Exponential Stability of Discrete Time Filters for Bounded Observation Noise

A. Budhiraja*
Department of Mathematics
Iowa State University
Ames, IA 50014
budhiraj@iastate.edu

D. Ocone
Department of Mathematics
Rutgers University
New Brunswick, NJ 08903
ocone@math.rutgers.edu

December 11, 2000

Abstract

This paper proves exponential asymptotic stability of discrete-time filters for the estimation of solutions to stochastic difference equations, when the observation noise is bounded. No assumption is made on the ergodicity of the signal. The proof uses the Hilbert projective metric, introduced into filter stability analysis by Atar and Zeitouni [1], [2]. It is shown that when the signal noise is sufficiently regular, boundedness of the observation noise implies that the filter update operation is, on average, a strict contraction with respect to the Hilbert metric. Asymptotic stability then follows.

Key Words: Hilbert metric, Birkhoff's contraction coefficient, nonlinear filtering, asymptotic stability.

*Research suppored by the Office of Naval Research grant no: N00014-96-1-0279

1 Introduction

In dynamic estimation of a Markov process observed in noise, the evolution of the conditional distribution of the signal, given the observations, is characterized by recursive filtering equations, such as the Kalman-Bucy equations or Zakai's equation. The filtering equations are stochastic evolution equations for a measure-valued process and require, as inputs, an observation path, which drives the equation, and an initial value. Without attempting a mathematically precise definition, we shall use the phrase *asymptotically stable* to desribe filtering equations whose long-time behavior is independent of initial condition, in the sense that some reasonable measure of distance between differently initialized solutions converges to zero in the infinite time limit. Identifying asymptotically stable filters is an interesting problem. The conditional distribution of the signal given the observations is the specific solution of the filtering equations whose initial value equals the initial probability law of the signal. Since one can rarely expect to know the initial signal law precisely, it is important to ask how sensitive solutions are to errors in initialization. In asymptotically stable filters, the observations correct mistakes in initialization by driving incorrectly initialized solutions towards the correct conditional distribution of the signal, as time evolves.

General results for asymptotic stability have been established in nonlinear filtering mostly for ergodic signals. In his seminal paper, Kunita [6] showed that when filtering a Markov process evolving in a compact state space, ergodicity of the signal "lifts" to ergodicity of the exact filter, which, as a measure valued process, will have an invariant law independent of initial condition. Kunita [7] and Stettner [12] generalize this result to signals with locally compact state space, and Stettner [13] and Ocone and Pardoux [11] derive consequences for asymptotic stability. Delyon and Zeitouni [5], and Atar and Zeitouni [1], [2] employ Lyapunov exponent and Hilbert projective metric techniques to prove exponential decay of the total variation norm between exact and incorrectly initialized filters for various ergodic signals. Without ergodicity assumptions, asymptotic stability is well known in linear filtering of linear systems. Given appropriate observability and detectability assumptions, calculation of the filter from Gaussian initial conditions reduces to the Kalman-Bucy equations for conditional mean and covariance, whose long-time behavior is independent of initial conditions; this is classical, see, for example Kwakernaak and Sivan [8]. Asymptotic stability extends as well to non-Gaussian initial conditions; see Makowski and Sowers [10] and Ocone and Pardoux [11]. Cerou [4] provides conditions for consistency in filtering noise free signals observed in white noise.

This paper presents a result on asymptotic stability of filters for a class of discrete time models which allow for transient signals. Intuition says that asymptotic stability should extend to filters of non-ergodic signals so long as the observations are sufficiently "good," as for the Kalman filter under observability and detectability assumptions. In this paper, our signal process $\{X_n\}$ is the solution of a stochastic difference equation and our model of "good" observations is

$$Y_n = X_n + v_n,$$

where $\{v_n\}$ is an X-independent sequence of independent, continuous, and uniformly bounded random variables. Throughout, we let M denote the uniform bound on the observation noise; thus, for all positive integers n,

$$P(|v_n| < M) = 1. \tag{1.1}$$

Our main theorem, Theorem 3.2. states that if the density of the signal noise is positive on a large enough interval, then differently initialized filters converge to one another at an exponential rate. The proof relies on using Hilbert's projective metric to measure the distance between solutions and showing that the operation of filter update is on average a strict contraction with respect to this metric. Our work is directly inspired by the papers of Atar and Zeitouni [2] [1], which introduce Hilbert metric techniques into the study of asymptotic filter convergence for filters of ergodic signals.

