta_1 - \beta_2 \frac{\sigma_2^2(\sigma_2 + \varepsilon)}{\sigma_1^3}\right) \frac{\beta_1(\sigma_1 - \varepsilon)}{\sigma_2^3}.$$

By assumption (c) the right-hand side is > 0 for a sufficiently small $\varepsilon > 0$. This shows that $w(t)$ goes to infinity at least as a linear function. By (3.3) and (3.6) the same holds for $z = uw - v^2$. Thus we conclude that $S_{\theta\theta} = u/z$ satisfies (3.2). $\square$

Remark. (1) When $\sigma, \beta > 0$ are constants we can choose $\beta_1 = \beta_2 = \beta$ and $\sigma_1 = \sigma_2 = \sigma$. Then conditions (a)–(c) are satisfied, so the estimator is consistent. This shows that Theorem 3.1 allows for some variability in β and σ.

(2) The boundedness of β from below by a strictly positive constant cannot be relaxed in general: Consider $\beta(t) = e^{-ct}$ with some $c > 0$, and suppose condition (b) from Theorem 3.1 is satisfied. Then it is not hard to verify that

$$E[(\theta_0 - \hat{\theta}_t)^2] \not\to 0 \quad \text{as } t \to \infty.$$

(3) The boundedness of β from above is probably not necessary. Consider for example $\beta(t) = t^n$ with $n \in \mathbb{N}$, and suppose condition (b) from Theorem 3.1 is satisfied. Then it is not hard to verify that

$$E[(\theta_0 - \hat{\theta}_t)^2] \leqslant \frac{const.}{t^{2n+1}} \to 0.$$

So the stronger β increases, the faster the estimation error goes to zero. This is intuitively plausible.

(4) The assumptions in Theorem 3.1 are fairly strong. With more refined arguments one can show that (3.2) remains valid without condition (c). Moreover, the L^2-consistency of $\hat{\theta}_t$ already follows when $|\beta|$ is bounded from below by a function which decreases slower to zero than $1/\sqrt{t}$. Even cases with oscillating β can be treated. A more detailed analysis of drift estimators (which also includes some non-linear SDEs) will be given elsewhere.

(5) It is well known [3] that the Kalman–Bucy theory remains valid if one replaces the BM (B_t, W_t) in (1.2) and (1.3) by an arbitrary centered orthogonal increment process of the same covariance structure, and simultaneously replaces (1.6) by the best linear L^2-estimator. Thus Theorem 3.1 remains valid under this replacement.

Theorem 3.1 gives conditions such that $\theta_0 - \hat{\theta}_t$ goes to 0 in L^2-sense. If the parameter θ would be a genuine Gaussian r.v. (so $\theta \equiv \theta_0$) then we would have a clear statistical interpretation for this convergence: First pick θ_0 at random, then let the dynamic system (1.2) run up to time t and simultaneously observe Y by (1.3), and finally compute $\hat{\theta}_t$ (by (3.9) given below). The quantity $(\theta_0(\omega) - \hat{\theta}_t(\omega))^2$ would then be the squared estimation error, for one particular experiment ω, and its statistical mean over many such experiments would go to 0 as $t \to \infty$. But since θ is a fixed parameter in our model, the statistical mean over different values of $\theta_0(\omega)$ has no experimental meaning (we can only "pick" $\theta_0(\omega) = \theta$).

The true estimation error is thus given by $\theta - \hat{\theta}_t$, not $\theta_0 - \hat{\theta}_t$. It is therefore desirable that the estimator $\hat{\theta}_t$ converges to θ_0 for "all fixed values $\vartheta = \theta_0$", almost surely. To establish such an assertion we work with a product space

$$(\mathbb{R} \times \Omega, \mathcal{B}(\mathbb{R}) \otimes \mathcal{F}, \mu \otimes P),$$

where μ denotes the law of θ_0, and $(\Omega, \mathcal{F}, P)$ is the underlying probability space for the BM $(B_t, W_t)_{t \geqslant 0}$. This space is most appropriate because one can make P-a.s. statements for fixed $\vartheta \in \mathbb{R}$. Notice that in this representation we have $\theta_0(\vartheta, \omega) = \vartheta$, for all $(\vartheta, \omega) \in \mathbb{R} \times \Omega$. The following consistency result (which applies in particular to the context of Theorem 3.1) assumes this underlying probability space:

