Robustness of Zakai's equation via Feynman-Kac representations

Rami Atar
Frederi Viens
Ofer Zeitouni

ABSTRACT: We propose to study the sensitivity of the optimal filter to its initialization, by looking at the distance between two differently initialized filtering processes in terms of the ratio between two simple Feynman-Kac integrals in the product space. We illustrate, by considering two simple examples, how this approach may be employed to study the asymptotic decay rate, as the difference between the growth rates of the two integrals. We apply asymptotic methods, such as large deviations, to estimate these growth rates. The examples we consider are the linear case, where we recover known results, and a case where the drift term in the state process is nonlinear. In both cases, only the small noise regime and only one-dimensional diffusions are studied.

KEYWORDS: Non-linear filtering; Feynman-Kac; Large deviations.

1 Introduction

Zakai's equation of nonlinear filtering theory represents the solution of the filtering problem consisting of evaluating the conditional law of a Markov process observed in white Gaussian noise. In this paper we study the stability of Zakai's equation with respect to perturbations in its initial conditions. It is known since the work of Kunita (1971) that under mild conditions, the conditional law, viewed as a random process taking values in the space of probability measures, is stationary when appropriately initialized. Stettner (1989) shows that whenever the state process is a Feller Markov process converging in law to its unique invariant measure, so is its conditional law. Actually, c.f. Stettner (1991), the joint law of the state and its filtering process is Markovian even if the filter is wrongly initialized. It thus seems natural to investigate the rate of convergence and to study the sensitivity of the optimal filter to its initialization with the wrong initial measure. This issue is also highly relevant for numerical and practical computation of the optimal filter or its approximations, for almost never does one have

access to the true initial distribution.
Several approaches exist to analyze this exponential sensitivity, taking full advantage of the linear structure of Zakai's equation. In a recent article Ocone and Pardoux (1996) have studied L^p type of convergence, and showed that the nonlinear filter initialized at the wrong initial condition converges (in an L^p sense) to the nonlinear filter initialized at the correct initial condition. In particular cases (most notably, the Kalman filter), this convergence is exponential (see also Ocone (1997a) for a study of the Beneš case from a different point of view). In general, however, no rates of convergence are given by this approach.
Another approach, which does yield exponential rate of convergence, extends earlier Lyapunov exponent techniques suitable for the finite state space case as in Atar and Zeitouni (1997a). It is based on evaluating the rate of contraction of solutions of Zakai's equation in the projective Hilbert metric, using the Birkhoff coefficient associated with the kernel of the solution, see Atar and Zeitouni (1997b) for a development of this idea, and more recent work Budhiraja and Ocone (1997), Borkar, Mitter and Tatikonda (1997). This technique can be applied rather well in the case of compact state space, and yields results which are usually not tight when the noise level in the observation is weak. A different procedure, which as a by-product yields contraction in the Hilbert projective metric by controlling logarithmic derivatives of Zakai's kernel, is announced in Da Prato, Fuhrman and Malliavin (1995). An approach based on studying the relative entropy is reported in Clark, Ocone and Coumarbatch (1997) and in Ocone (1997b). Model robustness over the infinite time interval, and the relation of this problem to the sensitivity to initial conditions is dealt with in Budhiraja and Kushner (1997a), Budhiraja and Kushner (1997b).
The results based on Hilbert Projective metric are restricted to the compact state space case, and are usually not tight in the limit of low observation noise, for reasons described in Atar and Zeitouni (1997b). Some exceptional cases where contraction results for one dimensional, $\mathbb{R}$-valued diffusions exist and are tight are described in Atar (1997). Our goal in this paper is to suggest a different point of view, looking at the contraction in the space of positive measures as a ratio of two expectations, for which a simple Feynman-Kac representation can be achieved. Asymptotic methods, such as large deviations, can be then applied to the estimation of the growth rate of the latter Feynman-Kac integrals. We content ourselves here with presenting the idea and analyzing a simple one dimensional Gaussian diffusions, for which results are available in more generality by different methods (see, e.g., Ocone and Pardoux (1996)). We also present some immediate consequences for a class of one dimensional nonlinear diffusions. The case of general diffusion processes, even in one dimension, requires additional work and ideas and at present is not resolved.

2 A Feynman-Kac representation for the decay rate

For any measurable space $(\Omega, \mathcal{F})$, let $M(\Omega, \mathcal{F})$ and $M_1(\Omega, \mathcal{F})$ denote the spaces of finite signed measures on $(\Omega, \mathcal{F})$ and of probability measures on $(\Omega, \mathcal{F})$, respectively. For $p \in M(\Omega, \mathcal{F})$ and f measurable on $(\Omega, \mathcal{F})$ let $p(f) := \int f dp$. Define on $M(\Omega, \mathcal{F})$ the norm $\|\cdot\|$ compatible with the variation distance i.e.,

$$\|p\| = \sup \{p(f) : \ f \text{ is measurable on } (\Omega, \mathcal{F}),\, |f| \leq 1\}, \quad p \in M(\Omega, \mathcal{F}).$$