The tools we use from the theory of Hilbert projective metrics are basic. Hilbert's projective metric can be defined on the positive cone of any partially ordered linear topological space. We shall use it only on the space $\mathcal{M}^+[a,b]$ of positive Borel measures on a bounded interval $[a,b]$, which is the positive cone in the space $\mathcal{M}[a,b]$ of finite, signed measures given the partial order, $\mu \leq \nu$ if and only if $\mu(A) \leq \nu(A)$ for all Borel A. We restrict the discussion to this setting. If $\mu, \nu \in \mathcal{M}^+$, the Hilbert projective distance between μ and ν is

$$h(\mu,\nu) := \ln\left[\sup_{A,A'} \frac{\mu(A)}{\nu(A)}\frac{\nu(A')}{\mu(A')}\right]. \tag{1.2}$$

In this definition, the form $\alpha/0$ is interpreted as ∞ if $\alpha > 0$. The supremum is evaluated over those Borel A and A' contained in $[a,b]$ for which $\mu(A) > 0$

and $\nu(A') > 0$. When $h(\mu, \nu) < \infty$, we say that μ and ν are *comparable*. For comparable μ and ν there exist constants $0 < c_1 < c_2$ such that $c_1\nu \leq \mu \leq c_2\nu$ and $h(\mu, \nu)$ is the infimum of $\ln(c_2/c_1)$ over all such choices of c_1 and c_2. Notice that the projective metric is independent of scaling factors; for any positive constants, c and d, $h(c\mu, d\nu) = h(\mu, \nu)$. Finally, when μ and ν are probability measures

$$||\mu - \nu||_{TV} \leq \frac{2}{\ln 3} h(\mu, \nu)\,; \tag{1.3}$$

where $||.||_{TV}$ denotes the total variation norm. For a proof see [1], Lemma 1.

Now consider the Hilbert projective metrics h_1 on $\mathcal{M}^+[a_1, b_1]$ and h_2 on $\mathcal{M}^+[a_2, b_2]$. A linear operator K from signed measures on $[a_1, b_1]$ to signed measures on $[a_2, b_2]$ is called *positive* if it maps $\mathcal{M}^+[a_1, b_1]$ to $\mathcal{M}^+[a_2, b_2]$. It is easy to check that for positive K, and any μ and ν in $\mathcal{M}^+[a_1, b_1]$,

$$h_2(K\mu, K\nu) \leq h_1(\mu, \nu).$$

The main tool is an inequality of Birkhoff identifying situations in which K is a strict contraction. Define

$$H(K) = \sup\left\{h_2(K\mu, K\nu)\,;\ \mu, \nu \in \mathcal{M}^+[a_1, b_1]\right\}; \tag{1.4}$$

this is the Hilbert diameter of the image of K. Then for any μ and ν in $\mathcal{M}^+[a_1, b_1]$,

$$h_2(K\mu, K\nu) \leq \tanh\left(\frac{H(K)}{4}\right) h_1(\mu, \nu). \tag{1.5}$$

A proof in the general setting of linear operators that leave a positive cone invariant may be found in Birkhoff [3]. The recent paper of Liverani [9] contains a very direct proof for the minor generalization used here to operators between different spaces. The factor $\tanh(H(K)/4)$ is called Birkhoff's contraction coefficient. When K is the integral operator

$$K\mu(A) = \int_A \int_{a_1}^{b_1} K(x, y)\, \mu(dy)\, dx \qquad A \in \mathcal{M}^+[a_2, b_2], \tag{1.6}$$

and K is positive, then

$$H(K) = \ln\left[\sup_{y,y' \in [a_1, b_1]} \operatorname{ess\,sup}_{x,x' \in [a_2, b_2]} \left(\frac{K(x, y)K(x', y')}{K(x, y')K(x', y)}\right)\right], \tag{1.7}$$

where the essential supremum is calculated with respect to Lebesgue measure. The proof of this known result is straightforward, using the fact,

$$\sup_A \frac{\int_A f(x)\,ds}{\int_A g(x)\,dx} = \operatorname{ess\,sup}_x \frac{f(x)}{g(x)}.$$

We remark here that, if $m\{x : K(x,y) > 0\} = 0$ for some $y \in [a_1, b_1]$, m being the Lebesgue measure on $[a_2, b_2]$, then K is no more a positive linear operator, since, for example, $K\delta_y = 0$. Nevertheless if one defines $H(K) = \infty$ for such K, Equation (1.5) remains valid, with the convention that $h(0,0) = 0$ and $h(0,\nu) = \infty$ if $\nu > 0$.

