The Annals of Probability
1998, Vol. 26, No. 4, 1552–1574

EXPONENTIAL STABILITY FOR NONLINEAR FILTERING OF DIFFUSION PROCESSES IN A NONCOMPACT DOMAIN[1]

BY RAMI ATAR

Technion–Israel Institute of Technology

The optimal nonlinear filtering problem for a diffusion process in a noncompact domain, observed in white noise, is considered. It is assumed that the process is ergodic, the diffusion coefficient is constant and the observation is linear. Using known bounds on the conditional density, it is shown that when the observation noise is sufficiently small, the filter is exponentially stable, and that the decay rate of the total variation distance between differently initialized filtering processes tends to infinity as the noise intensity approaches zero.

1. Introduction. For a pair of stochastic processes, $\{x_t, 0 \le t < \infty\}$ termed a *state process*, and $\{y_t, 0 \le t < \infty\}$ termed an *observation process*, with a given joint law, the filtering problem consists of the recursive computation of the conditional law of x_t given the observations in the past, $\{y_s, 0 \le s \le t\}$. The solution to the problem is called the optimal filter, and the conditional law, considered as a measure-valued process, is called the filtering process. In many cases the time-evolution of the filtering process can be described by equations such as the Kushner equation or the Zakai equation, and in some simple cases it can be recovered from the finite-dimensional Kalman–Bucy or Beneš equations. All these equations are driven by the observation process and are initialized in accordance with the prior law at time zero. In this article we study the sensitivity of the filter to perturbations in its initialization, in a certain class of problems where $\{(x_t, y_t)\}$ forms a diffusion process. Several works in the past have been devoted to related questions. Kunita [11] and Stettner [18, 19] have studied the ergodic properties of the filtering process and the convergence of the filter to its ergodic behavior. The stability of the filter to perturbations in its initial conditions has been studied by Delyon and Zeitonni [10] and by Ocone and Pardoux [16]. Bounds on the decay rate of the variation distance between the responses of the filter to different initializations have been provided by Atar and Zeitonni [1], [2]. Recently, Budhiraja and Ocone [7] proved exponential stability for models in which the state process may even be nonergodic. Le Gland and Mevel [13] proved exponential stability of the filter with respect to its initial condition under misspecification of the assumed model. Cérou [8] studied the asymptotic consistency of the filter for a noise-free state process. Some more recent related results appear in [3], [5], [6], [9], [14], [15].

Received April 1997; revised April 1998.
[1]Supported in part by a grant from the Israel Science Foundation.
AMS 1991 *subject classifications*. 93E11, 93E15, 60G35, 60J60.
Key words and phrases. Nonlinear filtering, exponential stability, conditional density bounds.

As for diffusion processes, it has been shown [2] that for a strictly elliptic diffusion on a compact manifold observed by a smooth $\mathbb{R}^d$-valued function with additive white noise, the filter is stable for all noise-level values. The technique used there, which involves Birkhoff's contraction coefficient, does not seem to be applicable when noncompact domains are considered. Intuitively, however, we expect that compactness of the domain is not essential for stability. The goal of this paper is to show that at least in the one-dimensional noncompact case, where the diffusion coefficient is constant and the observation function is linear, the filter is indeed stable. The reason that we pick out this model is that this is the only nontrivial case, where explicit upper and lower bounds on the conditional density are known [21], which are also tight when the observation noise-level is low. Using these bounds, we can show that the nonlinear filter is stable when the noise is sufficiently small, and moreover, that the decay rate tends to infinity as the noise-level approaches zero. An analogue of the latter result is yet unknown for the filtering of a diffusion process on a compact state space.

To control the variation distance between the responses of the filter to two given initializations $p_0^{(i)}$, $i = 1, 2$, we use an inequality (2.20), involving the two associated unnormalized conditional densities, $\rho_t^{(i)}$ (see definition in Section 2) and their exterior product, $\rho_t^{(1)} \wedge \rho_t^{(2)}$. This approach has formerly been taken in [2], Section 5. With this inequality, the problem of controlling the decay rate is translated into that of estimating the evolution rates of $\rho_t^{(i)}$ and of $\rho_t^{(1)} \wedge \rho_t^{(2)}$ that are both governed by *linear* flows. Both estimates are established by looking at the subsequences at times $n = 1, 2, \dots$. An upper bound on $\|\rho_t^{(1)} \wedge \rho_t^{(2)}\|_1$ is obtained by bounding the norms of the operators $\rho_{n-1}^{(1)} \wedge \rho_{n-1}^{(2)} \mapsto \rho_n^{(1)} \wedge \rho_n^{(2)}$, while a lower bound on $|\rho_n|_1$ is obtained by writing it as an n-fold integral and using the positivity of ρ_n. However, since explicit expressions for ρ_t are not available, we use instead Zeitouni's bounds on the densities ρ_t, which have explicit form. Ergodic considerations are then invoked to obtain an a.s. result.

In Section 2 we state our main result, quote Zeitouni's density bounds and sketch the proof. Section 3 is devoted to estimates on the evolution rates of the unnormalized densities and of their exterior product. In Section 4 we study the small noise asymptotics of the density bounds under some ergodic assumptions. Finally, all parts are combined in Section 5 to obtain the proof of the main result.

2. Setting and main result.

Our model for the state and observation processes is as follows:

$$dx_t = f(x_t)\,dt + dw_t, \quad x_t \in \mathbb{R}, \tag{2.1}$$

$$dy_t = x_t\,dt + N_0^{1/2}\,d\nu_t, \quad y_t \in \mathbb{R}, \quad y_0 = 0. \tag{2.2}$$

Here $\{w_t, 0 \le t < \infty\}$ and $\{\nu_t, 0 \le t < \infty\}$ are independent standard Brownian motions, and x_0 has density p_0, with respect to the Lebesgue measure on $\mathbb{R}$. We make the following assumption on f.

ASSUMPTION 1. The function $V(x) = f'(x) + f^2(x)$ is twice continuously differentiable with a bounded second derivative.

We denote by P the measure induced by the pair (2.1), (2.2) and the initial density p_0 and by E expectation with respect to it. The next assumption indicates stationarity and ergodicity of $\{x_t\}$ in an asymptotic sense. To state it, we define also P^q to be the measure induced by (2.1), (2.2) where the density of x_0 is q instead of p_0. Obviously, we have $P = P^{p_0}$.

ASSUMPTION 2(a). There exists an initial density $\tilde{p}$, such that under $P^{\tilde{p}}$, $\{x_t\}$ is stationary and ergodic.

We denote $\tilde{P} = P^{\tilde{p}}$ for one such $\tilde{p}$; $\tilde{E}$ denotes expectation with respect to it.

ASSUMPTION 2(b). For some $t \geq 0$, the marginal law of x_t under P is absolutely continuous with respect to that under $\tilde{P}$.

Furthermore, we shall assume the following.

ASSUMPTION 3. One has that $\tilde{E}x_t^2 < \infty$.

REMARKS. (a) We treat only the case where the diffusion coefficient is constant (2.1) and the observation function is linear (2.2), because there exist explicit upper and lower bounds on the conditional density (Theorem 2), applying only under these assumptions, that make it possible to control the decay rate. Under Assumption 1, these bounds are also tight as $N_0 \to 0$, in a sense explained below [see the remark after (2.12)].

(b) It is useful to note that Assumption 1 implies that f grows at most linearly that is, there exists a constant C_0 such that

$$|f(x)| \leq C_0 + C_0|x|. \tag{2.3}$$

(c) Assumption 2 implies that P is absolutely continuous with respect to $\tilde{P}$ on the tail σ-field.

Let us denote by P_0 the measure induced by

$$dx_t = f(x_t)\,dt + dw_t,$$
$$dy_t = N_0^{1/2}\,d\nu_t, \qquad y_0 = 0,$$

where $\{w_t\}$ and $\{\nu_t\}$ are as above, and the density of x_0 is again p_0. Let also $P^{(t)}$ ($P_0^{(t)}$) denote the restriction of P (P_0, respectively) to the sigma-field generated by the trajectories up to time t, namely, to $\sigma\{(x_s, y_s), 0 \leq s \leq t\}$. Then it is known that $P^{(t)}$ and $P_0^{(t)}$ are mutually absolutely continuous. Moreover, let us denote by $\mathscr{Y}_{s,t}$ the sigma-field generated by $\{y_\theta, s \leq \theta \leq t\}$ and by p_t the conditional density of x_t given $\mathscr{Y}_{0,t}$. Then if we define

$$\rho_t(z) = E_0\left(\frac{dP^{(t)}}{dP_0^{(t)}}\bigg|\mathscr{Y}_{0,t}, x_t = z\right)p_{x_t}(z),$$

where $p_{x_t}(\cdot)$ is the density of x_t under P_0, we have that P-a.s. $\rho_t(z)/\int_{\mathbb{R}}\rho_t(r)\,dr = p_t(z)$. We call ρ_t the unnormalized density, and under regularity conditions on p_0 (see, e.g., [17]), it is known to satisfy the Zakai equation [20].

By the smoothing property for the conditional expectation, one can show (see, e.g., [2]) that there is a linear relation between ρ_s and ρ_t, in the sense that if $s < t$, then there exists an integral kernel, namely, a real-valued measurable function $\mathsf{T}(s,\xi;t,z)$, for which

$$\rho_t(z) = \int_{\mathbb{R}} \mathsf{T}(s,\xi;t,z)\rho_s(\xi)\,d\xi.$$

This kernel is given by

$$\mathsf{T}(s,\xi;t,z) = E_0\left[\frac{dP}{dP_0}\middle|\mathscr{Y}_{s,t}, x_s=\xi, x_t=z\right]p_{x_t\,|\,x_s}(z\,|\,\xi), \tag{2.4}$$

where $p_{x_t\,|\,x_s}(\cdot\,|\,\xi)$ is the density of x_t conditioned on $x_s=\xi$ under P_0 (equivalently, under P).