One fixed notation, namely $\|\cdot\|$, will be used to denote the above norm for measures on different measurable spaces. Next, let $\mathcal{F} \otimes \mathcal{F}$ denote the product σ-field, and for $p, q \in M(\Omega, \mathcal{F})$, define the exterior product $p \wedge q \in M(\Omega \times \Omega, \mathcal{F} \otimes \mathcal{F})$ by

$$p \wedge q = \frac{1}{2}\,(p \times q - q \times p).$$

We then have

Lemma 1 *Let $(\Omega, \mathcal{F})$ be a measurable space and let $p, q \in M_1(\Omega, \mathcal{F})$. Then*

$$\|p \wedge q\| \leq \|p - q\| \leq 2\|p \wedge q\|. \tag{1.1}$$

Proof: Note that for $r, s \in M(\Omega, \mathcal{F})$ one has $\|r \times s\| = \|r\|\|s\|$. Indeed, let f be measurable and such that $|f| \leq 1$, then by Jordan's decomposition for a signed measure and Fubini's theorem,

$$(r \times s)(f) = \int \left[\int f(x,y) r(dx)\right] s(dy) =: \int g(y) s(dy).$$

But $|g(y)| \leq \|r\|$, so $(r \times s)(f) \leq \|r\|\|s\|$. The reverse inequality is trivial. The leftmost inequality in (1.1) follows from

$$2(p \wedge q) = (p - q) \times q - q \times (p - q),$$

implying that

$$2\|p \wedge q\| \leq \|(p - q) \times q\| + \|q \times (p - q)\| = 2\|p - q\|\,.$$

The rightmost inequality in (1.1) follows from

$$p(f) - q(f) = (p \times q)(f \times \mathbf{1}) - (q \times p)(f \times \mathbf{1}) = 2(p \wedge q)(f \times 1).$$

□

Let $(S, \mathcal{S})$ be a Polish space, and let Ω denote the set of RCLL paths $\omega_t,\ t \geq 0$ on S. Consider a strong Markov process, $x_t, t \geq 0$, with paths in Ω, possessing the Feller property. Let $g : S \to \mathbb{R}^d$ be measurable and define

$$y_t = \int_0^t g(x_s)ds + \sigma d\nu_t,$$

where $\nu_t, t \geq 0$ is a standard Brownian motion on $\mathbb{R}^d$, independent of x_t. In the context of filtering theory, x_t is referred to as the state process, and y_t as the observation process.

Let b_t, b'_t, $t \geq 0$ be two processes on $(S, \mathcal{S})$ with the same transition law as x_t, but possibly different initial laws $p, p' \in M_1(S, \mathcal{S})$, respectively, and which are independent of x_t, ν_t and of each other. Let E_b ($E_{b'}$, $E_{b,b'}$) denote expectation w.r.t. b (resp. b', (b, b')) alone.

Let P be the measure induced by (x_t, y_t), $t \geq 0$, and denote by $\mathcal{Y}_t$ the sigma-field generated by $\{y_s, 0 \leq s \leq t\}$. We assume that

$$E \int_0^t |g(x_s)|^2 ds < \infty, \quad t \geq 0. \tag{1.2}$$

Now let

$$\Lambda_t = \exp\left(\frac{1}{\sigma^2}\int_0^t (g(b_s), dy_s) - \frac{1}{2\sigma^2}\int_0^t |g(b_s)|^2 ds\right), \quad t \geq 0,$$

and define the measure valued processes ρ_t and p_t by

$$\rho_t(\phi) = E_b\left[\phi(b_t)\Lambda_t\right], \quad t \geq 0, \phi \in \mathcal{C}_b(\mathbb{R}),$$
$$p_t(\phi) = \rho_t(\phi)/\rho_t(\mathbf{1}). \tag{1.3}$$

Let also Λ'_t, ρ'_t and p'_t be the processes defined as above, with b replaced by b'. Then it is well known that in case where p equals the initial law of x_t, one has P-a.s. that for all $t \geq 0$, p_t equals the conditional law of x_t given $\mathcal{Y}_t$ under P. In this work however, we are interested mainly in the case where p is arbitrary. p_t may then be interpreted as the filtering process with perturbed initial condition p. In particular, we shall look at the decay rate of the distance between differently perturbed filtering processes:

$$\limsup_{t\to\infty} \frac{1}{t} \log \|p_t - p'_t\|.$$

In view of Lemma 1 and equation (1.3), the decay rate above may be studied in terms of $\|\rho_t \wedge \rho'_t\|$ and $\|\rho_t \times \rho'_t\|$, namely one has

$$\limsup_{t\to\infty} \frac{1}{t} \log \|p_t - p'_t\| = \limsup_{t\to\infty} \frac{1}{t} \log \frac{\|\rho_t \wedge \rho'_t\|}{\|\rho_t \times \rho'_t\|},$$

and a similar statement for the lim inf. While by definition the denominator can be written as:

$$\|\rho_t \times \rho'_t\| = E_{b,b'}\Lambda_t\Lambda'_t,$$

we also have the following representation for the numerator.