The invariance of the Hilbert metric to scalar multiplication, in combination with Birkhoff's contraction formula, make the Hilbert metric a very convenient tool for filtering. Consider filtering from observations $Y_n = X_n + v_n$, assuming condition (1.1). Then given Y_n, we know that the signal X_n must lie in the interval $[Y_n - M, Y_n + M]$, and hence any solution of the filtering equations at time n must be a measure supported in $[Y_n - M, Y_n + M]$. The filtering equations are defined by an updating formula that takes $\mu_{n-1} \in \mathcal{M}^+[Y_{n-1} - M, Y_{n-1} + M]$ to the measure

$$\mu_n = \widetilde{K}_n \mu_{n-1} := \frac{K_n \mu_{n-1}}{K_n \mu_{n-1}[Y_n - M, Y_n + M]}, \tag{1.8}$$

where K_n is an linear integral operator of the form (1.6) taking values in $\mathcal{M}^+[Y_n - M, Y_n + M] \cup \{0\}$; see Equation (3.2) below for the explicit definition of K_n. The normalization that appears in (1.8) and that renders $\widetilde{K}_n$ a non-linear operator, makes it difficult to derive its contraction properties directly with respect to the total variation norm. However, because of the invariance of Hilbert metric relative to normalization, the contraction(in Hilbert metric) resulting from the application of $\tilde{K}_n$ exactly equals that from the linear operator K_n. In the proof of Theorem 3.2 ,we will show that K_n is a strict contraction on average, which will imply exponential decay of the Hilbert metric distance between filters, and thus, by the bound (1.3), exponential decay of total variation distance.

The drawback of the metric defined in (1.2) is its reliance on the assumption of bounded observation noise. When the support of the noise density is not bounded, the filter solutions will not have bounded support either. However, (1.2) does not appear to be useful for comparing probability measures on the real line R. For example, two Gaussian distributions with identical variances, but different means, will not be comparable, so h will not even capture convergence of normal distributions by convergence in their means. Moreover, the filter update operator K_n will not in general be a strict contraction for measures with unbounded support. Hence theorems of asymptotic stability for unbounded observation noise models require a more delicate analysis.

The paper is organized as follows. In Section 2 we give our filtering model and the basic assumptions. Our main results, without proofs, are in Section 3. Finally, Section 4 is devoted to proofs.

2 Notation and Assumptions

Let $(\Omega, \mathcal{F}, P)$ be a probability space on which two i.i.d. sequences, $\{\xi_n\}_{n=1}^{\infty}$ and $\{\nu_n\}_{n=1}^{\infty}$, independent of each other, are defined. We assume that both ξ_1 and ν_1 have bounded densities, denoted by f and g respectively. Furthermore, we assume that:

(A.1) $$\text{supp} g \subseteq [-M, M], \qquad M < \infty.$$

Let X_0 be another random variable on the above probability space, independent of both $\{\xi_i\}$ and $\{\nu_i\}$, with law p_0. The signal $\{X_n\}_{n=0}^{\infty}$ is defined as:

$$X_{n+1} = m(X_n) + \sigma(X_n)\xi_{n+1}, \qquad n \geq 0 \tag{2.1}$$

In (2.1) m and σ are Borel measurable and are assumed to satisfy:

(A.2) $$0 < \underline{\sigma} := \inf_{x \in R} \sigma(x) \leq \sup_{x \in R} \sigma(x) =: \overline{\sigma} < \infty.$$

(A.3) $$C := \sup_{|z-z'| \leq 2M} |m(z) - m(z')| < \infty.$$

Note that (A.3) is satisfied if m is globally Lipschitz.