Theorem 3.2 (*Strong consistency*). *Assume $\hat{\theta}_t$ defined by* (1.6) *converges to θ_0 in $L^2(\mu \otimes P)$. Then there is a continuous version of the process $(\hat{\theta}_t)_{t \geqslant 0}$, such that for all $\vartheta \in \mathbb{R}$ this version satisfies*

$$\hat{\theta}_t(\vartheta, \cdot) \to \vartheta, \qquad P\text{-a.s., as } t \to \infty. \tag{3.7}$$

Moreover, for all $\vartheta \in \mathbb{R}$ the random variables $\hat{\theta}_t(\vartheta, \cdot)$ are normally distributed, and convergence (3.7) *also holds in $L^2(P)$.*

Proof. *Step* 1: We first show that (3.7) holds for all $\vartheta \in N^c$, where $\mu(N) = 0$. The Kalman–Bucy filter equations for system (1.4) are given by (cf. [3, Section 4.4])

$$\begin{pmatrix} d\hat{X}_t \\ d\hat{\theta}_t \end{pmatrix} = \begin{pmatrix} -S_{xx}(t) & \beta(t) \\ -S_{\theta x}(t) & 0 \end{pmatrix} \begin{pmatrix} \hat{X}_t \\ \hat{\theta}_t \end{pmatrix} dt + \begin{pmatrix} S_{xx}(t) \\ S_{\theta x}(t) \end{pmatrix} (X_t\, dt + dW_t), \tag{3.8}$$

subject to the initial conditions $\hat{X}_0 = E[X_0]$ and $\hat{\theta}_0 = E[\theta_0]$. If we denote by $\Phi(t, s)$ the matrix fundamental solution of the deterministic linear system

$$\begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \end{pmatrix} = \begin{pmatrix} -S_{xx}(t) & \beta(t) \\ -S_{\theta x}(t) & 0 \end{pmatrix} \begin{pmatrix} x(t) \\ y(t) \end{pmatrix}$$

then the solution to (3.8) is given by

$$\begin{pmatrix} \hat{X}_t \\ \hat{\theta}_t \end{pmatrix} = \Phi(t, 0) \cdot \begin{pmatrix} E[X_0] \\ E[\theta_0] \end{pmatrix} + \int_0^t \Phi(t, s) \cdot \begin{pmatrix} S_{xx}(s) \\ S_{\theta x}(s) \end{pmatrix} (X_s\, ds + dW_s). \tag{3.9}$$

By Kalman–Bucy theory $\hat{\theta}_t$ given in (1.6) coincides P-a.s. with the second component of this solution, which clearly defines a continuous version of (1.6). But (1.6) also shows that $(\hat{\theta}_t)_{t \geqslant 0}$ is a uniformly integrable martingale, so the martingale convergence theorem (and the supposed L^2-convergence) implies

$$\lim_{\mathbb{Q} \ni t \to \infty} \hat{\theta}_t = \theta_0, \qquad \mu \otimes P\text{-almost surely}.$$

By path continuity we can dispense with the restriction to $\mathbb{Q}$. So there is a set $M \subset \mathbb{R} \times \Omega$ of full measure 1, such that $\hat{\theta}_t(\vartheta, \omega) \to \theta_0(\vartheta, \omega) = \vartheta$, for all $(\vartheta, \omega) \in M$. An application of

Fubini's theorem to the indicator function 1_M shows that the set $M_\vartheta := \{\omega \in \Omega \mid (\vartheta, \omega) \in M\} \in \mathcal{F}$ has P-measure 1 for all $\vartheta \in N^c$, with $\mu(N) = 0$. For each $\vartheta \in N^c$ we have $(\vartheta, M_\vartheta) \subset M$, and thus

$$\hat{\theta}_t(\vartheta, \omega) \to \vartheta \qquad \forall \omega \in M_\vartheta \quad \forall \vartheta \in N^c.$$