Lemma 2 *Let V_t denote the event*

$$b_s \neq b'_s, \quad all\ s \in [0,t].$$

Then

$$(\rho_t \wedge \rho'_t)(A) = \frac{1}{2}\, E_{b,b'}\, \Lambda_t \Lambda'_t \mathbf{1}_{V_t} \left[\mathbf{1}_{(b_t,b'_t)\in A} - \mathbf{1}_{(b'_t,b_t)\in A}\right], \quad A \in \mathcal{S}\otimes\mathcal{S},$$

and hence

$$\|\rho_t \wedge \rho'_t\| \leq E_{b,b'}\, \Lambda_t \Lambda'_t \mathbf{1}_{V_t}\,, \tag{1.4}$$

implying that

$$\limsup_{t\to\infty} \frac{1}{t} \log \|p_t - p'_t\| \leq \limsup_{t\to\infty} \frac{1}{t} \log \left(\frac{E_{b,b'}\, \Lambda_t \Lambda'_t \mathbf{1}_{V_t}}{E_{b,b'}\, \Lambda_t \Lambda'_t}\right). \tag{1.5}$$

Proof: By definition we have

$$(\rho_t \wedge \rho'_t)(A) = \frac{1}{2}\, E_{b,b'}\, \Lambda_t \Lambda'_t \left[\mathbf{1}_{V_t} + \mathbf{1}_{V_t^c}\right] \left[\mathbf{1}_{(b_t,b'_t)\in A} - \mathbf{1}_{(b'_t,b_t)\in A}\right].$$

Using strong Markovity of the process (b_t, b'_t), a standard argument leads to

$$E_{b,b'}\, \Lambda_t \Lambda'_t \mathbf{1}_{V_t^c} \left[\mathbf{1}_{(b_t,b'_t)\in A} - \mathbf{1}_{(b'_t,b_t)\in A}\right] = 0, \tag{1.6}$$

and the result follows. Indeed, if we let $\{\mathcal{F}_t\}$ denote the filtration generated by $\{(b_t, b'_t)\}$, and

$$\tau = \inf\{s \geq 0 : b_s = b'_s\},$$

with $\inf \emptyset = \infty$, then τ is a stopping time on $\{\mathcal{F}_t\}$. We let $\theta_s,\ s \geq 0$ denote the family of shift operators $\theta_s : \Omega \times \Omega \to \Omega \times \Omega$. Namely, we have

$$(b_{s+t}, b'_{s+t})(\omega) = (b_t, b'_t)(\theta_s \omega), \quad \omega \in \Omega, \quad s,t \geq 0.$$

The strong Markov property then states that for any bounded measurable H,

$$E_{b,b'}\left[H_\tau \circ \theta_\tau \,|\, \mathcal{F}_\tau\right] = \phi(b_\tau, b'_\tau, \tau) \quad \text{on } \{\tau < \infty\}$$

where $\phi(x,x',u) = E_{b,b'}\left[H_u |\, b_0 = x, b'_0 = x'\right]$. Let us fix $t > 0$ and $A \in \mathcal{S}\otimes\mathcal{S}$ and for $n > 0$ define

$$H_s^n(b,b') = \\ \min\left\{n,\ \Lambda_{t-s}\Lambda'_{t-s}\right\} \mathbf{1}_{V_{t-s}^c} \left[\mathbf{1}_{(b(t-s),b'(t-s))\in A} - \mathbf{1}_{(b'(t-s),b(t-s))\in A}\right].$$

Note that on $\{\tau < \infty\}$ we have $H_\tau^n \circ \theta_\tau = H_0^n$. Moreover, for $x = x'$ and any u one has $\phi(x,x',u) = 0$. This proves that on $\{\tau < \infty\}$, $E_{b,b'}\left[H_0^n | \mathcal{F}_\tau\right] = 0$, while on $\{\tau = \infty\}$ this fact is trivial. Hence $E_{b,b'} H_0^n = 0$, and since assumption (1.2) and the independence of $\{x_t\}$ and $\{\nu_t\}$ imply that $\Lambda_t \Lambda'_t$ is $E_{b,b'}$-integrable, we obtain (1.6). □

In many cases the bound (1.4) is useless, in particular if $S = \mathbb{R}^d$, $d \geq 3$, and $\{x_t\}$ a standard Brownian motion, where V_t occurs almost surely for every t (given $P_{b,b'}(V_0) = 1$). However it is sharp in some other situations. For example, it holds with equality in the case $S = \mathbb{R}$ and $\{x_t\}$ a diffusion process, provided that $P_{b,b'}(b_0 > b'_0) = 1$. In high dimension, one needs to replace the coupling time τ by a more general coupling time, at which time the joint law of (b_t, b'_t) is exchangeable. For example, if $S = \mathbb{R}^d$ and the components of $b.$ are independent, one may take as coupling time the maximum of the collision times for each coordinate. In more generality, one may take as coupling time τ any stopping time at which the law of (b_τ, b'_τ) is the stationary law. For some examples where coupling times are explicitly constructed, see Lindvall (1992).