The observations on the signal are given as:

$$Y_n = X_n + \nu_n, \qquad n \geq 1. \tag{2.2}$$

Denote $\sigma(Y_1, \cdots, Y_n)$ by $\mathcal{Y}_n$ and the conditional distribution of X_n given $\mathcal{Y}_n$ by Π_n. Observe that $\text{supp}\Pi_n \subseteq [Y_n - M, Y_n + M]$. It can be shown(see the next section) that there exists a map: $\overline{\Pi}_n : R^n \times \mathcal{P}(R) \to \mathcal{P}(R)$, such that $\Pi_n = \overline{\Pi}_n(Y_1, \cdots, Y_n, p_0)$, a.s., where $\mathcal{P}(R)$ is the space of probability measures on R.

Finally let p_1 and p_2 be probability measures on R. In our analysis, they are two different initializations of the filter. Define for $i = 1, 2$,

$$\Pi_n^{(i)} := \overline{\Pi}_n(Y_1, \cdots, Y_n, p_i).$$

Observe that $\Pi_n^{(i)}$ corresponds to the filter with the initialization p_i. Our objective is to show that the distance between $\Pi_n^{(1)}$ and $\Pi_n^{(2)}$ converges to zero exponentially fast(in the sense, to be made precise in the next section) as $n \to \infty$.

3 Recursive filter and the main theorem

In order to derive an explicit formula for $\Pi_n(Y_1, \cdots, Y_n, p_0)$, we introduce the following sequence of random linear operators. For $n \geq 2$, define:

$$K_n : \mathcal{M}[Y_{n-1} - M, Y_{n-1} + M] \to \mathcal{M}[Y_n - M, Y_n + M], \tag{3.1}$$

by:

$$K_n\mu(A) = \int_A \int_{[Y_{n-1}-M, Y_{n-1}+M]} g(Y_n - x) f(\frac{x - m(z)}{\sigma(z)}) \sigma^{-1}(z) \mu(dz) dx, \tag{3.2}$$

$A \in \mathcal{B}[Y_n - M, Y_n + M]$.

Notice that K_n maps $\mathcal{M}^+[Y_{n-1} - M, Y_{n-1} + M]$ to $\mathcal{M}^+[Y_n - M, Y_n + M] \cup \{0\}$. Moreover it is a simple verification that K_n takes Π_{n-1} to Π_n, up to a normalizing factor. More precisely we mean the following.

For $\alpha \geq 0$, define $\alpha^{-1} := 1_{\{\alpha > 0\}}/\alpha$ and for $\mu \in \mathcal{M}^+[Y_{n-1} - M, Y_{n-1} + M] \cup \{0\}$, define: $\tilde{K_n}\mu := [K_n\mu([Y_n - M, Y_n + M])]^{-1} K_n\mu$. Then, $\forall n \geq 2$, $\tilde{K_n}\Pi_{n-1} = \Pi_n$.

To calculate the filter update at time $n = 1$, define:

$$\tilde{K_1} : \mathcal{P}(R) \to \mu^+[Y_1 - M, Y_1 + M] \cup \{0\},$$

as follows

$$(\tilde{K_1}p)(A) := \frac{\int_A \int_R g(Y_1 - x) f(\frac{x-m(z)}{\sigma(z)}) \sigma^{-1}(z) p(dz) dx}{\int_{[Y_1-M, Y_1+M]} \int_R g(Y_1 - x) f(\frac{x-m(z)}{\sigma(z)}) \sigma^{-1}(z) p(dz) dx}, \tag{3.3}$$

$\forall p \in \mathcal{P}(R)$ and $A \in \mathcal{B}[Y_1 - M, Y_1 + M]$.

Recalling that p_0 is the law of X_0, we have the following lemma.

Lemma 3.1 $\Pi_n \equiv \Pi_n(Y_1, \cdots Y_n, p_0) = \tilde{K_n} o \cdots o \tilde{K_1} p_0$.

