Elect. Comm. in Probab. **11** (2006), 129–140

ELECTRONIC COMMUNICATIONS in PROBABILITY

ON A ROLE OF PREDICTOR IN THE FILTERING STABILITY

PAVEL CHIGANSKY[1]
Department of Mathematics, The Weizmann Institute of Science, Rehovot 76100, Israel
email: `pavel.chigansky@weizmann.ac.il`

ROBERT LIPTSER
Department of Electrical Engineering Systems, Tel Aviv University, 69978 Tel Aviv, Israel
email: `liptser@eng.tau.ac.il`

Submitted November 26, 2005, *accepted in final form* July 6, 2006

AMS 2000 Subject classification: 93E11, 60J57
Keywords: nonlinear filtering, stability, martingale convergence

Abstract
When is a nonlinear filter stable with respect to its initial condition? In spite of the recent progress, this question still lacks a complete answer in general. Currently available results indicate that stability of the filter depends on the signal ergodic properties and the observation process regularity and may fail if either of the ingredients is ignored. In this note we address the question of stability in a particular weak sense and show that the estimates of certain functions are always stable. This is verified without dealing directly with the filtering equation and turns to be inherited from certain one-step predictor estimates.

1 Introduction

Consider the filtering problem for a Markov chain $(X,Y)=(X_n,Y_n)_{n\in\mathbb{Z}_+}$ with the signal X and observation Y. The signal process X is a Markov chain itself with the transition kernel $\Lambda(u,dx)$ and initial distribution ν. The observation process Y has the transition probability law

$$\mathsf{P}(Y_n\in B|X_{n-1},Y_{n-1})=\int_B\gamma(X_{n-1},y)\varphi(dy),\quad B\in\mathscr{B}(\mathbb{R}),$$

where $\gamma(u,y)$ is a density with respect to a σ-finite measure φ on $\mathbb{R}$. We set $Y_0=0$, so that, a priori information on the signal state at time $n=0$ is confined to the signal distribution ν. The random process (X,Y) is assumed to be defined on a complete probability space $(\Omega,\mathscr{F},\mathsf{P})$. Let $(\mathscr{F}^Y_n)_{n\ge0}$ be the filtration generated by Y:

$$\mathscr{F}^Y_0=\{\varnothing,\Omega\},\quad \mathscr{F}^Y_n=\sigma\{Y_1,\dots,Y_n\}.$$

[1] RESEARCH OF THE FIRST AUTHOR IS SUPPORTED BY A GRANT FROM THE ISRAEL SCIENCE FOUNDATION

It is well known that the regular conditional distribution $d\mathsf{P}(X_n \le x|\mathscr{F}_n^Y) =: \pi_n(dx)$ solves the recursive Bayes formula, called the *nonlinear filter*:

$$\pi_n(dx) = \frac{\int_{\mathbb{R}} \Lambda(u,dx)\gamma(u,Y_n)\pi_{n-1}(du)}{\int_{\mathbb{R}} \gamma(v,Y_n)\pi_{n-1}(dv)}, \quad n \ge 1, \tag{1.1}$$

subject to $\pi_0(dx) = \nu(dx)$. Clearly

$$\pi_n(f) := \int_{\mathbb{R}} f(x)\pi_n(dx)$$

is a version of the conditional expectation $\mathsf{E}\big(f(X_n)|\mathscr{F}_n^Y\big)$ for any measurable function $f = f(x)$, with $\mathsf{E}|f(X_n)| < \infty$.
Assume ν is unknown and the filter (1.1) is initialized with a probability distribution $\bar\nu$, different from ν and denote the corresponding solution by $\bar\pi = (\bar\pi_n)_{n\ge 0}$. Obviously, an arbitrary choice of $\bar\nu$ may not be admissible: it makes sense to choose $\bar\nu$ such that $\bar\pi_n(dx)$ preserves the properties of a probability distribution, i.e. $\int_B \bar\pi_n(dx) \ge 0$ for any measurable set $B \in \mathbb{R}$ and $\int_{\mathbb{R}} \bar\pi_n(dx) = 1$ for each $n \ge 1$ P-a.s. This would be the case if the right hand side of (1.1) does not lead to 0/0 uncertainty with a positive probability. As explained in the next section, the latter is provided by the relation $\nu \ll \bar\nu$, which is assumed to be in force hereafter. In fact it plays an essential role in the proof of main result.
The sequence $\bar\pi = (\bar\pi_n)_{n\ge 0}$ of random measures generally differs from $\pi = (\pi_n)_{n\ge 0}$ and the estimate $\bar\pi_n(f)$ of a particular function f is said to be stable if