3 A Gaussian Example

In this section we present a particularly simple example of a filtering problem where computations using Section 2 can be carried out rather explicitly. Let x_t denote a stationary Orenstein-Uhlenbeck process, i.e.

$$dx_t = -\frac{1}{2}x_t dt + dw_t \,, \tag{1.7}$$

observed linearly in Gaussian white noise of intensity σ:

$$dy_t = x_t dt + \sigma d\nu_t \,. \tag{1.8}$$

Here, w_t, ν_t denote independent standard Brownian motions. We let b, b' denote independent stationary solutions to (1.7), and use p_s to denote the density of the standard normal law. We let

$$\bar{\Lambda}_t = \exp\left(\frac{1}{\sigma}\int_0^t (b_s - x_s)\, d\nu_s - \frac{1}{2\sigma^2}\int_0^t (b_s - x_s)^2 ds\right), \quad t \geq 0,$$

and

$$\bar{\Lambda}'_t = \exp\left(\frac{1}{\sigma}\int_0^t (b'_s - x_s)\, d\nu_s - \frac{1}{2\sigma^2}\int_0^t (b'_s - x_s)^2 ds\right), \quad t \geq 0\,.$$

It is straightforward to check that now, (1.5) reads

$$\limsup_{t\to\infty}\frac{1}{t}\log\|p_t - p'_t\| \leq \limsup_{t\to\infty}\frac{1}{t}\log\left(\frac{E_{b,b'}\left(\frac{p(b_0)p'(b_0)}{p_s(b_0)p_s(b'_0)}\bar{\Lambda}_t\bar{\Lambda}'_t\mathbf{1}_{V_t}\right)}{E_{b,b'}\left(\frac{p(b_0)p'(b'_0)}{p_s(b_0)p'_s(b'_0)}\bar{\Lambda}_t\bar{\Lambda}'_t\right)}\right)\,. \tag{1.9}$$

In order not to be burdened by (inessential) technicalities, we make the following:

Assumption A There exists a constant $C > 0$ such that

$$C^{-1} \leq \inf_{x\in\mathbb{R}} \frac{p(x)}{p_s(x)} \leq \sup_{x\in\mathbb{R}} \frac{p(x)}{p_s(x)} \leq C\,,\; C^{-1} \leq \inf_{x\in\mathbb{R}} \frac{p'(x)}{p_s(x)} \leq \sup_{x\in\mathbb{R}} \frac{p'(x)}{p_s(x)} \leq C\,.$$

Our goal is to prove the:

Theorem 1 *Assume p, p' satisfy assumption A. Then,*

$$\limsup_{\sigma\to 0} \limsup_{n\to\infty} \frac{\sigma}{n} \log \|p_n - p'_n\| < 0\,,\ a.s.. \tag{1.10}$$

In fact, there exist $\alpha, \beta > 0$ such that for all t large enough,

$$P(\|p_t - p'_t\| > \exp(-\alpha t/\sigma)) \leq \exp(-\beta t/\sigma)\,. \tag{1.11}$$

A control of the right hand side of (1.10) is possible, however since our bounds are not expected to be particularly tight we do not try to make it explicit.

Proof of Theorem 1

Throughout this proof, we use t to denote the time index, but except when computing expectations we will always think of $t = 0, 1, \ldots$. Thus, the statement

$$\limsup_{t\to\infty} a_t = 0, \text{a.s.}\,,$$

is taken to say that $\limsup_{n\to\infty} a_n = 0$, a.s. We also note that (1.11) follows readily from our proof of (1.10) by a Chebycheff inequality. Thus, we concentrate here on proving the later.

Obviously, under assumption A, the right hand side of (1.9) reads

$$\begin{aligned}
\limsup_{t\to\infty} \frac{1}{t} \log \|p_t - p'_t\| &\leq \limsup_{t\to\infty} \frac{1}{t} \log \left(\frac{E_{b,b'}\left(\bar{\Lambda}_t \bar{\Lambda}'_t \mathbf{1}_{V_t}\right)}{E_{b,b'}\left(\bar{\Lambda}_t \bar{\Lambda}'_t\right)} \right) \\
&:= \limsup_{t\to\infty} \frac{1}{t} \log \frac{B_1^t}{B_2^t}\,. \qquad (1.12)
\end{aligned}$$

For $t < \infty$, let $(\lambda_i^t)^2$ denote the Karhunen-Löeve eigenvalues and let $\varphi_i^t(\cdot)$ denote the associated eigenfunctions. Let $\{a_i\}, \{a'_i\}$ denote independent sequences of independent standard normal random variables, and let $\phi_i^t = \int_0^t \varphi_i^t(s) x_s ds$, $\nu_i^t = \int_0^t \varphi_i^t(s) d\nu(s)$. Then, ϕ_i^t are independent normal variables of zero mean and variance $1/(\lambda_i^t)^2$, ν_i^t are independent standard normal variables, and one has the identities

$$\begin{aligned}
I_1^t := \int_0^t (b_s - x_s)^2 ds &= \sum_{i=1}^\infty \left(\frac{a_i}{\lambda_i^t} - \phi_i^t \right)^2\,, \\
I_1'^t := \int_0^t (b'_s - x_s)^2 ds &= \sum_{i=1}^\infty \left(\frac{a'_i}{\lambda_i^t} - \phi_i^t \right)^2\,,
\end{aligned}$$

$$\begin{aligned} I_2^t := \int_0^t (b_s - x_s) d\nu_s &= \sum_{i=1}^{\infty} \left(\frac{a_i}{\lambda_i^t} - \phi_i^t \right) \nu_i^t \,, \\ I_2'^t := \int_0^t (b_s' - x_s) d\nu_s &= \sum_{i=1}^{\infty} \left(\frac{a_i'}{\lambda_i^t} - \phi_i^t \right) \nu_i^t \,. \end{aligned}$$

Further, one knows that $\lambda_i^t \sim \pi i/t$. The definitions above are used in the proof of the following lemma, which is presented later.