Step 2: Integration of (1.4) gives $X_s = X_0 + \theta_0 \int_0^s \beta(u)\, du + \int_0^s \sigma(u)\, dB_u$. Putting this into (3.9) shows that $\hat{\theta}_t$ can be expressed as

$$\hat{\theta}_t(\vartheta, \omega) = \vartheta \cdot f(t) + Z_t(\omega), \tag{3.10}$$

where *f* is a deterministic continuous function and (Z_t) is a continuous Gaussian process on $(\Omega, \mathcal{F}, P)$. This shows that the r.v.s $\hat{\theta}_t(\vartheta, \cdot)$ are Gaussian on $(\Omega, \mathcal{F}, P)$. By Step 1 we can also pick $\vartheta_1, \vartheta_2 \in N^c$ such that $\vartheta_1 \neq \vartheta_2$. For each $\omega \in M_{\vartheta_1} \cap M_{\vartheta_2}$ (which has P-measure 1) we have $\hat{\theta}_t(\vartheta_i, \omega) \to \vartheta_i$. From (3.10) we obtain

$$\hat{\theta}_t(\vartheta_1, \omega) - \hat{\theta}_t(\vartheta_2, \omega) = (\vartheta_1 - \vartheta_2) f(t) \to \vartheta_1 - \vartheta_2 \quad \text{as } t \to \infty.$$

Thus we conclude

$$f(t) \to 1 \quad \text{as } t \to \infty. \tag{3.11}$$

Now $\hat{\theta}_t(\vartheta_1, \omega) \to \vartheta_1$ for all $\omega \in M_{\vartheta_1}$ implies

$$Z_t(\omega) \to 0 \qquad \forall \omega \in M_{\vartheta_1}. \tag{3.12}$$

So (3.11) and (3.12) show that $\hat{\theta}_t(\vartheta, \omega)$, given by (3.10), in fact converges to ϑ, for all $\vartheta \in \mathbb{R}$ and all $\omega \in M_{\vartheta_1}$.

Step 3: Since $E[\hat{\theta}_t] = E[\theta_0]$ Eq. (3.10) gives $E[\theta_0] = E[\theta_0] f(t) + E[Z_t]$. This and (3.11) implies $E[Z_t] = E[\theta_0](1 - f(t)) \to 0$ as $t \to \infty$. Using this and the independence of θ_0 and Z_t one easily verifies

$$E[(\hat{\theta}_t - \theta_0)^2] = (1 - f(t))^2 (E[\theta_0^2] - 2E[\theta_0]^2) + E[Z_t^2].$$

By assumption this quantity goes to zero, so with (3.11) we can conclude that $E[Z_t^2] \to 0$, as $t \to \infty$. This implies the last assertion. □

Remark. (1) By the factorization lemma we can write $\hat{\theta}_t$ as a function of the data (1.3), i.e. there is a measurable function $\hat{S}_t$ on the space of continuous paths $C(\mathbb{R}_+)$ (equipped with its standard σ-algebra $\mathcal{B}$) such that $\hat{\theta}_t(\vartheta, \omega) = \hat{S}_t(Y_\cdot(\vartheta, \omega))$, where $Y_\cdot(\vartheta, \omega)$ denotes the continuous path $(Y_s(\vartheta, \omega))_{s \geqslant 0}$. In view of (1.4) a continuous process (Y_t^ϑ) is defined on $(\Omega, \mathcal{F}, P)$ for each fixed ϑ by $Y_t^\vartheta(\omega) := Y_t(\vartheta, \omega)$, and thus the probability measure $P_\vartheta := P_{Y^\vartheta}$ (the law of Y^ϑ) is induced on the sample space $(C(\mathbb{R}_+), \mathcal{B})$. Now (3.7) can be stated as

$$\hat{S}_t(Y) \to \vartheta, \qquad P_\vartheta\text{-}a.s., \ \text{as } t \to \infty,$$

where now Y is the identity map on $C(\mathbb{R}_+)$. This is the more conventional form of strong consistency for the statistical model $(C(\mathbb{R}_+), \mathcal{B}, \{P_\vartheta, \vartheta \in \mathbb{R}\})$.