In order to deal with perturbed initial conditions, it is necessary first to define the response p_t^p of the filter to an arbitrary initial condition p. A definition in that spirit appears already in [11], and later also, for example, in [16] and [2], and in the present context it can be described as follows. Let the density p be an arbitrary nonnegative element of $L_1(\mathbb{R})$ whose L_1 norm is $|p|_1 = 1$. Then we let

$$p_t^p(z) = \frac{\rho_t^p(z)}{\int_{\mathbb{R}}\rho_t^p(r)\,dr},$$

where

$$\rho_t^p(z) = \int_{\mathbb{R}} \mathsf{T}(0,\xi;t,z)p(\xi)\,d\xi.$$

Note that $\mathsf{T}(s,\xi;t,z)$ is also the fundamental solution to the Zakai equation and that $\rho_t^p(z)$ ($p_t^p(z)$) agrees with the solution to the Zakai equation (respectively, the Kushner equation) initialized with p, under further conditions that guarantee existence and uniqueness of solutions.

Let now $p_0^{(i)}$, $i=1,2$ be two nonnegative elements of $L_1(\mathbb{R})$ with $|p_0^{(i)}|_1 = 1$ and define $\rho_t^{(i)} = \rho_t^{p_0^{(i)}}$ and $p_t^{(i)} = p_t^{p_0^{(i)}}$. Our main result is the following.

THEOREM 1. *Under Assumptions 1, 2 and 3 there exists a nonrandom constant C_1, independent of N_0 and of $p_0^{(i)}$, such that P-a.s.,*

$$\limsup_{t\to\infty}\frac{1}{t}\log\left|p_t^{(1)} - p_t^{(2)}\right|_1 \le \frac{1}{4}\log N_0 + C_1.$$

REMARKS. (a) Definitely, in order that the filter may produce the correct result, namely, have its solution p_t^p equal to the conditional density p_t, its initial condition must be set to p_0. However, one should not confuse the initial condition of the filter, which is *perturbed* in the above discussion, with the initial density p_0 of $\{x_t\}$, which is kept *fixed* throughout.

(b) We know that the bound in Theorem 1 is not tight as $N_0 \to 0$. In fact, in the special case of the Kalman–Bucy filter, where one has explicit expressions for p_t^p, the decay rate can be precisely computed, and the result is of the order of $N_0^{-1/2}$, rather than of logarithmic order. Some recent results that exhibit behavior of the order of $N_0^{-1/2}$ in cases where the dynamics is nonlinear appear in [3]. Also, see [15] for a related study of the Beneš filter.

Bounds on the conditional density. For the model (2.1) and (2.2), bounds on the conditional density are provided in [21], Corollary 1. We repeat these bounds in what follows. Let

$$\Phi(z) = \exp \int_0^z f(\theta)\, d\theta.$$

Throughout, the presence of the sign $\circ$ in a stochastic integral denotes a Stratonovitch integral, while its absence denotes an Itô integral.

THEOREM 2 (Zeitouni). *Let there exist $P_{i,t}$, $Q_{i,t}$ and $k_{i,t}$ $(i = 1, 2,\ t \geq 0)$ that are $\mathscr{Y}_{0,t}$-adapted processes, satisfying*

$$\tfrac{1}{2} P_{1,t} x^2 + Q_{1,t} x + k_{1,t} \leq \tfrac{1}{2} V(x) \leq \tfrac{1}{2} P_{2,t} x^2 + Q_{2,t} x + k_{2,t}. \tag{2.5}$$

Assume

$$p_0(z) \propto \Phi(z) \exp\bigl(-(z-\xi)^2/2\gamma\bigr)$$

for some $\xi \in \mathbb{R}$ and $\gamma > 0$. Then the following bounds hold P-a.s.:

$$\frac{\Phi(z)}{\Phi(\xi)} u_2(t, z) \leq \rho_t(z) \leq \frac{\Phi(z)}{\Phi(\xi)} u_1(t, z). \tag{2.6}$$

Above,

$$u_i(t, z) = K_i(t) \exp \frac{-\bigl(z - \mu_i(t)\bigr)^2}{2\Lambda_i(t)},$$

and $\Lambda_i(t)$, $\mu_i(t)$ and $K_i(t)$ are the solutions of

$$-\frac{d}{dt}\Lambda_i^{-1}(t) = \Lambda_i^{-2}(t) - \alpha_i^2, \qquad \Lambda_i^{-1}(0) = \gamma^{-1}, \tag{2.7}$$

$$\Lambda_i^{-1}(t)\, d\mu_i(t) = -\alpha_i^2 \mu_i(t)\, dt - Q_{i,t}\, dt + N_0^{-1} \circ dy_t, \qquad \mu_i(0) = \xi, \tag{2.8}$$

$$\begin{aligned} K_i^{-1}(t) \circ dK_i(t) &= \Lambda_i^{-1}(t)\mu_i(t) \circ d\mu_i(t) + \tfrac{1}{2}\alpha_i^2 \mu_i^2(t)\, dt \\ &\quad - \tfrac{1}{2}\Lambda_i^{-1}(t)\, dt - k_{i,t}\, dt, \qquad K_i(0) = \kappa, \end{aligned} \tag{2.9}$$

with

$$\alpha_i = \sqrt{N_0^{-1} + P_i}$$

$$\kappa = \frac{\Phi(\xi)}{\int \Phi(z) \exp(-(z-\xi)^2/2\gamma)\, dz}.$$

Under Assumption 1, let V denote a constant such that $|V''(z)/2| \leq V$ for all z. In Theorem 3 of [21], the choice for the coefficients P_i, Q_i and k_i is taken to be compatible with the expansion of $V(z)$ around $z = \mu_1(t)$; namely, they are chosen as follows:

$$2k_{i,t} = V\big(\mu_1(t)\big) - \mu_1(t)V'\big(\mu_1(t)\big) + (-1)^i V\mu_1^2(t), \tag{2.10}$$

$$Q_{i,t} = \tfrac{1}{2}V'(\mu_1(t)) - (-1)^i V\mu_1(t), \tag{2.11}$$

$$P_i = P_{i,t} = (-1)^i V. \tag{2.12}$$

The existence of coefficients P_i, Q_i and k_i satisfying (2.5) follows from Assumption 1. Furthermore, it is shown in Theorems 2 and 3 of [21] that under Assumption 1, with the above choice for the coefficients, the bounds (2.6) are tight. In particular, as $N_0 \to 0$, one has $\Lambda_i(t) \to 0$, $\Lambda_1(t)/\Lambda_2(t) \to 1$ and P-a.s. (conditioned on $\mathscr{Y}_{0,t}$), $K_1(t)/K_2(t) \to 1$ and $(\mu_i(t) - \hat{x}_t)/\Lambda_1^{1/2}(t) \to 0$, where $\hat{x}_t = E(x_t|\mathscr{Y}_{0,t})$. Although we do not use this result here directly, we adopt this choice for the coefficients in what follows, in a way to be explained below.

For our purpose, it is desired to have bounds on the kernel $\mathrm{T}(s,\xi;t,z)$, rather than on the density $\rho_t(z) = \int \mathrm{T}(0,\xi;t,z)p_0(\xi)\,d\xi$. This may be obtained by a slight modification of Theorem 2 as follows. We associate the initial conditions of (2.7), (2.8) and (2.9) with time s (rather than 0), take $\gamma = 0$, and look at the bounds as functions of s, ξ, t and z. More precisely, let us write the solutions to (2.7) and (2.8) with initial conditions $\Lambda_i(s) = 0$ and $\mu_i(s) = \xi$, respectively:

$$\begin{aligned}
\Lambda_i(s;t) &= \frac{1}{\alpha_i}\tanh\big(\alpha_i(t-s)\big),\\
\mu_i(t) &= \mu_i(s,\xi;t)\\
&= \frac{1}{\cosh(\alpha_i(t-s))}\bigg\{\xi + \int_s^t \cosh(\alpha_i(\theta-s))\big(-Q_{i,\theta}\Lambda_i(s;\theta)\,d\theta\\
&\qquad\qquad + \Lambda_i(s;\theta)N_0^{-1}\circ dy_\theta\big)\bigg\}.
\end{aligned} \tag{2.13}$$

Next, let us define

$$\begin{aligned}
\Psi_i(s,\xi;t) &= \frac{1}{\Phi(\xi)\sqrt{2\pi\sinh(\alpha_i(t-s))}}\\
&\quad\times\exp\bigg\{-\int_s^t\bigg(k_{i,\theta} + \frac{\alpha_i^2}{2}\mu_i^2(\theta) + Q_{i,\theta}\mu_i(\theta)\bigg)\,d\theta\\
&\qquad\qquad + \int_s^t N_0^{-1}\mu_i(\theta)\circ dy_\theta\bigg\}.
\end{aligned} \tag{2.14}$$

Then as a consequence of Theorem 2 we have the following.

COROLLARY 1. *Let $P_{i,t}$, $Q_{i,t}$ and $k_{i,t}$ be any $\mathscr{Y}_{s,t}$-adapted processes satisfying (2.5), and let $\Lambda_i(s;t)$, $\mu_i(s,\xi;t)$ and $\Psi_i(s,\xi;t)$ be as above. Define*

$$T_i(s,\xi;t,z) = \Psi_i(s,\xi;t)\Phi(z)\exp\frac{-(z-\mu_i(s,\xi;t))^2}{2\Lambda_i(s;t)}.$$

Then for all $t > s$ we have P-a.s.,

$$T_2(s,\xi;t,z) \le \mathsf{T}(s,\xi;t,z) \le T_1(s,\xi;t,z). \tag{2.15}$$

The proof in [21] of Theorem 2 above is based on the stochastic Feynman–Kac formula [22]. To prove the corollary along the same lines, one needs to consider a degenerate case, where ρ_t is initialized with a Dirac measure at ξ, and then to explore the dependence on ξ. In relation to this approach, see [12]. We choose an alternative method, using uniqueness of weak solutions to the associated filtering equation (based on a result of [17]).