Lemma 3 *For* $p \in [1, \infty)$,

$$\limsup_{\sigma \to 0} \limsup_{t \to \infty} \frac{\sigma}{t} \log E_b(\bar{\Lambda}_t^p) = \frac{1}{4} \left(p^{3/2} - 3\sqrt{p} \right) \,, \; a.s.. \tag{1.13}$$

In particular,

$$\limsup_{\sigma \to 0} \limsup_{t \to \infty} \frac{\sigma}{t} \log B_2^t = -1 \,, \; a.s.. \tag{1.14}$$

The key to the evaluation of a bound on the right hand side of (1.12) lies in restricting the domain of integration in B_1^t. For K a fixed constant (later taken as $K = 1 + \eta$ with $\eta > 0$ arbitrary small), define the events

$$A_t := \{\{a_i\} : \, I_1^t \le K \sigma t\} \,, \quad A_t' := \{\{a_i'\} : \, I_1'^t \le K \sigma t\} \,.$$

Then,

$$\begin{aligned} \frac{B_1^t}{B_2^t} &= \frac{E_{b,b'}(\bar{\Lambda}_t \bar{\Lambda}_t' \mathbf{1}_{V_t})}{E_b(\bar{\Lambda}_t)^2} \\ &\le \frac{E_{b,b'}(\bar{\Lambda}_t \bar{\Lambda}_t' \mathbf{1}_{A_t} \mathbf{1}_{A_t'} \mathbf{1}_{V_t})}{E_b(\bar{\Lambda}_t)^2} + 2 \frac{E_b(\bar{\Lambda}_t \mathbf{1}_{A_t^c})}{E_b(\bar{\Lambda}_t)} := \frac{B_{11}^t}{B_2^t} + 2 C_1^t \,. \end{aligned} \tag{1.15}$$

The proof of the following lemma is postponed to the end of this section:

Lemma 4 *With* K *as above, and any* $\eta > 0$,

$$\limsup_{\sigma \to 0} \limsup_{t \to \infty} \frac{\sigma}{t} \log C_1^t < 0 \,, \; a.s.. \tag{1.16}$$

Equipped with Lemmas 3 and 4, our next task is to achieve a control on B_{11}^t, which is contained in the following.

$$B_{11}^t := E_{b,b'}(\bar{\Lambda}_t \bar{\Lambda}_t' \mathbf{1}_{A_t} \mathbf{1}_{A_t'} \mathbf{1}_{V_t}) \le E_b^{2/p}(\bar{\Lambda}_t^p) E_{b,b'}^{1/q}(\mathbf{1}_{A_t} \mathbf{1}_{A_t'} \mathbf{1}_{V_t}) \,. \tag{1.17}$$

In view of (1.12), (1.15) and Lemma 4, Lemma 3), (1.17) and the following large deviations computation complete the proof of Theorem 1:

Lemma 5

$$\limsup_{\sigma \to 0} \limsup_{t \to \infty} \frac{\sigma}{t} \log E_{b,b'}(\mathbf{1}_{A_t} \mathbf{1}_{A_t'} \mathbf{1}_{V_t}) \le -K^{-1} \left(\frac{3 + \sqrt{3}}{4} \right)^2 \,, \; a.s.$$

Indeed, taking $p = 1+\delta$ with δ small enough and $\eta > 0$ small enough one obtains that

$$\limsup_{\sigma\to 0}\limsup_{t\to\infty}\frac{\sigma}{t}\log\left(\frac{B^t_{11}}{B^t_2}\right)\le \delta\left(1-\left(\frac{3+\sqrt{3}}{4}\right)^2+g_\eta\right)+o(\delta)<0\,,\ \text{a.s.},$$

where $g_\eta \to 0$ as $\eta\to 0$. The proof of Lemma 5 is presented below. $\square$

Proof of Lemma 5: It is enough to show that

$$\limsup_{n\to\infty}\frac{\sigma}{n}\log E\left(E_{b,b'}(\mathbf{1}_{A_n}\mathbf{1}_{A'_n}\mathbf{1}_{V_n})\right)\le -K^{-1}\left(\frac{3+\sqrt{3}}{4}\right)^2+g(\sigma)\,,\quad (1.18)$$

where $g(\sigma)\to 0$ as $\sigma\to 0$. Indeed, assume (1.18). Then, for any $0<\beta<1$,

$$\begin{aligned}&P\left(E_{b,b'}(\mathbf{1}_{A_n}\mathbf{1}_{A'_n}\mathbf{1}_{V_n})>\exp\left(-\frac{\beta n}{\sigma}\left(K^{-1}\left(\frac{3+\sqrt{3}}{4}\right)^2+g(\sigma)\right)\right)\right)\\&\le\exp\left(-\frac{(1-\beta)n}{\sigma}\left(K^{-1}\left(\frac{3+\sqrt{3}}{4}\right)^2+g(\sigma)+\bar g(n)\right)\right)\,,\end{aligned}$$

where $\bar g(n)\to 0$ as $n\to\infty$, implying by the Borel-Cantelli lemma that

$$\limsup_{\sigma\to 0}\limsup_{n\to\infty}\frac{\sigma}{n}\log E_{b,b'}(\mathbf{1}_{A_n}\mathbf{1}_{A'_n}\mathbf{1}_{V_n})\le -K^{-1}\left(\frac{3+\sqrt{3}}{4}\right)^2\,,\ \text{a.s.}$$