$$\mathsf{E}\big|\pi_n(f) - \bar\pi_n(f)\big| \xrightarrow[n\to\infty]{} 0 \tag{1.2}$$

holds for any admissible pair $(\nu, \bar\nu)$.
The verification of (1.2) in terms of $\Lambda(u,dx)$, $\gamma(x,y)$, $\varphi(dy)$ is quite a nontrivial problem, which is far from being completely understood in spite of the extensive research during the last decade.
For a bounded f, (1.2) is closely related to ergodicity of $\pi = (\pi_n)_{n\ge 0}$, viewed as a Markov process on the space of probability measures. In the late 50's D. Blackwell, motivated by the information theory problems, conjectured in [5] that π has a unique invariant measure in the particular case of ergodic Markov chain X with a finite state space and noiseless observations $Y_n = h(X_n)$, where h is a fixed function. This conjecture was found to be false by T. Kaijser, [15]. In the continuous time setting, H. Kunita addressed the same question in [16] for a filtering model with general Feller-Markov process X and observations

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

where the Wiener process $W = (W_t)_{t\ge 0}$ is independent of X. According to [16], the filtering process $\pi = (\pi_t)_{t\ge 0}$ inherits ergodic properties from X, if the tail σ-algebra of X is P-a.s. empty. Unfortunately this assertion remains questionable due to a gap in its proof (see [4]).

Notice that (1.2) for bounded f also follows from

$$\big\|\pi_n - \bar\pi_n\big\|_{\mathsf{tv}} \xrightarrow[n\to\infty]{} 0, \quad \mathsf{P}-a.s., \tag{1.4}$$

where $\|\cdot\|_{\mathsf{tv}}$ is the total variation norm. Typically this stronger type of stability holds when X is an ergodic Markov chain with the state space $\mathbb{S} \subseteq \mathbb{R}$ (or $\mathbb{R}^d$, $d \ge 1$) and its transition probability kernel $\Lambda(u, dx)$ is absolutely continuous with respect to a σ-finite reference measure $\psi(dx)$,

$$\Lambda(u, dx) = \lambda(u, x)\psi(dx),$$

while the density λ satisfies the so called *mixing* condition:

$$0 < \lambda_* \le \lambda(u, x) \le \lambda^*, \quad \forall\, x, u \tag{1.5}$$

with a pair of positive constants λ_* and λ^*. Then (see [2], [18], [11], [8]),

$$\varlimsup_{n\to\infty} \frac{1}{n} \log \big\|\pi_n - \bar{\pi}_n\big\|_{\mathsf{tv}} \le -\frac{\lambda_*}{\lambda^*}, \quad \mathsf{P}\text{-}a.s. \tag{1.6}$$