For the proof of Corollary 1, see the Appendix.

To make the bounds explicit, it is left to determine the choice of the coefficients $P_{i,t}$, $Q_{i,t}$ and $k_{i,t}$ for each value of ξ. Therefore, we let ξ_0 be a nonrandom constant and for $t \ge s$ denote

$$\mu_i^0(s;t) = \mu_i(s,\xi_0;t).$$

Then, for s *fixed*, we choose the coefficients to be those that are defined in (2.10), (2.11) and (2.12) with $\mu_1^0(s;t)$ substituted instead of $\mu_1(t)$, for all values of ξ.

It is important to observe that by this definition, the coefficients $Q_{i,t}$ and $k_{i,t}$ depend upon the value of s that we have fixed. To denote this dependence, we introduce the notation $Q_{i,t}^{(s)}$, $k_{i,t}^{(s)}$ to be used when s to which the coefficients correspond is not clear and discarded otherwise.

REMARK. There is no circularity in the definition of Q_i and μ_i, as one may suspect at a first glance. In fact, once s is fixed, $Q_{1,t}$ and $\mu_1^0(s;t)$ are just the solution to the set of equations (2.8) and (2.11) for $t > s$, with the initial condition $\mu_1^0(s;s) = \xi_0$. Occasionally, $Q_{2,t}$ is determined by (2.11) and $\mu_i(s,\xi;t)$ by (2.8) with $\mu_i(s,\xi;s) = \xi$ [alternatively, by (2.13)].

A useful representation of (2.14) is obtained if one writes (2.13) as

$$\mu_i(s,\xi;t) = \mu_i^0(s;t) + \frac{\xi-\xi_0}{\cosh(\alpha_i(t-s))} \tag{2.16}$$

and concludes that the exponent in (2.14) is quadratic in ξ, namely

$$\Psi_i(s,\xi;t) = \frac{1}{\Phi(\xi)}\exp\biggl(\frac{-\bigl(\xi-a_i(s;t)\bigr)^2}{2b_i(s;t)} + c_i(s;t)\biggr). \tag{2.17}$$

Solving for a, b and c, one gets

$$b_i(s;t) = \big(\alpha_i \tanh(\alpha_i(t-s))\big)^{-1},$$

$$a_i(s;t) = \xi_0 + b_i(s;t)\bigg(-\int_s^t \frac{\big(\alpha_i^2 \mu_i^0(s;\theta) + Q_{i,\theta}\big)\, d\theta}{\cosh\big(\alpha_i(\theta - s)\big)} + \int_s^t \frac{dy_\theta}{N_0 \cosh\big(\alpha_i(\theta - s)\big)}\bigg), \tag{2.18}$$

$$c_i(s;t) = \log\big(\Psi_i(s, \xi_0; t)\Phi(\xi_0)\big) + \frac{\big(a_i(s;t) - \xi_0\big)^2}{2b_i(s;t)}. \tag{2.19}$$

Notation. The following notation will be used throughout. For an $L_p(\mathbb{R})$ function $\phi(\cdot)$ we denote by $|\phi|_p$ its L_p norm ($p = 1$ or $p = \infty$). Let $\mathscr{T}$ be an operator on $L_1(\mathbb{R})$, that possesses an integral kernel $T(\cdot, \cdot)$, namely, it is given by

$$\mathscr{T}\phi(y) = \int_{\mathbb{R}} T(x, y)\phi(x)\, dx.$$

Then we denote by $|T|_{\mathrm{op}}$ its operator norm on $L_1(\mathbb{R})$. For $D \subset \mathbb{R}$, $\phi|_D$ is the restriction of ϕ to D, and $T|_D$ is the restriction of T to D with respect to the first argument (i.e., $T|_D(x, y) = T(x, y)$, $x \in D$, $y \in \mathbb{R}$). For $r > 0$, $\mathscr{B}_r(x)$ denotes the interval $(x - r, x + r)$. For two real-valued measurable functions u and v, we let the exterior product $u \wedge v$ be the $\mathbb{R}^2 \to \mathbb{R}$ mapping defined by $(x, y) \mapsto u(x)v(y) - u(y)v(x)$, and let $\|u \wedge v\|_1$ denote its $L_1(\mathbb{R}^2)$ norm (see, e.g., [4], page 61 for the usual definition). Now, $\wedge^2 L_1(\mathbb{R})$ will be the subspace $\{u \wedge v\colon \|u \wedge v\|_1 < \infty\}$ of $L_1(\mathbb{R}^2)$. For a linear operator $\mathscr{A}$ on $\wedge^2 L_1(\mathbb{R})$ that possesses an integral kernel $A(\cdot, \cdot)$, $\|A\|_{\mathrm{op}}$ will denote the operator norm with respect to $\|\cdot\|_1$. By $T \wedge T$ we denote the integral kernel of the $\wedge^2 L_1(\mathbb{R})$ operator $\mathscr{T} \wedge \mathscr{T}$ defined by $(\mathscr{T} \wedge \mathscr{T})(u \wedge v) = \mathscr{T}u \wedge \mathscr{T}v$. We write $h(x) = O(x)$ as $x \to l$, if h is a *deterministic* function of x and $\limsup_{x \to l} |h(x)|/x < \infty$. By C we denote a nonrandom constant, independent of N_0 and of the spatial and temporal variables (ξ, z, s and t), whose value may change from one line to another.

SKETCH OF THE PROOF OF THEOREM 1. The naive approach of using the asymptotics as $t \to \infty$ of the bounds on the kernel to control the contraction is useless, since the bounds are not tight for large values of t. To take advantage of the tightness for fixed t and small N_0, we rather look at the bounds $T_i(n-1, \xi; n, z)$, $i = 1, 2$, $n = 1, 2, \ldots$ derived in Corollary 1, on each of the kernels $\mathrm{T}(n-1, \xi; n, z)$.

For $s = n - 1$ and $t \ge s$ (usually, we shall have $t \in [n-1, n]$), $k_{i,t}^{(s)}$ and $Q_{i,t}^{(s)}$ are used in conjunction with $\mu_i^0(s;t)$ to define, as explained above, $\mu_i(s, \xi; t)$, $\Psi_i(s, \xi; t)$ and, in turn, also $a_i(s;t)$ and $c_i(s;t)$. The following abbreviations are used:

$$\mathsf{T}_n(\xi,z)=\mathsf{T}(n-1,\xi;n,z),\quad T_{i,n}(\xi,z)=T_i(n-1,\xi;n,z),$$

$$\mu_{i,n}(\xi)=\mu_i(n-1,\xi;n),\qquad \mu^0_{i,n}=\mu^0_i(n-1;n),\qquad \Psi_{i,n}(\xi)=\Psi_i(n-1,\xi;n),$$

$$a_{i,n}=a_i(n-1;n),\qquad c_{i,n}=c_i(n-1;n),$$

$$\Lambda_{i,t}=\Lambda_i(0;t),\qquad \Lambda_i=\Lambda_i(1),\qquad b_{i,t}=b_i(0;t),\qquad b_i=b_i(1).$$

Our estimate is based on the following lemma whose proof is straightforward.

LEMMA 1.

$$|p_t^{(1)}-p_t^{(2)}|_1\le\frac{\|\rho_t^{(1)}\wedge\rho_t^{(2)}\|_1}{|\rho_t^{(1)}|_1|\rho_t^{(2)}|_1}. \tag{2.20}$$

We use the above inequality to bound the exponential decay rate of $|p_t^{(1)}-p_t^{(2)}|_1$ by upper bounding $\|\rho_t^{(1)}\wedge\rho_t^{(2)}\|_1$ and lower bounding $|\rho_t^{(i)}|_1$, using the fact that both $\rho_t^{(1)}\wedge\rho_t^{(2)}$ and $\rho_t^{(i)}$ are governed by linear flows. To bound $\|\rho_t^{(1)}\wedge\rho_t^{(2)}\|_1$ we write

$$\|\rho_n^{(1)}\wedge\rho_n^{(2)}\|_1\le\|\rho_0^{(1)}\wedge\rho_0^{(2)}\|_1\prod_{j=1}^n\|\mathsf{T}_j\wedge\mathsf{T}_j\|_{\mathrm{op}}, \tag{2.21}$$

and develop bounds on $\|\mathsf{T}_n\wedge\mathsf{T}_n\|_{\mathrm{op}}$. To this end, we write T_n as a perturbation of a rank-1 operator. The following lemma shows how to control its norm by controlling the perturbation.