Observe that $\tilde{K_n}, n \geq 1$ does not depend on p_0 and is a function of system parameters and observations alone. A filter initialized with p_i instead of p_0 leads us to

$$\Pi_n^{(i)} \equiv \Pi_n(Y_1, \cdots Y_n, p_i) = \tilde{K_n} o \cdots o \tilde{K_1} p_i.$$

The central result of this work is the following theorem. Define for $x \in R$,

$$\gamma(x) := \inf\{f(u) : |u| \leq [2M + C + \overline{\sigma}|x|]/\underline{\sigma}\}.$$

Let h_n denote the Hilbert projective metric and $||.||_{TV}$ the total variation norm on $\mathcal{M}^+$ $[Y_n - M, Y_n + M]$. Define for $i = 1, 2$:

$$S_i := \inf\{k : \Pi_k^{(i)} = 0\}.$$

Theorem 3.2 *Assume that*

$$P[\gamma(\xi_1) > 0] > 0. \tag{3.4}$$

Then, almost everywhere on $\{S_1 \wedge S_2 = \infty\}$*:*

$$\overline{\lim}_{n\to\infty} \frac{1}{n} \ln ||\Pi_n^{(1)} - \Pi_n^{(2)}||_{TV} \leq \quad \overline{\lim}_{n\to\infty} \frac{1}{n} \ln h_n(\Pi_n^{(1)}, \Pi_n^{(2)}) < \quad 0. \tag{3.5}$$

Remark 3. 1 *Note that the condition (3.4) is satisfied if f is nonzero on a sufficiently large interval (for example, $[-K, K]$, where $K > (2M + C)/\underline{\sigma}$.) The condition (3.4) applies in particular to the case when the support of the observation noise density equals $[-M, M]$. If the support is actual some proper subset of $[-M, M]$, one can refine the theorem by an appropriate modification of γ.*

Remark 3. 2 *The conclusion of the above theorem (assuming (3.4)) clearly holds on $\{S_1 \vee S_2 < \infty\}$, since in this case after a finite(random) number of steps both $\Pi_n^{(1)}$ and $\Pi_n^{(2)}$ become zero.*

Remark 3. 3 *If* $p_0 \ll p_i$ *for* $i = 1, 2$ *then it can be easily verified that* $\forall n > 1$, $\Pi_n \ll \Pi_n^i, i = 1, 2$. *Observing that* $\Pi_n > 0, \forall n > 1$, *we have that in this case* $P(S_1 \wedge S_2 = \infty) = 1$ *and therefore under the assumption (3.4), the conclusion (3.5) holds a.e.*

The following corollary says that an incorrectly initialized filter, either becomes identically zero after a finite number of steps(in which case we can discard that initialization) or it converges to the correct filter exponentially fast.

Corollary 3. 3 *Let the condition (3.4) be true and let* $p_1 \neq p_0$ *be an incorrect initialization. Then exactly one of the following can occur.*
(i) $\overline{\lim}_{n\to\infty} \frac{1}{n} \ln h_n(\Pi_n, \Pi_n^{(1)}) < \quad 0.$
(ii) $S_1 < \infty$.

Remark 3. 4 *The above theorem can be easily modified to cover the case where the observation process is* $Y_n = h(X_n) + \nu_n$ *with* h *being a* $1:1$ *function with a globally Lipschitz inverse. For example if* $supp(f) = (-\infty, \infty)$, *(3.5) holds (with the above stated observation process) without any additional assumption of the form (3.4). For the general case, (3.4) needs to be modified appropriately.*

In special cases, as the following proposition shows, one can obtain additional information on the rate of convergence of $h_n(\Pi_n^{(1)}, \Pi_n^{(2)})$ than provided by the above theorem.

Proposition 3.4 *Let* $suppf = (-\infty, \infty), \sigma \equiv 1$ *and* $\sup_x |(\ln f)''(x)| := \kappa < \infty$. *Then*

$$\overline{\lim}_{n\to\infty} \frac{1}{n} \ln h_n(\Pi_n^{(1)}, \Pi_n^{(2)}) < \quad \ln \tanh(\kappa MC/2). \tag{3.6}$$

The proofs of the above results are based on the following lemmas.

The first lemma that we state will be used in observing that even though K_n is not necessarily a positive operator, the essential results of Birkhoff introduced in Section 1 can still be applied. We begin by giving a slightly extended definition of $H(K_n)$.