The condition (1.5) was recently relaxed in [8], where (1.6) was verified with λ_* replaced by

$$\lambda_\circ := \int_{\mathbb{S}} \operatorname*{essinf}_{x\in\mathbb{S}} \lambda(u, x)\mu(u)\psi(du),$$

with $\mu(u)$ being the invariant density of the signal relative to $\psi(du)$.
The mixing condition, including its weaker form, implies geometric ergodicity of the signal (see [8]). However, in general the ergodicity (and even geometrical ergodicity) itself does not imply stability of the filter (see counterexamples in [15], [10], [4]). If the signal process X is compactly supported, the density $\lambda(u, x)$ usually corresponds to the Lebesgue measure or purely atomic reference measure $\psi(dx)$. Signals with non compact state space do not fit the mixing condition framework since an appropriate reference measure is hard to find and sometimes it doesn't exist (as for the Kalman-Bucy filter).
In non-compact or non-ergodic settings, the filtering stability can be verified under additional structural assumptions on (X, Y). In this connection, we mention the Kalman-Bucy filter being stable for controllable and observable linear systems (see e.g. [10], [20], [19], Sections 14.6 and 16.2). Similarly, in the nonlinear case certain relations between $\lambda(x, u)$ and $\gamma(x, y)$ provide (1.4) (see e.g. [6], [7], [1], [17], [4]).
In summary, stability of the nonlinear filter stems from a delicate interplay of the signal ergodic properties and the observations "quality". If one of these ingredients is removed, the other should be strengthened in order to keep the filter stable. Notably all the available results verify (1.2) via (1.4) and, thus, require restricting assumptions on the signal structure. Naturally, this raises the following question: are there functions f for which (1.2) holds with "minimal" constraints on the signal model ?
In this note, we give examples of functions for which this question has an affirmative answer. It turns out that (1.2) holds if $\nu \ll \bar{\nu}$ and the integral equation with respect to g,

$$f(x) = \int_{\mathbb{R}} g(y)\gamma(x, y)\varphi(dy), \tag{1.7}$$

has a bounded solution. The proof of this fact relies on the martingale convergence theorem rather than direct analysis of filtering equation (1.1).
The precise formulations and other generalizations with their proofs are given in Section 2. Several nonstandard examples are discussed in Section 3.

2 Preliminaries and the main result

For notational convenience, we assume that the pair (X, Y) is a coordinate process defined on the canonical measurable space $(\Omega, \mathscr{F})$ with $\Omega = (\mathbb{R}^\infty \times \mathbb{R}^\infty)$ and $\mathscr{F} = \mathscr{B}(\mathbb{R}^\infty \times \mathbb{R}^\infty)$, where

$\mathscr{B}$ stands for the Borel σ-algebra. Let P be a probability measure on $(\Omega,\mathscr{F})$ such that (X,Y) is Markov a process with the transition kernel $\gamma(u,y)\Lambda(u,dx)\varphi(dy)$ and the initial distribution $\nu(dx)\delta_{\{0\}}(dy)$, where $\delta_{\{0\}}(dy)$ is the point measure at zero. Let $\bar{\mathsf{P}}$ be another probability measure on $(\Omega,\mathscr{F})$ such that (X,Y) is Markov process with the same transition law and the initial distribution $\bar{\nu}(dx)\delta_{\{0\}}(dy)$. Hereafter, E and $\bar{\mathsf{E}}$ denote expectations relative to P and $\bar{\mathsf{P}}$ respectively. By the Markov property of (X,Y),

$$\nu \ll \bar{\nu} \Rightarrow \mathsf{P} \ll \bar{\mathsf{P}} \quad \text{and} \quad \frac{d\mathsf{P}}{d\bar{\mathsf{P}}}(x,y) = \frac{d\nu}{d\bar{\nu}}(x_0), \quad \bar{\mathsf{P}}\text{-a.s.}$$

We assume that $\mathscr{F}^Y_0$ is completed with respect to $\bar{\mathsf{P}}$. Denote $\mathscr{F}^Y_\infty = \bigvee_{n\ge 0}\mathscr{F}^Y_n$ and let P^Y, $\bar{\mathsf{P}}^Y$ and P^Y_n, $\bar{\mathsf{P}}^Y_n$ be the restrictions of P, $\bar{\mathsf{P}}$ on $\mathscr{F}^Y_\infty$ and $\mathscr{F}^Y_n$ respectively. Obviously, $\mathsf{P} \ll \bar{\mathsf{P}} \Rightarrow \mathsf{P}^Y \ll \bar{\mathsf{P}}^Y$ and $\mathsf{P}^Y_n \ll \bar{\mathsf{P}}^Y_n$ with the densities