LEMMA 2. *Let $\mathscr{T}$ be a linear operator with integral kernel given by*

$$T(\xi,z)=\psi(\xi)\phi(z)+e(\xi,z).$$

Then the following bound holds

$$\|T\wedge T\|_{\mathrm{op}}\le 2|\psi|_\infty|\phi|_1|e|_{\mathrm{op}}+|e|^2_{\mathrm{op}}.$$

PROOF. This is an immediate consequence of the following identity

$$\begin{aligned}(\mathscr{T}u\wedge\mathscr{T}v)(z,z')&\\ =\int\int\big[&\psi(\xi)\phi(z)e(\xi',z')+\psi(\xi')\phi(z')e(\xi,z)+e(\xi,z)e(\xi',z')\big]\\ &\times(u\wedge v)(\xi,\xi')\,d\xi\,d\xi'.\end{aligned}$$ □

We therefore write

$$\begin{aligned}\mathsf{T}_n(\xi,z)&=\Psi_{1,n}(\xi)\Phi(z)\exp\{-(z-\mu_{1,n}(\xi))^2/2\Lambda_1\}-\varepsilon_n(\xi,z)\\ &=\Psi_{1,n}(\xi)\Phi(z)\exp\{-(z-\beta_n)^2/2\Lambda_1\}+\delta_n(\xi,z)-\varepsilon_n(\xi,z),\end{aligned}$$

where β_n are constants that do not depend on ξ and z (whose values we determine later) and the above equalities define δ_n and ε_n. Corollary 1 asserts that

$$0 \le \varepsilon_n(\xi, z) \le T_{1,n}(\xi, z) - T_{2,n}(\xi, z),$$

while by definition,

$$\delta_n(\xi, z) = \Psi_{1,n}(\xi)\Phi(z)\{\exp\{-(z-\mu_{1,n}(\xi))^2/2\Lambda_1\} - \exp\{-(z-\beta_n)^2/2\Lambda_1\}\}.$$

We define

$$\tilde{\Phi}_\Lambda(u) = \int_{\mathbb{R}} \Phi(z)\exp(-(z-u)^2/2\Lambda)\,dz.$$

Using Lemma 2 with $e_n = \delta_n - \varepsilon_n$, we therefore obtain

$$\|T_n \wedge T_n\|_{\mathrm{op}} \le 2|\Psi_{1,n}|_\infty \tilde{\Phi}_{\Lambda_1}(\beta_n)|e_n|_{\mathrm{op}} + |e_n|^2_{\mathrm{op}}. \tag{2.22}$$

Now, it follows from (2.3) and (2.17) that for all ξ,

$$\begin{aligned} \log\Psi_{1,n}(\xi) &\le -(\xi - a_{1,n})^2/2b_1 + c_{1,n} + \tfrac{1}{2}C_0 + C_0\xi^2 \\ &\le -(\xi - a_{1,n})^2/2b_1 + c_{1,n} + \tfrac{1}{2}C_0 + (C_0+1)\xi^2. \end{aligned} \tag{2.23}$$

While the first inequality will be used below to bound $\Psi_{1,n}(\xi)$, we define γ_n to be the value of ξ maximizing the RHS of the second inequality. The reason for the definition with the extra term ξ^2 is that it makes it possible to conclude later the inequality (5.3). One checks that as $N_0 \to 0$,

$$\gamma_n = a_{1,n}(1 + O(N_0^{1/2})) + O(N_0^{1/2}). \tag{2.24}$$

We now set

$$\beta_n = \mu_{1,n}(\gamma_n).$$

The considerations behind our technique to control $|e_n|_{\mathrm{op}}$ (in Lemma 3 below) are as follows. First, in view of Lemma 5(iii), $a_{1,n}$ can be interpreted as a good estimate for x_{n-1} (for small values of N_0), and in turn, by (2.24), so can γ_n. Also, by Lemma 5(i), one expects that β_n is a good estimate for x_n. We notice, moreover, that because of the nearly Gaussian "shape" of the bounds T_i, both in z (see Corollary 1) and in ξ [see (2.17)], $e_n(\xi, z)$ has the property that most of its mass is concentrated around $z = x_n$, while the value of ξ that maximizes that mass is in the vicinity of x_{n-1}. We therefore estimate $|\varepsilon_n|_{\mathrm{op}} = \sup_\xi \int |\varepsilon_n(\xi, z)|\,dz$ by integrating separately on $\mathscr{B}_1(\beta_n)$ and on its complement, where that mass is negligible, and taking the supremum separately on $\mathscr{B}_1(\gamma_n)$ and on its complement, where the maximum is rarely achieved. In a similar way we estimate $|\delta_n|_{\mathrm{op}}$.

In Lemma 4, a lower bound on $|\rho_n|_1$ is obtained by writing it as an n-fold integral and using the positivity of ρ_n. The ergodic assumptions are exploited in Lemma 5 to obtain tightness of the coefficients appearing in the expressions for the bounds T_i, in an L^p sense, as $N_0 \to 0$.

3. Auxiliary bounds. The purpose of this section is to introduce bounds on the numerator and the denominator of the RHS of (2.20).

LEMMA 3. *Under Assumption 1, one has* $|e_n|_{\mathrm{op}} \le B_n^{(1)} + B_n^{(2)} + B_n^{(3)}$, *where*

$$\begin{aligned}
B_n^{(1)} &= q_n \big|\Psi_{1,n}\big|_\infty O\left(\exp \frac{-N_0^{-1/2}}{2}\right),\\
B_n^{(2)} &= 2 \sup\big\{\Psi_{1,n}(\xi)\big(\tilde\Phi_{\Lambda_1}(\beta_n) + \tilde\Phi_{\Lambda_1}(\mu_{1,n}(\xi))\big)\colon \xi \notin \mathscr{B}_1(\gamma_n)\big\},\\
B_n^{(3)} &= \frac{C|\Phi|_{\mathscr{B}_2(\beta_n)}|_\infty}{\inf\big\{\Phi(\xi)\colon \xi \in \mathscr{B}_1(\gamma_n)\big\}} \exp(c_{1,n}) N_0^{1/4} r_n,\\
q_n &= \big|\Phi|_{\mathscr{B}_2(\beta_n)}\big|_\infty + \sup\big\{\tilde\Phi_{\Lambda_1}(\mu_{1,n}(\xi) + j)\colon \xi \in \mathscr{B}_1(\gamma_n),\ j = \pm 1\big\},\\
r_n &= N_0^{-1/4}|a_{1,n} - a_{2,n}| + |c_{1,n} - c_{2,n}| + O(N_0).
\end{aligned}$$

PROOF. Let us write

$$|\delta_n|_{\mathrm{op}} = \max\big\{\big|\delta_n|_{\mathscr{B}_1(\gamma_n)}\big|_{\mathrm{op}}, \big|\delta_n|_{B_1^c(\gamma_n)}\big|_{\mathrm{op}}\big\}.$$

Note that $|\beta_n - \mu_{1,n}(\xi)| \le 1$ for all $\xi \in \mathscr{B}_1(\gamma_n)$. Let us denote the indicator function by $1\{\cdot\}$. We then have

$$\exp\left(\frac{-z^2}{2\Lambda}\right) 1\big\{|z| > 1\big\} \le \exp\left(\frac{-1}{2\Lambda}\right)\left\{\exp \frac{-(z-1)^2}{2\Lambda} + \exp \frac{-(z+1)^2}{2\Lambda}\right\}.$$

Therefore we have

$$\big|\delta_n|_{\mathscr{B}_1(\gamma_n)}\big|_{\mathrm{op}} \le |\Psi_{1,n}|_\infty \big|\Phi|_{\mathscr{B}_2(\beta_n)}\big|_\infty A_n^{(1)} + |\Psi_{1,n}|_\infty \exp\left(\frac{-1}{2\Lambda_1}\right) A_n^{(2)},$$

where

$$A_n^{(1)} = \sup_{\xi \in \mathscr{B}_1(\gamma_n)} \int \left|\exp \frac{-(z - \mu_{1,n}(\xi))^2}{2\Lambda_1} - \exp \frac{-(z-\beta_n)^2}{2\Lambda_1}\right| dz$$

and

$$A_n^{(2)} = 2 \sup\big\{\tilde\Phi_{\Lambda_1}(\mu_{1,n}(\xi) + j)\colon \xi \in \mathscr{B}_1(\gamma_n),\ j = \pm 1\big\}.$$

Denoting by F the standard normal distribution on $\mathbb{R}$, and using the fact that $\mu_{1,n}(\xi)$ is a nondecreasing function of ξ, we have

$$A_n^{(1)} = \sqrt{2\pi\Lambda_1} \max_{j=\pm 1}\left\{-2 + 4F\left(\frac{|\mu_{1,n}(\gamma_n + j) - \beta_n|}{2\sqrt{\Lambda_1}}\right)\right\} \le \frac{2}{\cosh(\alpha_1)}.$$

We have therefore shown that $\left|\delta_n|_{\mathscr{B}_1(\gamma_n)}\right|_{\mathrm{op}} \le B_n^{(1)}/2$. On the other hand, it is straightforward to show that $\left|\delta_n|_{B_1^c(\gamma_n)}\right|_{\mathrm{op}} \le B_n^{(2)}/2$. Similarly,

$$\left|\varepsilon_n|_{\mathscr{B}_1(\gamma_n)}\right|_{\mathrm{op}} \le \left|\Phi|_{\mathscr{B}_2(\beta_n)}\right|_\infty A_n^{(3)} + |\Psi_{1,n}|_\infty \exp\left(\frac{-1}{2\Lambda_1}\right)\frac{A_n^{(2)}}{2},$$

where

$$A_n^{(3)} = \sup\left\{\sqrt{2\pi\Lambda_1}\Psi_{1,n}(\xi) - \sqrt{2\pi\Lambda_2}\Psi_{2,n}(\xi)\colon \xi \in \mathscr{B}_1(\gamma_n)\right\}.$$

One has by (2.17) that

$$A_n^{(3)} \le \frac{C}{\inf\{\Phi(\xi)\colon \xi\in\mathscr{B}_1(\gamma_n)\}} \sup_{\xi\in\mathbb{R}}\left\{\sqrt{\Lambda_1}\exp\left(\frac{-(\xi-a_{1,n})^2}{2b_1}+c_{1,n}\right)\right.$$
$$\left. - \sqrt{\Lambda_2}\exp\left(\frac{-(\xi-a_{2,n})^2}{2b_2}+c_{2,n}\right)\right\}.$$

Now, bounding the partial derivatives of $\exp(-(\xi-a)^2/2b+c)$ with respect to a, b and c, and using the fact that both b_i and Λ_i are $O(N_0^{1/2})$, one obtains that for all N_0 small enough,

$$A_n^{(3)} \le \frac{C}{\inf\{\Phi(\xi)\colon \xi\in\mathscr{B}_1(\gamma_n)\}}\left\{|a_{1,n}-a_{2,n}|\exp(c_{1,n}) + \frac{b_1-b_2}{\sqrt{b_2}}\exp(c_{1,n})\right.$$
$$\left. + \sqrt{\Lambda_1}\exp(c_{1,n})\left|c_{1,n}-c_{2,n}+\frac{1}{2}\log\frac{\Lambda_1}{\Lambda_2}\right|\right\}$$

and hence

$$\left|\varepsilon_n|_{\mathscr{B}_1(\gamma_n)}\right|_{\mathrm{op}} \le B_n^{(3)} + B_n^{(1)}/2.$$

Finally, one has that $\left|\varepsilon_n|_{B_1^c(\gamma_n)}\right|_{\mathrm{op}} \le B_n^{(2)}/2$, and the lemma follows. $\square$

We proceed with a lower bound on $|\rho_n^{(i)}|_1$. Fix i (either $i=1$ or $i=2$) and let D_i be any fixed compact set on which $\rho_0^{(i)} = p_0^{(i)}$ is positive and such that $|D_i| > 0$. For $z_0 \in D_i$ let $m_0(z_0) = z_0$ and for $n = 1, 2, \dots$ let

$$m_n(z_0) = \mu_{2,n}(m_{n-1}(z_0)).$$

Then we have the following.