Define

$$H(K_n) = \begin{cases} \infty, \\ \text{if } \exists \mu \in \mathcal{M}^+ \ [Y_{n-1} - M, Y_{n-1} + M], \text{ such that } K_n\mu = 0 \\ \\ \ln\left[\sup_{y,y'} \operatorname{ess\,sup}_{x,x'} \left(\frac{K_n(x,y)K_n(x',y')}{K_n(x,y')K_n(x',y)}\right)\right], \\ \text{otherwise} \end{cases}$$

In the above definition, for $(x, y) \in [Y_n - M, Y_n + M] \times [Y_{n-1} - M, Y_{n-1} + M]$, $K_n(x, y) := f(\frac{x-m(y)}{\sigma(y)})$. The essential supremum is taken with respect to the Lebesgue measure for x and x' in $[Y_n - M, Y_n + M]$, and the supremum is taken with respect to y and y' in $[Y_{n-1} - M, Y_{n-1} + M]$.

Lemma 3.5 *For* $n \geq 2$

$$h_n(\Pi_n^{(1)}, \Pi_n^{(2)}) \leq \tanh\left(\frac{H(K_n)}{4}\right) h_{n-1}(\Pi_{n-1}^{(1)}, \Pi_{n-1}^{(2)}). \tag{3.7}$$

The next lemma will be used to derive an upper bound for $H(K_n)$. Let $x \in [Y_n - M, Y_n + M]$, $z \in [Y_{n-1} - M, Y_{n-1} + M]$. Observing that $X_{n-1} \in [Y_{n-1} - M, Y_{n-1} + M]$, we have that $|z - X_{n-1}| \leq 2M$ and therefore:

$$\begin{aligned} \left|\frac{x - m(z)}{\sigma(z)}\right| &\leq \frac{1}{\underline{\sigma}}(|x - Y_n + Y_n - m(z)|) \\ &= \frac{1}{\underline{\sigma}}(|x - Y_n + m(X_{n-1}) - m(z) + \sigma(X_{n-1})\xi_n + \nu_n|) \\ &\leq \frac{1}{\underline{\sigma}}(2M + C + \overline{\sigma}|\xi_n|), \end{aligned}$$

where C is as in (A.3).

The above computation yields that for $x \in [Y_n - M, Y_n + M]$, $z \in [Y_{n-1} - M, Y_{n-1} + M]$, $|f\left(\frac{x-m(z)}{\sigma(z)}\right)| \geq \gamma(\xi_n)$. The consequence of this observation is the following lemma. Define $\bar{f} = \sup_x f(x)$ and recall that we assume $\bar{f} < \infty$.

Lemma 3.6 $H(K_n) \leq \ln\left(\frac{\bar{f}^2}{\gamma^2(\xi_n)}\right)$.

The next lemma follows immediately from Lemma 3.5.

Lemma 3.7 *If* $\gamma(\xi_n) > 0$, $\Pi_n^{(1)} > 0$ *and* $\Pi_n^{(2)} > 0$, *then* $h_n(\Pi_n^{(1)}, \Pi_n^{(2)}) < \infty$.

4 Proofs

Proof of Lemma 3.1: In what follows we will use the notation: $\Phi_{U/V}$ for the conditional density of a random vector U given the random vector V. An application of Bayes formula gives that for $n > 1$:

$$\Pi_n(x) = c.\Phi_{(X_n,Y_n)/\mathbf{Y}_{n-1}}(x, Y_n),$$

where c is a normalizing constant and for $m \geq 1$, $\mathbf{Y}_m := (Y_1, \cdots, Y_m)$. Next observe that

$$\Phi_{(X_n,Y_n)/\mathbf{Y}_{n-1}}(x, Y_n) = \Phi_{X_n/\mathbf{Y}_{n-1}}(x).\Phi_{Y_n/(\mathbf{Y}_{n-1},X_n=x)}(Y_n).$$

Clearly, $\Phi_{Y_n/(\mathbf{Y}_{n-1},X_n=x)}(Y_n) = g(Y_n - x)1_{[x-M,x+M]}(Y_n)$ and

$$\begin{aligned}
\Phi_{X_n/\mathbf{Y}_{n-1}}(x) &= \int_R \Phi_{X_n/(\mathbf{Y}_{n-1},X_{n-1}=y}(x).\Phi_{X_{n-1}/\mathbf{Y}_{n-1}}(y)dy \\
&= \int_{[Y_{n-1}-M,Y_{n-1}+M]} \sigma^{-1}(y) f\left(\frac{x - m(y)}{\sigma(y)}\right)\Pi_{n-1}(y)dy,
\end{aligned}$$

where the last equality follows on noting that $\text{supp}\Pi_{n-1} = [Y_{n-1} - M, Y_{n-1} + M]$.