$$\frac{d\mathsf{P}^Y}{d\bar{\mathsf{P}}^Y} = \bar{\mathsf{E}}\Big(\frac{d\nu}{d\bar{\nu}}(X_0)\big|\mathscr{F}^Y_\infty\Big) \quad \text{and} \quad \frac{d\mathsf{P}^Y_n}{d\bar{\mathsf{P}}^Y_n} = \bar{\mathsf{E}}\Big(\frac{d\nu}{d\bar{\nu}}(X_0)\big|\mathscr{F}^Y_n\Big) := \varrho_n.$$

Let $\bar{\pi}_n(dx)$ be the solution of (1.1) subject to $\bar{\nu}$ considered on $\big(\Omega,\mathscr{F},\bar{\mathsf{P}}\big)$, so that, it is a version of the conditional distribution $\bar{\mathsf{P}}(X_n \le x|\mathscr{F}^Y_n)$. Since $\mathsf{P} \ll \bar{\mathsf{P}}$, $\bar{\pi}_n$ satisfies (1.1) on $\big(\Omega,\mathscr{F},\mathsf{P}\big)$ as well.
In the sequel, we have to operate with $\varrho_n\pi_n(dx)$ as a random object defined on $(\Omega,\mathscr{F},\bar{\mathsf{P}})$. Since $\bar{\nu} \ll \nu$ is not assumed, π_n cannot be defined properly on $(\Omega,\mathscr{F},\bar{\mathsf{P}})$ by applying the previous arguments. However, the product $\varrho_n\pi_n$ is well defined on $(\Omega,\mathscr{F},\bar{\mathsf{P}})$. Indeed, let $\varGamma$ denote the set, where $\varrho_n\pi_n$ is well defined. Notice that $\varGamma \in \mathscr{F}^Y_n$ and so, $\bar{\mathsf{P}}(\varGamma) = \bar{\mathsf{P}}_n(\varGamma)$. Now, by the Lebesgue decomposition of $\bar{\mathsf{P}}_n$ with respect to P_n,

$$\bar{\mathsf{P}}_n(\varGamma) = \int_{\varGamma\cap\{\varrho_n>0\}} \varrho_n^{-1} d\mathsf{P}_n + \bar{\mathsf{P}}_n\big(\{\varrho_n = 0\}\cap\varGamma\big) \ge \int_{\varGamma\cap\{\varrho_n>0\}} \varrho_n^{-1} d\mathsf{P}_n.$$

Since both π_n and ϱ_n are defined P-a.s., $\mathsf{P}_n(\varGamma) = 1$ holds. Moreover, $\mathsf{P}_n(\varrho_n > 0) = 1$ since $\mathsf{P}_n(\varrho_n = 0) = \int_{\{\varrho_n=0\}} \varrho_n d\bar{\mathsf{P}}_n = 0$. Hence,

$$\int_{\varGamma\cap\{\varrho_n>0\}} \varrho_n^{-1} d\mathsf{P}_n = \int_\Omega \varrho_n^{-1} d\mathsf{P}_n = \int_\Omega \varrho_n^{-1}\varrho_n d\bar{\mathsf{P}}_n = 1,$$

that is, $\bar{\mathsf{P}}_n(\varGamma) = 1$.