LEMMA 4. *Let Assumption 1 hold and let there exist some constant $C_2 > -\infty$ independent of N_0 and of z_0, such that the following holds:*

$$\liminf_{n\to\infty}\frac{1}{n}\log\inf\left\{\Phi(z)\colon \left|z-m_n(z_0)\right| < \sqrt{\Lambda_2}\right\} \ge C_2. \tag{3.1}$$

Then

$$
\begin{aligned}
&\liminf_{n\to\infty} \frac{1}{n} \log\bigl|\rho_n^{(i)}\bigr|_1 \\
&\quad \geq C + \frac{1}{2}\log\Lambda_2 \\
&\qquad + \liminf_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n-1}\biggl\{c_{2,\,j+1} - N_0^{-1/2}\biggl(\biggl|\sum_{k=2}^{j-1}\cosh^{k-j}(\alpha_2)\bigl(\mu_{2,\,k}(\xi_0)-\xi_0\bigr)\biggr| \\
&\qquad\qquad + \bigl|\mu_{2,\,j}(\xi_0) - a_{2,\,j+1}\bigr|\biggr)^2\biggr\},
\end{aligned}
$$

where by convention, $\sum_m^n = 0$ *if* $n < m$.

PROOF. By repeated application of the lower bound we have

$$
\begin{aligned}
\bigl|\rho_n^{(i)}\bigr|_1 &= \int\cdots\int \rho_0^{(i)}(z_0)\prod_{j=1}^{n}\mathsf{T}_n(z_{j-1}, z_j)\,dz_0\cdots dz_n \\
&\geq \int\cdots\int \rho_0^{(i)}(z_0)1\{z_0\in D_i\}\prod_{j=1}^{n}\Psi_{2,\,j}(z_{j-1})\Phi(z_j) \\
&\quad\times\exp\bigl(-(z_j-\mu_{2,\,j}(z_{j-1}))^2/2\Lambda_2\bigr) \\
&\quad\times 1\bigl\{|z_j - m_j(z_0)| < \sqrt{\Lambda_2}\bigr\}\,dz_0\cdots dz_n \\
&\geq \inf_{z_0\in D_i}\rho_0^{(i)}(z_0)\inf_{z_0\in D_i}\Psi_{2,1}(z_0)\inf\bigl\{\Phi(z)\colon \bigl|z - m_n(z_0)\bigr| < \sqrt{\Lambda_2}\bigr\} \qquad (3.2)\\
&\quad\times\inf_{z_0\in D_i}\prod_{j=2}^{n}\inf\bigl\{\Psi_{2,\,j}(z)\Phi(z)\colon |z - m_{j-1}(z_0)| < \sqrt{\Lambda_2}\bigr\} \\
&\quad\times\int\cdots\int 1\{z_0\in D_i\}\prod_{j=1}^{n}1\bigl\{|z_j - m_j(z_0)| < \sqrt{\Lambda_2}\bigr\} \\
&\quad\times\exp\bigl(-(z_j-\mu_{2,\,j}(z_{j-1}))^2/2\Lambda_2\bigr)\,dz_0\cdots dz_n.
\end{aligned}
$$

Now by (2.13) we know that $0 \leq d\mu_{2,\,j}(\xi)/d\xi \leq 1$. Therefore, within the integration region, the following holds:

$$
\begin{aligned}
\bigl|z_j - \mu_{2,\,j}(z_{j-1})\bigr| &\leq \bigl|z_j - m_j(z_0)\bigr| + \bigl|m_j(z_0) - \mu_{2,\,j}(m_{j-1}(z_0))\bigr| \\
&\quad + \bigl|\mu_{2,\,j}\bigl(m_{j-1}(z_0)\bigr) - \mu_{2,\,j}\bigl(z_{j-1}\bigr)\bigr| \leq 2\sqrt{\Lambda_2}.
\end{aligned}
\tag{3.3}
$$

Hence the multiple integral in (3.2) is greater than or equal to $|D_i|(2\sqrt{\Lambda}e^{-2})^n$, and we obtain

$$\begin{aligned}\frac{1}{n}\log|\rho_n^{(i)}|_1 \geq \frac{1}{n}\log\inf_{D_i}\rho_0^{(i)}(z_0) + \frac{1}{n}\log\inf_{D_i}\Psi_1(z_0)\\ + \frac{1}{n}\log\,\inf\{\Phi(z)\colon |z-m_n(z_0)| < \sqrt{\Lambda_2}\}\\ + \frac{1}{n}\log|D_i| + \log 2 - 2 + \frac{1}{2}\log\Lambda_2 + \frac{1}{n}\sum_{j=2}^{n}\big(A_j^{(4)} + c_{2,j}\big),\end{aligned}$$

where

$$A_n^{(4)} = \inf_{z_0\in D_i}\inf\{\log\Psi_{2,n}(z)\Phi(z)\colon |z-m_{n-1}(z_0)| < \sqrt{\Lambda_2}\} - c_{2,n}.$$

The lemma now follows from the fact that

$$A_n^{(4)} \geq -\sup_{z_0\in D_i}\big(1 + N_0^{-1/2}\big(m_{n-1}(z_0) - a_{2,n}\big)^2\big)$$

and

$$\big|m_{n-1}(z_0) - a_{2,n}\big| \leq \big|m_{n-1}(z_0) - \mu^0_{2,n-1}\big| + \big|\mu^0_{2,n-1} - a_{2,n}\big|,$$

while

$$\begin{aligned}m_{n-1}(z_0) - \mu^0_{2,n-1} = \sum_{k=1}^{n-2}\cosh^{k-n+1}(\alpha_2)\big(\mu^0_{2,k} - \xi_0\big)\\ + \cosh^{1-n}(\alpha_2)(z_0 - \xi_0).\end{aligned} \tag{3.4}$$

□

4. Ergodic behavior. Stationarity and ergodicity of the discrete-time processes $\mu^0_{i,n}, n = 1, 2, \ldots$ under $\tilde{P}$ are implied by Assumption 2, since $\mu^0_{i,n}$ is determined by the paths $\{(x_t, \nu_t - \nu_{n-1}), n-1 \leq t \leq n\}$ alone. The same is true for the processes $a_{i,n}$, $c_{i,n}$, $K_{i,n}$ and $|\Psi_{i,n}|_\infty$.

Some of the consequences of Assumptions 2 and 3 are summarized below.

LEMMA 5. *For* $0 < \varepsilon < 1/2$ *fixed, as* $N_0 \to 0^+$ *one has*

(i) $$\tilde{E}\big(\mu_i^0(n; n+t) - x_{n+t}\big)^2 = \begin{cases} O(1), & \textit{uniformly in } t \in [0, N_0^{1/2-\varepsilon}],\\ O(N_0^{1/2}), & \textit{uniformly in } t \in [N_0^{1/2-\varepsilon}, 1],\end{cases}$$

(ii) $$\begin{aligned}&\tilde{E}\big(\mu_1^0(n; n+t) - \mu_2^0(n; n+t)\big)^2\\ &\quad= \begin{cases} O(N_0^2), & \textit{uniformly in } t \in [0, N_0^{1/2-\varepsilon}],\\ O(N_0^{5/2}), & \textit{uniformly in } t \in [N_0^{1/2-\varepsilon}, 1],\end{cases}\end{aligned}$$

(iii) $\tilde{E}(a_{i,n} - x_{n-1})^2 = O\big(N_0^{1/2}\big)$,

(iv) $\tilde{E}(a_{1,n} - a_{2,n})^2 = O(N_0^2)$,

(v) $\tilde{E}\log\dfrac{K_{1,n}}{K_{2,n}} = O(N_0^{1/4})$,

(vi) $\tilde{E}\big|c_{1,n} - c_{2,n}\big| = O(N_0^{1/4})$,

(vii) $\tilde{E}(\log|\Psi_{1,n}|_\infty - c_{1,n}) = O(1)$.

REMARK. In [21], analogues of (i), (ii) and (v) of the lemma have been shown for convergence in an a.s. sense.