For $t\in \mathbb{R}$, let $\lfloor t\rfloor$ ($\lceil t\rceil$, respectively) denote the greatest (smallest) integer smaller than or equal to t (greater than or equal to t). Then the conclusion of the lemma follows by noting that, from monotonicity,

$$\frac{\sigma}{t}\log E_{b,b'}(\mathbf{1}_{A_t}\mathbf{1}_{A'_t}\mathbf{1}_{V_t})\le\frac{\sigma}{\lceil t\rceil}\log E_{b,b'}(\mathbf{1}_{A_{\lfloor t\rfloor}}\mathbf{1}_{A'_{\lfloor t\rfloor}}\mathbf{1}_{V_{\lfloor t\rfloor}})\,.$$

Turning to the proof of the key estimate (1.18), the following transformation reduces the three-dimensional problem into two-dimensional one:

$$\begin{pmatrix} b_. - x_. \\ b'_. - x_. \end{pmatrix}=\begin{pmatrix}\sqrt{3/2}u_. + v_./\sqrt{2}\\ \sqrt{3/2}u_. - v_./\sqrt{2}\end{pmatrix}\,,\qquad (1.19)$$

where $u_., v_.$ are two independent Orenstein-Uhlenbeck processes satisfying the SDE (1.7). Rewrite now (1.18) in terms of $u_., v_.$ to obtain that

$$\begin{aligned}&\limsup_{n\to\infty}\frac{\sigma}{n}\log E\left(E_{b,b'}(\mathbf{1}_{A_n}\mathbf{1}_{A'_n}\mathbf{1}_{V_n})\right)\\&\le\limsup_{n\to\infty}\frac{\sigma}{n}\log P\Big(\frac{1}{t}\int_0^t(\sqrt{3}u_s-v_s)^2ds\le 2K\sigma,\end{aligned}$$

$$\begin{aligned}
&\frac{1}{t}\int_0^t (\sqrt{3}u_s+v_s)^2 ds \le 2K\sigma, \frac{1}{t}\int_0^t \mathbf{1}_{v_s<0} ds = 0\Big) \\
&\le -\inf\Big\{\frac{1}{8}\iint \frac{f_x^2+f_y^2}{f}(x,y)m(dx,dy) : \\
&\qquad f\ge 0, \iint f(x,y)m(dx,dy)=1, \\
&\qquad \iint (\sqrt{3}x+y)^2 f(x,y)m(dx,dy) \le 2K\sigma, \\
&\qquad \iint (\sqrt{3}x-y)^2 f(x,y)m(dx,dy) \le 2K\sigma, \\
&\qquad f(x,y)=f(x,y)\mathbf{1}_{y\ge 0}\Big\}, \qquad (1.20)
\end{aligned}$$

where $m(dx,dy) = \exp(-(x^2+y^2)/2)dxdy/2\pi$, f_x, f_y denote the partial derivatives of $f(x,y)$ w.r.t. x, y, respectively, and the last inequality in (1.20) is a consequence of the Donsker-Varadhan theorem for occupation measures, c.f. Deuschel and Stroock (1989) pg. 241 for this version. The conclusion now follows by solving the calculus of variations problem (1.20), as outlined below.

The cost function in the constrained problem is

$$J(f) = \iint \left(\frac{1}{8}\frac{f_x^2+f_y^2}{f} + r(x,y)f\right) m(dx,dy)$$

where

$$r(x,y) = \lambda_1(\sqrt{3}x+y)^2 + \lambda_2(\sqrt{3}x-y)^2 + \mu.$$

The Euler-Lagrange equation then has the form

$$\frac{1}{8}\|\nabla \log f\|^2 + \frac{1}{4}(\Delta \log f - x(\log f)'_x - y(\log f)'_y) = r(x,y)f.$$

The solution to the above, which is supported on $\{y\ge 0\}$ and satisfies

$$\iint f(x,y)m(dx,dy)=1, \quad \iint (\sqrt{3}x\pm y)^2 f(x,y)m(dx,dy) = 2K\sigma$$

is given by

$$f(x,y) = c_1 y^2 e^{-(\gamma x^2+y^2)/2\alpha^2}\mathbf{1}_{y\ge 0}$$

where c_1 is a normalization constant, $\gamma = -\alpha^2 + \sqrt{\alpha^4 + 3(1+2\alpha^2)}$ and for small values of σ,

$$\alpha^2 = \frac{2K\sigma}{3+\sqrt{3}} + o(\sigma).$$

Substituting into the rate function one obtains

$$\frac{1}{8}\iint \frac{f_x^2+f_y^2}{f}(x,y)m(dx,dy) = \left(\frac{3+\sqrt{3}}{4}\right)^2 (K\sigma)^{-1} + o(\sigma^{-1})$$

and (1.18) follows. $\square$

Proof of Lemma 3 Write

$$\bar{\Lambda}_t^p = \exp\left(\frac{p}{\sigma}\sum_{i=1}^{\infty}(\frac{a_i}{\lambda_i^t} - \phi_i^t)\nu_i^t - \frac{p}{2\sigma^2}\sum_{i=1}^{\infty}(\frac{a_i}{\lambda_i^t} - \phi_i^t)^2\right).$$