Combining the above two observations we have the lemma. $\square$

Proof of Theorem 3.2: Let $T = \min\{n : \gamma(\xi_n) > 0\}$ and $S := S_1 \wedge S_2$. By (3.4) and the independence of $\{\xi_i\}$; $P[T < \infty] = 1$. It suffices therefore to prove (3.5) on $\{T = k\} \cap \{S = \infty\}$ for an arbitrary k. Fixing k, we have, using Lemma 3.4, that on the set $\{T = k, S = \infty\}, \forall n \geq k$,

$$\begin{aligned}
\frac{1}{n}\ln h_n(\Pi_n^{(1)}, \Pi_n^{(2)}) &\leq \frac{1}{n}\left[\sum_{j=k+1}^{n} \ln\tanh\left(\frac{H(K_j)}{4}\right) + \ln h_k(\Pi_k^{(1)}, \Pi_k^{(2)})\right] \\
&\leq \frac{1}{n}\sum_{j=k+1}^{n} \ln\tanh\left(\frac{\ln\left(\frac{\overline{f}^2}{\gamma^2(\xi_j)}\right)}{4}\right) + \frac{\ln h_k(\Pi_k^{(1)}, \Pi_k^{(2)})}{n},
\end{aligned}$$

where the last inequality follows from Lemma 3.5. Observing from Lemma 3.6 that $h_k(\Pi_k^{(1)}, \Pi_k^{(2)}) < \infty$, we have that on $[T = k, S = \infty]$,

$$\limsup_{n\to\infty} \frac{1}{n}\ln h_n(\Pi_n^{(1)}, \Pi_n^{(2)}) \leq \limsup_{n\to\infty} \frac{1}{n}\sum_{j=k+1}^{n} \ln\tanh\left(\frac{\ln\left(\frac{\overline{f}^2}{\gamma^2(\xi_j)}\right)}{4}\right).$$

The above inequality yields from the strong law of large numbers that for almost every $\omega \in [T = k, S = \infty]$:

$$\limsup_{n\to\infty} \frac{1}{n} \ln h_n(\Pi_n^{(1)}, \Pi_n^{(2)}) \leq E[\ln \tanh \left(\frac{\ln\left(\frac{\bar{f}^2}{\gamma^2(\xi_j)}\right)}{4} \right)]. \qquad (4.1)$$

The result now follows on observing that by (3.4) the right side of (4.1) is strictly less than 0. □

Remark 4. 1 *No use was made in the proof of the above theorem that the ν_k's are identically distributed.*

Proof of Proposition 3.3: As in the above theorem, we have that on $[T = k]$:

$$\limsup_{n\to\infty} \frac{1}{n} \ln h_n(\Pi_n^{(1)}, \Pi_n^{(2)}) \leq \limsup_{n\to\infty} \frac{1}{n} \sum_{j=k+1}^{n} \ln \tanh \left(\frac{H(K_j)}{4} \right). \qquad (4.1)$$

Next observe that:

$$\begin{aligned}
&\ln \left[\frac{f(x - m(y))f(x' - m(y'))}{f(x - m(y'))f(x' - m(y))} \right] \\
&= \ln f(x - m(y)) + \ln f(x' - m(y')) - \ln f(x - m(y')) - \ln f(x' - m(y)) \\
&= \int_{x'}^{x} [(\ln f)'(u - m(y)) - (\ln f)'(u - m(y')]du \\
&\leq \kappa |m(y) - m(y')||x - x'|.
\end{aligned}$$

The above expression is clearly bounded by: $2\kappa CM$ for $x, x' \in [Y_n - M, Y_n + M]$and $y, y' \in [Y_{n-1} - M, Y_{n-1} + M]$. Using this observation in (4.1) we have the result. □

Proof of Lemma 3.6: If $\gamma(\xi_n) > 0$, we have from Lemma 3.5 that $H(K_n) < \infty$ and therefore

$$\begin{aligned}
h_n(\Pi_n^{(1)}, \Pi_n^{(2)}) &= h_n(K_n \Pi_{n-1}^{(1)}, K_n \Pi_{n-1}^{(2)}) \\
&\leq \sup_{\mu,\nu} [h_n(K_n \mu, K_n \nu) : \nu, \mu \in \mathcal{M}^+[Y_{n-1} - M, Y_{n-1} + M]] \\
&= H(K_n) \qquad < \infty. \qquad \square
\end{aligned}$$