PROOF. (i) Substituting (2.2) in (2.13), we can write

$$\mu_i^0(0;t) - x_t = \frac{1}{\cosh(\alpha_i t)}\biggl(\xi_0 - \int_0^t \frac{\sinh(\alpha_i s)}{\alpha_i} Q_{i,s}\,ds + \int_0^t \frac{\sinh(\alpha_i s)}{N_0\alpha_i}(x_s - u x_t)\,ds\biggr) + V_{i,t},$$

where

$$u = \frac{N_0\alpha_i\cosh(\alpha_i t)}{\int_0^t \sinh(\alpha_i s)\,ds} = 1 + O(N_0)$$

uniformly in $t \in [N_0^{1/2-\varepsilon}, 1]$ and

$$V_{i,t} = \frac{1}{\cosh(\alpha_i t)}\int_0^t \frac{\sinh(\alpha_i s)}{N_0^{1/2}\alpha_i}\,d\nu_s.$$

We have $EV_{i,t}^2 = O(N_0^{1/2})$ uniformly in $t\in[0,1]$. Therefore, using Minkowski's inequality and the inequality $E(\int_0^t Z_s\,ds)^2 \le (\int_0^t \sqrt{EZ_s^2}\,ds)^2$, we have

$$\begin{aligned}\tilde{E}\bigl(\mu_i^0(0;t) - x_t\bigr)^2 &\le CN_0\biggl(\int_0^t\sqrt{\tilde{E}Q_{i,s}^2}\,ds\biggr)^2 \\ &\quad + CN_0^{-1}\left(\frac{\int_0^t \sinh(\alpha_i s)\sqrt{\tilde{E}(x_t - x_s)^2}\,ds}{\cosh(\alpha_i t)}\right)^2 + O\bigl(N_0^{1/2}\bigr)\end{aligned} \tag{4.1}$$

uniformly in $t \in [N_0^{1/2-\varepsilon}, 1]$. Now, since f grows at most linearly [cf. (2.3)], it follows from (2.1) that $\tilde{E}(x_t - x_s)^2 \le C|t-s|$. Since we also have

$$\frac{\int_0^t \sinh(\alpha_i s)(t-s)^{1/2}\,ds}{\cosh(\alpha_i t)} = O\bigl(N_0^{3/4}\bigr)$$

uniformly in $t \in [0,1]$, it follows that the second term on the RHS of (4.1) is $O(N_0^{1/2})$ uniformly in $t \in [0,1]$. As for the first term on the RHS, recall first, that x_t has finite second moment and that $Q_{i,t}$ is at most linear in $\mu_i^0(0;t)$. Denoting $\bar{\mu}_{i,t} = \tilde{E}(\mu_i^0(0;t))^2$, we have just shown that $\bar{\mu}_{i,t} \le C + CN_0(\int_0^t \bar{\mu}_{i,s}^{1/2}\,ds)^2 \le C + CN_0 t\int_0^t \bar{\mu}_{i,s}\,ds$. By Gronwall's lemma, $\bar{\mu}_{i,t}$ is therefore finite, and in fact, bounded uniformly in $t \in [0,1]$ and $N_0 \in (0,1]$. Therefore, (i) is deduced from (4.1).

(ii) In [21], equation (3.7), an expression for $\mu_1^0(0;t) - \mu_2^0(0;t)$ is obtained. Namely, let

$$\lambda_t = \frac{\Lambda_1(t)}{N_0} + V\Lambda_2(t)$$

and $\tilde{\xi}_t = \Lambda_1(t) - \Lambda_2(t)$. Then

$$\begin{aligned}
&\mu_1^0(0;t) - \mu_2^0(0;t) \\
&\quad = \int_0^t \exp\biggl(-\int_s^t \lambda_\theta \, d\theta\biggr)\biggl(V'\bigl(\mu_1^0(0;s)\bigr)\frac{\tilde{\xi}_s}{N_0^{1/2}} + \bigl(x_s - \mu_2^0(0;s)\bigr)\frac{\tilde{\xi}_s}{N_0}\biggr)\,ds \\
&\qquad + \int_0^t \exp\biggl(-\int_s^t \lambda_\theta \, d\theta\biggr)\frac{\tilde{\xi}_s d\nu_s}{N_0^{1/2}}.
\end{aligned}$$

One checks that $\tilde{\xi}_t = O(N_0^{3/2})$ uniformly in $t \in [0,1]$ (note that we have $\Lambda_i(0) = 0$ rather than > 0 as in [21]) and that

$$\int_0^t \exp\biggl(-\int_s^t \lambda_\theta \, d\theta\biggr)\,ds = O(N_0^{1/2})$$

uniformly in $t \in [0,1]$, as well as

$$\int_0^t \exp\biggl(-2\int_s^t \lambda_\theta \, d\theta\biggr)\,ds = O(N_0^{1/2})$$

uniformly in $t \in [0,1]$. We obtain

$$\begin{aligned}
\tilde{E}\bigl(\mu_1^0(0;t) - \mu_2^0(0;t)\bigr)^2 &= O(N_0)\biggl(\int_0^t \exp\biggl(-\int_s^t \lambda_\theta \, d\theta\biggr)\sqrt{\tilde{E}\bigl(x_s - \mu_2^0(0;s)\bigr)^2}\,ds\biggr)^2 \\
&\quad + O(N_0^{5/2}).
\end{aligned}$$

Noting also that for $s < N_0^{1/2-\varepsilon/2}$ and $t > N_0^{1/2-\varepsilon}$, one has $\int_s^t \lambda_\theta \, d\theta > N_0^{-\varepsilon}/2$, the result follows from (i).

(iii) By (2.18) and (2.2) we have

$$\begin{aligned}
a_{i,1} - x_0 &= \xi_0 - x_0 + b_i \int_0^1 \frac{N_0^{-1} x_s - \alpha_i^2 \mu_i^0(0;s) - Q_{i,s}}{\cosh(\alpha_i s)}\,ds \\
&\quad + \frac{b_i}{N_0^{1/2}} \int_0^1 \frac{d\nu_s}{\cosh(\alpha_i s)} \qquad (4.2) \\
&= W_1 + W_2,
\end{aligned}$$

where

$$W_1 = \frac{b_i}{N_0}\int_0^1 \frac{x_s}{\cosh(\alpha_i s)}\,ds - b_i\alpha_i^2\int_0^1 \frac{\mu_i^0(0;s) - \xi_0/\cosh(\alpha_i s)}{\cosh(\alpha_i s)}\,ds - x_0,$$

$$W_2 = -b_i\int_0^1 \frac{Q_{i,s}}{\cosh(\alpha_i s)}\,ds + \frac{b_i}{N_0^{1/2}}\int_0^1 \frac{d\nu_s}{\cosh(\alpha_i s)}.$$

We have $\tilde{E}W_2^2 = O(N_0^{1/2})$. On the other hand, by (2.13) we have

$$W_1 = \frac{b_i}{N_0}\int_0^1 \frac{x_s - x_0}{\cosh(\alpha_i s)}\,ds - \frac{b_i\alpha_i}{N_0}\int_0^1 \frac{\int_0^s \sinh(\alpha_i\theta)(x_\theta - x_0)\,d\theta}{\cosh(\alpha_i s)^2}\,ds + W_3,$$

where

$$W_3 = \frac{P_i}{\alpha_i^2}x_0 - \frac{1}{\tanh(\alpha_i)}\int_0^1 \frac{\int_0^s \sinh(\alpha_i\theta)\bigl(-Q_{i,\theta}\,d\theta + N_0^{-1/2}\,d\nu_\theta\bigr)}{\cosh(\alpha_i s)^2}\,ds,$$

and one checks that $\tilde{E}W_3^2 = O(N_0)$. By considerations similar to those in the proof of (i) we then obtain $\tilde{E}W_1^2 = O(N_0^{1/2})$, which proves (iii).

(iv) In order to use (4.2) let us first write

$$a_{1,1} - a_{2,1} = (b_1 - b_2)\frac{a_{1,1} - \xi_0}{b_1} + b_2\left(\frac{a_{1,1} - \xi_0}{b_1} - \frac{a_{2,1} - \xi_0}{b_2}\right).$$

Noting that $b_1 - b_2 = O(N_0^{3/2})$ and using (iii), we have that the second moment of the first term is $O(N_0^2)$. The second term is equal to $b_2(W_4 + W_5 + W_6)$ where

$$\begin{aligned} W_4 &= \int_0^1 \left(\frac{1}{\cosh(\alpha_2 s)} - \frac{1}{\cosh(\alpha_1 s)}\right) \\ &\qquad \times \left(N_0^{-1}\bigl(x_s - \mu_1^0(0;s)\bigr) - \frac{1}{2}V'\bigl(\mu_1^0(0;s)\bigr)\right)ds, \\ W_5 &= \int_0^1 \frac{1}{\cosh(\alpha_2 s)}(N_0^{-1} - V)\bigl(\mu_2^0(0;s) - \mu_1^0(0;s))\bigr)\,ds, \\ W_6 &= N_0^{-1/2}\int_0^1 \left(\frac{1}{\cosh(\alpha_2 s)} - \frac{1}{\cosh(\alpha_1 s)}\right)d\nu_s. \end{aligned}$$

One checks that

$$\int_0^1 \left(\frac{1}{\cosh(\alpha_2 s)} - \frac{1}{\cosh(\alpha_1 s)}\right)ds = O(N_0^{3/2}),$$

while

$$\int_0^1 \left(\frac{1}{\cosh(\alpha_2 s)} - \frac{1}{\cosh(\alpha_1 s)}\right)^2 ds = O\bigl(N_0^{5/2}\bigr).$$

Hence it follows from (i) that $\tilde{E}W_4^2 = O(N_0)$, from (ii) that $\tilde{E}W_5^2 = O(N_0)$ and directly that $\tilde{E}W_6^2 = O(N_0^{3/2})$. As a result, (iv) follows.

(v) Note that by definition, $K_{i,n} = \Psi_{i,n}(\xi_0)\Phi(\xi_0)$. Therefore, by (2.14),

$$\begin{aligned}\log\frac{K_{1,1}}{K_{2,1}} &= -\frac{1}{2}\log\frac{\sinh(\alpha_1)}{\sinh(\alpha_2)} + \int_0^1\biggl(\frac{N_0^{-1}+V}{2}\bigl(\mu_2^0(0;s)-\mu_1^0(0;s)\bigr)^2 \\ &\qquad + \frac{1}{2}V'\bigl(\mu_1^0(0;s)\bigr)\bigl(\mu_2^0(0;s)-\mu_1^0(0;s)\bigr)\biggr)\,ds + W_7,\end{aligned} \tag{4.3}$$

where

$$\begin{aligned}W_7 &= \int_0^1 N_0^{-1}\bigl(\mu_1^0(0;s)-\mu_2^0(0;s)\bigr)\circ dy_s \\ &= \frac{1}{N_0}\int_0^1\biggl(\bigl(\mu_1^0(0;s)-\mu_2^0(0;s)\bigr)x_s + \frac{1}{2}\bigl(\Lambda_1(s)-\Lambda_2(s)\bigr)\biggr)\,ds \\ &\quad + \frac{1}{\sqrt{N_0}}\int_0^1\bigl(\mu_1^0(0;s)-\mu_2^0(0;s)\bigr)\,d\nu_s.\end{aligned}$$

The first term in (4.3) is $O(N_0^{1/2})$. From (ii), expectation of the second term is $O(N_0^{5/4})$, while $\tilde{E}W_7 = O(N_0^{1/4})$. Therefore (v) follows.