Noting that the exponent is nothing but a quadratic form in the random variables a_i, taking the expectation reveals that

$$E_b\bar{\Lambda}_t^p = \prod_{i=1}^{\infty}\frac{1}{\sqrt{1+p(\sigma\lambda_i^t)^{-2}}}\exp\left(\sum_{i=1}^{\infty}\frac{p(\nu_i^t)^2 - 2\phi_i^t\nu_i^t\sigma(\lambda_i^t)^2 - (\phi_i^t\lambda_i^t)^2}{2(\sigma^2\lambda_i^2/p+1)}\right).$$

An analysis of the first term reveals that

$$\begin{aligned}\log\prod_{i=1}^{\infty}\frac{1}{\sqrt{1+p(\sigma\lambda_i^t)^{-2}}} &\sim -\frac{1}{2}\sum_{i=1}^{\infty}\log(1+pt^2/\pi^2\sigma^2 i^2)\\ &\sim -\frac{t\sqrt{p}\int_0^\infty \log(1+x^{-2})dx}{2\sigma\pi} + o(t)) = -\frac{t\sqrt{p}}{2\sigma}.\end{aligned}$$

while for the second term one notes that

$$E_t := \sum_{i=1}^{\infty}\frac{p(\nu_i^t)^2 - 2\phi_i^t\nu_i^t\sigma(\lambda_i^t)^2 - (\phi_i^t\lambda_i^t)^2}{2(\sigma^2\lambda_i^2/p+1)}$$

is a sum of independent, random variables, whose variance is of order t, and

$$\lim_{t\to\infty} E(E_t)/t = \frac{(p-1)\sqrt{p}\int_0^\infty(1+x^2)^{-1}dx}{2\pi\sigma} = \frac{(p-1)\sqrt{p}}{4\sigma}.$$

Hence, $E_t/t \to (p-1)\sqrt{p}/4\sigma$, in probability, and a.s. on the sequence $t_n = n$ because $E(E_t - E(E_t))^4 = O(t^2)$ (for fixed σ). $\square$

Proof of Lemma 4 The key to the proof is Chebycheff's inequality: Fix $\beta > 0$, then

$$E_b(\bar{\Lambda}_t\mathbf{1}_{A_t^c}) \le e^{\frac{-\beta K t}{2\sigma}}E_b(\bar{\Lambda}_t\exp(\beta I_1^t/2\sigma^2)).$$

One now repeats the analysis of Lemma 3 to conclude that

$$\begin{aligned}\limsup_{t\to\infty}\frac{2\sigma}{t}\log E_b(\bar{\Lambda}_t\mathbf{1}_{A_t^c}) &\le -\beta K - \frac{\sqrt{1-\beta}\int_0^\infty\log(1+x^{-2})dx}{\pi}\\ &\quad + \frac{\beta}{\sqrt{1-\beta}\pi}\int_0^\infty(1+x^2)^{-1}dx\\ &= -\beta K - \sqrt{1-\beta} + \frac{\beta}{2\sqrt{1-\beta}}.\end{aligned}$$

Optimization over $\beta \ge 0$ and an application of Lemma 3 with $p = 1$ yield the result, as soon as $K > 1$. $\square$

4 A non-Gaussian Example

Consider next the stochastic process x_t solution of the SDE

$$dx_t = c(x_t)dt + dw_t\,, \tag{1.21}$$

and let y_t be as in (1.8). Define $F(x) = \int_0^x (c(\theta)+\theta/2)d\theta$. To avoid unessential technical difficulties, we use the following assumption:
Assumption B The functions $|F(x)|$, $|c'(x)|$ and $|c^2(x)-x^2/2|$ are bounded by a global constant C.
We now have the

Corollary 1 *Assume assumptions A and B. Then,*

$$\limsup_{\sigma\to 0}\limsup_{n\to\infty}\frac{\sigma}{n}\log\|p_n - p'_n\| < 0\,,\ \ a.s.. \tag{1.22}$$

Proof: Using assumptions A and B, one knows that the solution to (1.21) is absolutely continuous w.r.t. to the Orenstein-Uhlenbeck $b_\cdot$, with Radon-Nikodym derivative equal to

$$\begin{aligned}&\frac{p(b_0)}{p_s(b_0)}\exp\left(\int_0^t (c(b_s)+b_s/2)db_s - \frac{1}{2}\int_0^t (c^2(b_s)-(b_s/2)^2)ds\right)\\ &\quad\le Ce^{2C}\exp 2Ct\,,\end{aligned} \tag{1.23}$$

where the inequality above uses Ito's formula:

$$F(b_t) = F(b_0) + \int_0^t (c(b_s)+b_s/2)db_s + \frac{1}{2}\int_0^t (c'(b_s)+1/2)ds\,.$$