(2) It is well known that $\hat{\theta}_t$ is a Gaussian r.v. on the space $(\mathbb{R}\times\Omega, \mathcal{B}(\mathbb{R})\otimes\mathcal{F}, \mu\otimes P)$, but it appears not to be obvious that for fixed ϑ also $\hat{\theta}_t(\vartheta, \cdot)$ must be Gaussian on $(\Omega, \mathcal{F}, P)$. This property trivially implies that $\hat{\theta}_t$ is an *asymptotically normal estimator*. But it is not to be expected that $\sqrt{t}(\hat{\theta}_t(\vartheta, \cdot) - \vartheta)$ converges to a normal distribution, because our process (1.2) is non-stationary. Also notice that the L^2-convergence (3.7) implies that $\hat{\theta}_t$ is *asymptotically unbiased*. (It is well known that Bayes estimators for *finite* time are always biased, except in trivial cases.)

(3) Theorem 3.2 may be summarized as "$L^2(\mu \otimes P)$-*consistency implies strong consistency*" (without additional conditions on β and σ). Our proof of "μ-a.s. strong consistency" (Step 1) is based on martingale convergence. In the context of time discrete martingales this argument was introduced by Doob [5]. Extensions (again for sequences) were given by Schwartz [14].

(4) The proof of Step 1 in Theorem 3.2 only requires (besides L^2-convergence) that the martingale $(\hat{\theta}_t)_{t\geqslant 0}$ has a continuous version. This property also holds in the context of nonlinear filtering theory [4], so strong consistency (up to a set of μ-measure zero) also holds in that context.

(5) The proof of Theorem 3.2 requires that $\hat{\theta}_t$ is a martingale, and thus does not generalize without modifications to arbitrary orthogonal increment processes (as it was the case with Theorem 3.1).

Acknowledgments

It is a pleasure to thank Thomas Theting for fruitful discussions about some aspects of this work.

References

[1] M. Arató, Linear Stochastic Systems with Constant Coefficients, Lecture Notes in Control and Information Sciences, vol. 45, Springer, Berlin, 1982.
[2] I.V. Basawa, B.L.S. Prakasa Rao, Statistical Inference for Stochastic Processes, Academic Press, London, 1980.
[3] M.H.A. Davis, Linear Estimation and Stochastic Control, Chapman & Hall, London, 1977.
[4] T. Deck, T. Theting, Robust parameter estimation for stochastic differential equations, Acta Appl. Math. 84 (2004) 279–314.
[5] J.L. Doob, Applications of the theory of martingales, Colloque Internat. CNRS, 1949, 22–28.
[6] G. Kallianpur, R.S. Selukar, Parameter estimation in linear filtering, J. Multivariate Anal. 39 (2) (1991) 284–304.
[7] P.S. Krishnaprasad, S. Marcus, M. Hazewinkel, Current algebras and the identification problem, Stochastics 11 (1–2) (1983) 65–101.
[8] H. Kunita, Asymptotic behavior of the nonlinear filtering errors of Markov processes, J. Multivariate Anal. 1 (1971) 365–393.
[9] Yu A. Kutoyants, Parameter Estimation for Stochastic Processes, Heldermann Verlag, Berlin, 1984.
[10] R.S. Lipster, N. Shiryayev, Statistics of Random Processes I and II, second ed., Springer, New York, 2001.
[11] A. O'Hagan, Bayesian Inference, Kendalls Advanced Theory of Statistics, vol. 2B, Arnold, New York, 1999.

[12] B. Øksendal, Stochastic Differential Equations, Springer, Berlin, 1998.
[13] B.L.S. Prakasa Rao, Statistical Inference for Diffusion Type Processes, Kendall's Library of Statistics, vol. 8, Arnold, London, 1999.
[14] L. Schwartz, On Bayes procedures, Z. Wahrsch. Theo. 4 (1965) 10–26.
[15] W. Walter, Gewöhnliche Differentialgleichungen, Springer, Heidelberg, 1972.