(vi) Note that $K_{i,n} = \exp(-(\xi_0 - a_{i,n})^2/2b_i + c_{i,n})$. Therefore

$$c_{1,n} - c_{2,n} = \log\frac{K_{1,n}}{K_{2,n}} + \frac{(\xi_0 - a_{1,n})^2}{2b_1} - \frac{(\xi_0 - a_{2,n})^2}{2b_2}.$$

By (v) and since $b_1 > b_2$, we have

$$\tilde{E}\bigl|c_{1,n} - c_{2,n}\bigr| \le O(N_0^{1/4}) + \frac{1}{2b_2}\tilde{E}^{1/2}\bigl(2\xi_0 - a_{1,n} - a_{2,n}\bigr)^2\tilde{E}^{1/2}\bigl(a_{1,n} - a_{2,n}\bigr)^2,$$

and (vi) follows from (iv).

(vii) This is directly implied by (2.23) and (2.24). □

5. Proof of Theorem 1. Let us return to Lemma 3 and denote

$$\bar{\Phi}_n = \max\bigl\{q_n, \tilde{\Phi}_{\Lambda_1}(\beta_n) + \exp\bigl((\mu_{1,n}^0)^2\bigr)\bigr\}.$$

With this notation we have that

$$B_n^{(1)} \le |\Psi_{1,n}|_\infty \bar{\Phi}_n O\bigl(\exp(-N_0^{-1/2}/2)\bigr).$$

Moreover, using the bound

$$|\log\Phi(x)| \le C_0(|x| + \tfrac{1}{2}x^2) \le C_0(\tfrac{1}{2} + x^2), \tag{5.1}$$

implied by (2.3), we have that

$$\tilde{\Phi}_\Lambda(x) \le \Lambda^{1/2}\exp(C + C\Lambda x^2), \tag{5.2}$$

and as a consequence,

$$B_n^{(2)} \le C\sup_{\xi\notin\mathscr{B}_1(\gamma_n)}\Psi_{1,n}(\xi)\bigl\{\tilde{\Phi}_{\Lambda_1}(\beta_n) + \exp(C\Lambda_1(\mu_{1,n}(\xi))^2)\bigr\}.$$

From (2.16), one has that $C\Lambda_1(\mu_{1,n}(\xi))^2 \le (\mu_{1,n}^0)^2 + \xi^2 + \xi_0^2$ if N_0 is small enough. Hence, from the definition of γ_n, based on (2.23), we obtain

$$\Psi_{1,n}(\xi)\exp(\xi^2) \le \exp(-\tilde{\alpha}_{1,n}(\xi-\gamma_n)^2 + \tilde{\alpha}_{2,n}), \tag{5.3}$$

where $\tilde{\alpha}_{1,n}$, $\tilde{\alpha}_{2,n}$ are defined by the identity $-\tilde{\alpha}_{1,n}(\xi-\gamma_n)^2 + \tilde{\alpha}_{2,n} = -(\xi - \mu_{1,n})^2/2b_1 + c_{1,n} + C_0/2 + (C_0+1)\xi^2$. Solving for $\tilde{\alpha}_{1,n}, \tilde{\alpha}_{2,n}$, and using also (2.24) and the fact that $b_1 > (1-\varepsilon)N_0^{1/2}$, one obtains

$$B_n^{(2)} \le \exp\bigl(C\gamma_n^2 + c_{1,n}\bigr)\bar{\Phi}_n O\bigl(\exp(-N_0^{-1/2}/2)\bigr).$$

Since $|\Psi_{1,n}|_\infty \ge \Psi_{1,n}(\gamma_n)$ we have $\exp(c_{1,n}) \le \exp((a_{1,n}-\gamma_n)^2/2b_1 + C + C\gamma_n^2)|\Psi_{1,n}|_\infty$, and using again (2.24) we have

$$B_n^{(2)} \le \exp(C\gamma_n^2)|\Psi_{1,n}|_\infty \bar{\Phi}_n O\bigl(\exp(-N_0^{-1/2}/2)\bigr).$$

Similarly,

$$B_n^{(3)} \le C\exp(C\gamma_n^2)|\Psi_{1,n}|_\infty \bar{\Phi}_n N_0^{1/4} r_n,$$

hence

$$|e_n|_{\mathrm{op}} \le CN_0^{1/4}\exp(C\gamma_n^2)|\Psi_{1,n}|_\infty \bar{\Phi}_n r_n.$$

From the inequalities (2.20), (2.21) and (2.22) we therefore conclude that

$$\begin{aligned} &\bigl|p_n^{(1)} - p_n^{(2)}\bigr|_1 \\ &\quad \le \frac{\prod_{j=1}^n CN_0^{1/2}|\Psi_{1,j}|_\infty^2(\bar{\Phi}_j + \bar{\Phi}_j^2)\exp\bigl(C(\beta_j^2+\gamma_j^2)\bigr)(r_j + r_j^2)}{|\rho_n^{(1)}|_1|\rho_n^{(2)}|_1} \\ &\qquad \times \bigl\|\rho_0^{(1)} \wedge \rho_0^{(2)}\bigr\|_1. \end{aligned} \tag{5.4}$$

Now, condition (3.1) of Lemma 4 holds P-a.s. by the following argument. By (5.1), and since $\Lambda_2 < 1$ for N_0 small enough, it suffices to show that

$$\limsup_{n\to\infty} n^{-1}m_n^2(z_0) \le C, \tag{5.5}$$

where $C > 0$ is some constant independent of N_0 and of z_0. It follows from (3.4) that for n large enough,

$$m_n^2(z_0) \le 2(\mu_{2,n}^0)^2 + 2s_n^2 + 1,$$

where

$$s_n = \sum_{k=1}^{n-1}\cosh^{k-n}(\alpha_2)\bigl|\mu_{2,k}^0 - \xi_0\bigr|.$$

Note that $\{\mu_{2,n}^0, n = 1, 2, \ldots\}$ is a square integrable stationary process, and therefore, by the last display, $P(2(\mu_{2,n}^0)^2 + 2s_n^2 + 1 > Cn) \le \lim_{k\to\infty} P(2(\mu_{2,k}^0)^2 + 2s_k^2 + 1 > Cn)$ is summable. It follows that $P(m_n^2(z_0) > Cn)$ is summable, and (5.5) follows from the Borel–Cantelli lemma.

We now refer to the conclusion of Lemma 4 and note that the means can be replaced by the expectations, using Birkhoff's ergodic theorem. The second moment of the expression

$$\sum_{k=2}^{j-1} \cosh^{k-j}(\alpha_2)\big(\mu_{2,k}(\xi_0) - \xi_0\big)$$

is $O(\exp(-N_0^{-1/2}))$ for $j = 1, 2, \ldots$ (in fact it is zero for $j = 1, 2$). Also, by Lemma 5(i), (iii), the second moment of $(\mu_{2,j} - x_j) + (x_j - a_{2,j+1})$ is $O(N_0^{1/2})$. It follows that for N_0 small enough, one has P-a.s.,

$$\liminf_{n\to\infty} \frac{1}{n} \log\big\{|\rho_n^{(1)}|_1 |\rho_n^{(2)}|_1\big\} \geq \log \Lambda_2 + 2\tilde{E}c_{2,1} + C. \tag{5.6}$$

Let us now return to the inequality (5.4). Note that the factor $N_0^{1/2}$ cancels out with the term $\log \Lambda_2$ in the lower bound of the denominator (up to a constant); moreover, we show below that the factor $|\Psi_{1,j}|^2$ balances with the term $2\tilde{E}c_{2,1}$, the terms $\bar{\Phi}_j$, β_j^2 and γ_j^2 have no effect, while the terms r_j decay rapidly enough to cause stability. Indeed, by Lemma 5 (vi), (vii) we have that $\tilde{E}(2\log|\Psi_{1,1}|_\infty - 2c_{2,1}) = O(1)$. Combining the expressions in (5.4) and (5.6) we have that P-a.s.,

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n} \log\big|p_n^{(1)} - p_n^{(2)}\big|_1 \leq{}& \frac{1}{2}\log N_0 - \log\Lambda_2 + C\tilde{E}(\beta_1^2 + \gamma_1^2) \\ &+ \tilde{E}\log^+\big(\bar{\Phi}_1 + \bar{\Phi}_j^2\big) + \tilde{E}\log\big(r_1 + r_1^2\big) + C.\end{aligned}$$

Furthermore, $\tilde{E}\log^+(\bar{\Phi}_1 + \bar{\Phi}_1^2)$ is bounded as $N_0 \to 0$ by the following argument. If Z is a random variable satisfying $\tilde{E}Z^2 < \infty$, then by (5.1), $\tilde{E}\log^+\big|\Phi|_{\mathscr{B}_2(Z)}\big|_\infty < \infty$. Moreover, by (5.2), $\sup_{0<\Lambda<\Lambda_0} \tilde{E}\log^+\big|\tilde{\Phi}_\Lambda|_{\mathscr{B}_2(Z)}\big|_\infty < \infty$ for Λ_0 small enough. Indeed, by Lemma 5(iii), β_n is square integrable. Next, since we have by Jensen's inequality and Lemma 5(iv), (vi) that $\tilde{E}\log(1 + r_n)$ is bounded while $\tilde{E}\log r_n \leq \log CN_0^{1/4}$, we have shown that P-a.s.,

$$\limsup_{n\to\infty} \frac{1}{n}\log\big|p_n^{(1)} - p_n^{(2)}\big|_1 \leq \frac{1}{4}\log N_0 + C.$$

To carry on with the proof for continuous parameter, note first that the law of

$$\sup\big\{\log|\mathrm{T}(s,\cdot;t,\cdot)|_{\mathrm{op}}\colon s \text{ and } t \text{ for which } n \leq s < s+1 \leq t \leq n+2\big\}$$

under $\tilde{P}$ does not depend on $n = 0, 1, \ldots$. We next prove the following.