Hence, exactly as in the argument leading to (1.9),

$$\|p_t - p'_t\| \le C^3 e^{4C} e^{4Ct}\frac{E_{b,b'}(\bar\Lambda_t\bar\Lambda'_t\mathbf{1}_{V_t})}{(E_b\bar\Lambda_t)^2}\,, \tag{1.24}$$

with the difference from the Gaussian case lying in the fact that in the R.H.S. of (1.24), the random variables ϕ_i^t are neither independent nor Gaussian or uncorrelated. However, due to the proof of (1.11), there exist $\alpha,\beta>0$ such that

$$P_G\left(\frac{E_{b,b'}(\bar\Lambda_t\bar\Lambda'_t\mathbf{1}_{V_t})}{(E_b\bar\Lambda_t)^2} > \exp(-\alpha t/\sigma)\right) \le \exp(-\beta t/\sigma)\,, \tag{1.25}$$

where P_G denotes the measure under which ϕ_i^t are normal independent of variance $(\lambda_i^t)^2$. Applying now again a change of measure as in (1.23), we see that for t large enough,

$$\begin{aligned}&P\left(\frac{E_{b,b'}(\bar\Lambda_t\bar\Lambda'_t\mathbf{1}_{V_t})}{(E_b\bar\Lambda_t)^2} > \exp(-\alpha t/\sigma)\right)\\ &\quad\le e^{3Ct}P_G\left(\frac{E_{b,b'}(\bar\Lambda_t\bar\Lambda'_t\mathbf{1}_{V_t})}{(E_b\bar\Lambda_t)^2} > \exp(-\alpha t/\sigma)\right)\,,\end{aligned}$$

and the conclusion follows from (1.25). □

Remark Obviously, assumption B is far from optimal, and one may relax it by using appropriate Hölder inequalities. However, even at best, the technique of this section is quite limited and does not seem to allow one to go beyond the case of linear observation functions $h(\cdot)$ or, more generally, beyond the case of non-constant diffusion coefficients in the state process. What is needed in order to better employ the bound (1.5) is a direct way of controlling $E(\bar{\Lambda}_t^p)$ which does not involve the Karhunen-Löeve expansion. Rather, a conditional large deviation principle for the joint occupation measure of (b, b', x, ν) is needed. This will be reported elsewehere.

Acknowledgments: The work of R. Atar was supported in part by the Office of Naval Research (ONR-N00014-96-1-0276). The work of O. Zeitouni was partially supported by a grant from the basic research fund administered by the Israeli Academy of Sciences.

References

R. Atar (1997) *Exponential Stability for Nonlinear Filtering of Diffusion Processes in Non-Compact Domain,* submitted.

R. Atar and O. Zeitouni (1997a) *Lyapunov Exponents for Finite State Nonlinear Filtering,* Siam J. Contr. Optim., 35, pp. 36–55.

R. Atar and O. Zeitouni (1997b) *Exponential Stability for Nonlinear Filtering,* Ann. Inst. Hen. Poincare, 33, pp. 697–725.

V. S. Borkar, S. K. Mitter and S. Tatikonda (1997) *Optimal Sequential Vector Quantization of Markov Sources,* preprint.

A. Budhiraja and H. J. Kushner (1997a) *Robustness of Nonlinear Filters Over the Infinite Time Interval*, to appear in Siam J. Contr. Optim.

A. Budhiraja and H. J. Kushner (1997b) *Approximation and Limit Results for Nonlinear Filters Over an Infinite Time Interval*, preprint.

A. Budhiraja and D. L. Ocone (1997) *Exponential Stability of Discrete Time Filters for Bounded Observation Noise*, Systems and Control Letters, 30, pp. 185–193.

J. M. C. Clark, D. L. Ocone and C. Coumarbatch (1997) *Relative Entropy and Error Bounds for Filtering of Markov Process,* preprint.

J. D. Deuschel and D. W. Stroock (1989) Large Deviations. Academic Press, Boston.

G. Da Prato, M. Fuhrman, P. Malliavin (1995) *Asymptotic Ergodicity for the Zakai Filtering Equation*, C. R. Acad. Sci. Paris, t. 321, Série I, pp. 613–616.

H. Kunita (1971) *Asymptotic Behavior of the Nonlinear Filtering Errors of Markov Processes,* J. Multivariate Anal., 1, pp. 365–393.

T. Lindvall (1992) Lectures on the coupling method, Wiley, New York.

D. L. Ocone (1997a) *Asymptotic Stability of Beneš Filters*, preprint.

D. L. Ocone (1997b) this volume.

D. L. Ocone and E. Pardoux (1996) *Asymptotic Stability of the Optimal Filter with respect to its Initial Condition*, Siam J. Contr. Optim., 34, pp. 226–243.

L. Stettner (1989) *On Invariant Measures of Filtering Processes,* Stochastic Differential Systems, Proc. 4th Bad Honnef Conf., 1988, Leture Notes in Control and Inform. Sci. 126, edited by Christopeit, N., Helmes, K. and Kohlmann, M., Springer, pp. 279–292.

L. Stettner (1991) *Invariant Measures of the Pair: State, Approximate Filtering Process,* Colloq. Math., LXII, pp. 347–351.

Rami Atar, Division of Applied Mathematics, Brown University, Providence, RI 02912, USA. rami@cfm.brown.edu
Frederi Viens, Department of Mathematics, University of North Texas, Denton, TX 76203-5118, USA. fviens@unt.edu
Ofer Zeitouni, Department of Electrical Engineering, Technion, Haifa 32000, ISRAEL. zeitouni@ee.technion.ac.il