Proof of Lemma 3.5: It suffices to consider the case, $\gamma(\xi_n) > 0$. This implies from the definition of γ that $f\left(\frac{x-m(z)}{\sigma(z)}\right) > 0$ for all $x \in [Y_n - M, Y_n + M]$and $z \in [Y_{n-1} - M, Y_{n-1} + M]$. Consequently, K_n is a positive operator. The lemma is now an immidiate consequence of the definition of $H(K_n)$ and γ. $\square$

Proof of Lemma 3.4: It suffices to consider the case in which

$$h_{n-1}(\Pi_{n-1}^{(1)}, \Pi_{n-1}^{(2)}) < \infty.$$

In this case $\Pi_{n-1}^{(1)}$ and $\Pi_{n-1}^{(2)}$ are comparable. Let α and β be such that:

$$\alpha\Pi_{n-1}^{(1)} \leq \Pi_{n-1}^{(2)} \leq \beta\Pi_{n-1}^{(1)}.$$

¿From the definition of K_n it is clear then that

$$\alpha K_n\Pi_{n-1}^{(1)} \leq K_n\Pi_{n-1}^{(2)} \leq \beta K_n\Pi_{n-1}^{(1)}.$$

The above computation implies that

$$h_n(\Pi_n^{(1)}, \Pi_n^{(2)}) \leq h_{n-1}(\Pi_{n-1}^{(1)}, \Pi_{n-1}^{(2)}).$$

Hence (3.7) clearly holds if $H(K_n) = \infty$. On the other hand, if $H(K_n) < \infty$, then K_n is a positive operator and we have the result from the classical Birkhoff's contraction estimate(see (1.5)). $\square$

References

[1] R. Atar and O. Zeitouni. Exponential stability for nonlinear filtering. *Annales de l'Institut H. Poincare Probabilites et Statistique*, 33:697–726, 1997.

[2] R. Atar and O. Zeitouni. Lyapunov exponents for finite state nonlinear filtering. *SIAM Journal of Control and Optimization*, 35:36–55, 1997.

[3] G. Birkhoff. *Lattice Theory.* Am. Math. Soc. Publ.25, 3rd ed., 1967.

[4] F. Cérou. Long time asymptotics for some dynamical noise free non linear filtering problems. *Rapport de Recherche 2446, INRIA*, December, 1994.

[5] B. Delyon and O. Zeitouni. Lyapunov exponents for filtering problems. In M.H.A. Davis and R.J. Elliot, editors, *Applied Stochastic Analysis*, pages 511–521, 1991.

[6] H. Kunita. Asymptotic behavior of the nonlinear filtering errors of Markov processes. *Journal of Multivariate Analysis*, 1:365–393, 1971.

[7] H. Kunita. Ergodic properties of nonlinear filtering processes. In K.C. Alexander and J.C. Watkins, editors, *Spatial Stochastic Processes*, 1991.

[8] H. Kwakernaak and R. Sivan. *Linear Optimal Control Systems.* Wiley, Interscience, New York, 1972.

[9] C. Liverani. Decay of correlations. *Annals of Mathematics*, 142:239–301, 1995.

[10] A.M. Makowski and R.B. Sowers. Discrete time filtering for linear systems with non-gaussian initial conditions. *IEEE Trans. Automat. Control*, 33:114–121, 1992.

[11] D. Ocone and E. Pardoux. Asymptotic stability of the optimal filter with respect to its initial condition. *SIAM Journal of Control and Optimization*, 34:226–243, 1996.

[12] L. Stettner. On invariant measures of filtering processes. In K. Helmes, N. Christopeit, and M. Kohlmann, editors, *Stochastic differential systems, Proc. 4th Bad Honnef Conf.,1988, Lecture Notes in Control and Inform Sci.*, pages 279–292, 1989.

[13] L. Stettner. Invariant measures of pair: State, approximate filtering process. *Colloq. Math. LXII*, pages 347–352, 1991.