LEMMA 6. *One has*

$$\tilde{E}\sup\big\{\log|\mathrm{T}(s,\cdot;t,\cdot)|_{\mathrm{op}}\colon 0 \leq s < s+1 \leq t \leq 2\big\} < \infty.$$

PROOF. Using the upper bound we have

$$\left|\mathsf{T}(s,\cdot;t,\cdot)\right|_{\mathrm{op}} \le \sup_{\xi} \Psi_1(s,\ \xi;\ t)\ \sup_{\xi} \tilde{\Phi}_{\Lambda_1(s;t)}(\mu_1(s,\xi;t)).$$

Square integrability of $\mu_1^0(s;t)$, uniformly in $0 \le s < s+1 \le t \le 2$, is obtained by an argument similar to that in the proof of Lemma 5(i). From this, using (2.17), (2.18) and (2.19), it follows that $\tilde{E} \log \sup_{\xi} \Psi_1(s,\xi;t) < \infty$, and from (5.2), that $\tilde{E} \log \sup_{\xi} \tilde{\Phi}_{\Lambda_1(s;t)}(\mu_1(s,\xi;t)) < \infty$. $\square$

Let now $n = n(t) = [t]$. Since for any $L_1(\mathbb{R})$ operator with kernel $\mathsf{T}(\cdot,\cdot)$ we have $\|\mathsf{T} \wedge \mathsf{T}\|_{\mathrm{op}} \le |\mathsf{T}|_{\mathrm{op}}^2$, it follows from (2.20) that $|p_t^{(1)} - p_t^{(2)}|_1 \le z_n \tau_n$, where

$$z_n = \frac{\left\|\rho_{n-1}^{(1)} \wedge \rho_{n-1}^{(2)}\right\|_1}{\left|\rho_{n+2}^{(1)}\right|_1 \left|\rho_{n+2}^{(2)}\right|_1}$$

and

$$\tau_n = \sup_{t \in [n-1,\,n]} \left|\mathsf{T}(n-1,\cdot;t,\cdot)\right|_{\mathrm{op}}^2 \left|\mathsf{T}(t,\cdot;n+2,\cdot)\right|_{\mathrm{op}}^2.$$

It follows from the proof above that P-a.s., $\limsup_{n\to\infty}(1/n)\log z_n \le \frac{1}{4}\log N_0 + C$ (e.g., by replacing the LHS in (5.6) with $\liminf_{n\to\infty}(1/n)\log(|\rho_{n+3}^{(1)}|_1|\rho_{n+3}^{(2)}|_1)$). Since it follows from Lemma 6 that $\limsup_{t\to\infty}(1/t)\log \tau_n = 0$ P-a.s., the theorem follows. $\square$

APPENDIX

Proof of Corollary 1. As is the case in [21], looking at the expression (2.4), using the Girsanov theorem to write $dP^{(t)}/dP_0^{(t)}$ explicitly and then using (2.5), one obtains that the inequalities (2.15) hold true with T_i replaced by $\tilde{T}_i$, that are defined by

$$\tilde{T}_i(s,\xi;t,z) = \frac{\Phi(z)}{\Phi(\xi)} E_0\big(L_{i,s,t}\big|\mathscr{Y}_{s,t},\, x_s = \xi,\, x_t = z\big)\, p_{x_t\,|\,x_s}(z\,|\,\xi)$$

and

$$L_{i,s,t} = \exp\left\{\int_s^t \left(-\frac{P_{i,\theta}x_\theta^2}{2} - Q_{i,\theta}x_\theta - k_{i,\theta} - \frac{x_\theta^2}{2N_0}\right) d\theta + \int_s^t \frac{x_\theta\, dy_\theta}{N_0}\right\}.$$

Consider now the weak solutions to the corresponding backward filtering equation (see, e.g., [17], Section 6.3), that is, functions $v_t(z)$, satisfying for every $\varphi \in C_0^\infty(\mathbb{R})$, P-a.s. [denoting $' = \partial/\partial z$ and $I(z) \equiv z$],

$$(v_t,\varphi) = \varphi(\xi) - \int_s^t \big(\tfrac{1}{2}(v_\theta',\varphi') + (fv_\theta' + f'v_\theta,\varphi)\big)\, d\theta + \int_s^t (N_0^{-1} I v_\theta, \varphi)\, dy_\theta.$$

Then Theorem 1 of [17], Section 4.1, states that such a solution is unique, provided that it exists. However, applying Itô's formula to the identity

$$\big(\tilde{T}_i(s,\xi;t,\cdot),\varphi\big)=\frac{1}{\Phi(\xi)}E_0\big(\varphi(x_t)\Phi(x_t)L_{i,s,t}\,|\,\mathscr{Y}_{s,t},x_s=\xi\big),$$

it is easy to verify that $\tilde{T}_i$ is a weak solution to the backward equation and directly that so is T_i. Therefore $\tilde{T}_i=T_i$ P-a.s. and (2.15) follows. □

Acknowledgments. This work is a part of the author's thesis under the supervision of Professor Ofer Zeitouni at the Technion. I am indebted to Professor Zeitouni for his guidance and continuing support. I thank him also for posing the problem studied in this article, for the idea of using [21], for the separation idea behind Lemma 3 and for many enlightening discussions. I am also grateful to Dr. Amarjit Budhiraja and to Professor Daniel Ocone for providing a preprint of [7]. Finally, the referee is gratefully acknowledged for several valuable comments.

REFERENCES

[1] ATAR, R. and ZEITOUNI, O. (1997). Lyapunov exponents for finite state nonlinear filtering. *Siam J. Control Optim.* 35 36–55.

[2] ATAR, R. and ZEITOUNI, O. (1997). Exponential stability for nonlinear filtering. *Ann. Inst. H. Poincaré Probab. Statist.* 33 697–725.

[3] ATAR, R., VIENS, F. and ZEITOUNI, O. (1998). Robustness of Zakai's equation via Feyman–Kac representations. In *Stochastic Analysis, Control, Optimization and Applications: A Volume in Honor of W. H. Fleming* (W. M. McEneaney, G. Yin and Q. Zhang eds.) Birkhäuser, Boston. To appear.

[4] BOUGEROL, P. and LACROIX, J. (1985). *Products of Random Matrices with Applications to Schrödinger Operators.* Birkhäuser, Boston.

[5] BUDHIRAJA, A. and KUSHNER, H. J. (1998). Robustness of nonlinear filters over the infinite time interval. *Siam J. Control Optim.* To appear.

[6] BUDHIRAJA, A. and KUSHNER, H. J. (1997). Approximation and limit results for nonlinear filters over an infinite time interval. Preprint.

[7] BUDHIRAJA, A. and OCONE, D. (1997). Exponential stability of discrete time filters for bounded observation noise. *Systems Control Lett.* 30 185–193.

[8] CÉROU, F. (1995). Long time asymptotics for some dynamical noise free filtering problems. *C. R. Acad. Sci. Paris Ser. I* 321 469–472.

[9] CLARK, J. M. C., OCONE, D. L. and COUMARBATCH, C. (1998). Relative entropy and error bounds for filtering of Markov process. *Math. Control Signals Systems.* To appear.

[10] DELYON, B. and ZEITOUNI, O. (1991). *Lyapunov exponents for filtering problems.* In *Applied Stochastic Analysis* (M. H. A. Davis and R. J. Elliot, eds.) 511–521. Gordon and Breach, London.

[11] KUNITA, H. Asymptotic behavior of the nonlinear filtering errors of Markov processes. *J. Multivariate Anal.* 1 365–393.

[12] KUNITA, H. (1981). Cauchy problem for stochastic partial differential equations arizing in nonlinear filtering theory. *Systems Control Lett.* 1 37–41.

[13] LE GLAND, F. and MEVEL, L. (1996). Geometric ergodicity in hidden Markov models. Publication interne 1028, IRISA.

[14] OCONE, D. (1998). Entropy inequalities and entropy dynamics in nonlinear filtering of diffusion processes. In *Stochastic Analysis, Control, Optimization and Applications: A Volume in Honor of W. H. Fleming* (W. M. McEneaney, G. Yin and Q. Zhang eds.) Birkhäuser, Boston. To appear.

[15] Ocone, D. (1997). Asymptotic stability of Beneš filters. Preprint.
[16] Ocone, D. and Pardoux, E. (1996). Asymptotic stability of the optimal filter with respect to its initial condition. *Siam J. Contr. Optim.* 34 226–243.
[17] Rozovskii, B. L. (1990). Stochastic evolution systems. In *Stochastic Evolution Systems* 315. Kluwer, Dordrecht.
[18] Stettner, L. (1989). On invariant measures of filtering processes. *Stochastic Differential Systems. Lecture Notes in Control and Inform. Sci.* 126 279–292. Springer, New York.
[19] Stettner, L. (1991). Invariant measures of the pair: state, approximate filtering process. *Colloq. Math.* 52 347–351.
[20] Zakai, M. (1969). On the optimal filtering of diffusion processes. *Z. Wahrsch Verw. Gebiete* 11 230–243.
[21] Zeitouni, O. (1988). Approximate and limit results for nonlinear filters with small observation noise: the linear sensor and constant diffusion coefficient case. *IEEE Trans. Automat. Control* 33 595–599.
[22] Zeitouni, O. and Bobrovsky, B. Z. (1986). On the reference probability approach to the equations of non-linear filtering. *Stochastics* 19 133–149.

Department of Electrical Engineering
Technion–Israel Institute of Technology
Haifa 32000
Israel
E-mail: rami@tx.technion.ac.il