For $g:\mathbb{R}\mapsto\mathbb{R}$ with $\mathsf{E}|g(Y_n)| < \infty$ and $\bar{\mathsf{E}}|g(Y_n)| < \infty$, let us define predicting estimates: $\eta_{n|n-1}(g) = \mathsf{E}\big(g(Y_n)|\mathscr{F}^Y_{n-1}\big)$ and $\bar{\eta}_{n|n-1}(g) = \bar{\mathsf{E}}\big(g(Y_n)|\mathscr{F}^Y_{n-1}\big)$. We fix the following versions of these conditional expectations

$$\begin{aligned}\eta_{n|n-1}(g) &= \int_{\mathbb{R}}\int_{\mathbb{R}} g(y)\gamma(x,y)\varphi(dy)\pi_{n-1}(dx)\\ \bar{\eta}_{n|n-1}(g) &= \int_{\mathbb{R}}\int_{\mathbb{R}} g(y)\gamma(x,y)\varphi(dy)\bar{\pi}_{n-1}(dx).\end{aligned}$$

Similarly to $\bar{\pi}_n$, the predictor $\bar{\eta}_{n|n-1}(g)$ is well defined P- and $\bar{\mathsf{P}}$-a.s. while only $\varrho_{n-1}\eta_{n|n-1}(g)$ makes sense with respect to both measures.

Theorem 2.1 *Assume $\nu \ll \bar{\nu}$ and any of the following conditions:*

i. *g is bounded*;

ii. $\dfrac{d\nu}{d\bar{\nu}}$ *is bounded and the family* $(g(Y_n))_{n\ge1}$ *is* $\bar{\mathsf{P}}$*-uniformly integrable*;

iii. *for* $p,q>1$, $\dfrac{1}{p}+\dfrac{1}{q}=1$, $\bar{\mathsf{E}}\Big(\dfrac{d\nu}{d\bar{\nu}}\Big)^p<\infty$ *and the family* $\big(|g(Y_n)|^q\big)_{n\ge1}$ *is* $\bar{\mathsf{P}}$*-uniformly integrable.*

Then,

$$\lim_{n\to\infty}\mathsf{E}\big|\eta_{n|n-1}(g)-\bar{\eta}_{n|n-1}(g)\big|=0. \tag{2.1}$$

Proof. Suppose that α is $\mathscr{F}^Y_n$-measurable random variable defined on $(\Omega,\mathscr{F},\mathsf{P})$ with $\mathsf{E}|\alpha|<\infty$. Then $\bar{\mathsf{E}}|\alpha|\varrho_n<\infty$ and

$$\varrho_{n-1}\mathsf{E}\big(\alpha|\mathscr{F}^Y_{n-1}\big)=\bar{\mathsf{E}}\big(\alpha\varrho_n|\mathscr{F}^Y_{n-1}\big),\quad \bar{\mathsf{P}}\text{-a.s.} \tag{2.2}$$

(i) For $\alpha:=g(Y_n)$, (2.2) reads:

$$\varrho_{n-1}\eta_{n|n-1}(g)=\bar{\mathsf{E}}\big(g(Y_n)\varrho_n|\mathscr{F}^Y_{n-1}\big),\quad \bar{\mathsf{P}}\text{-a.s.}$$

Therefore, (here $|g|\le C$ is assumed for definiteness)

$$\begin{aligned}
\mathsf{E}\big|\eta_{n|n-1}(g)-\bar{\eta}_{n|n-1}(g)\big| &= \bar{\mathsf{E}}\frac{d\mathsf{P}^Y_{n-1}}{d\bar{\mathsf{P}}^Y_{n-1}}\big|\eta_{n|n-1}(g)-\bar{\eta}_{n|n-1}(g)\big|\\
&=\bar{\mathsf{E}}\varrho_{n-1}\big|\eta_{n|n-1}(g)-\bar{\eta}_{n|n-1}(g)\big|\\
&=\bar{\mathsf{E}}\big|\varrho_{n-1}\eta_{n|n-1}(g)-\varrho_{n-1}\bar{\eta}_{n|n-1}(g)\big|\\
&=\bar{\mathsf{E}}\big|\bar{\mathsf{E}}\big(g(Y_n)\varrho_n|\mathscr{F}^Y_{n-1}\big)-\bar{\mathsf{E}}\big(g(Y_n)\varrho_{n-1}|\mathscr{F}^Y_{n-1}\big)\big|\\
&=\bar{\mathsf{E}}\Big|\bar{\mathsf{E}}\Big(g(Y_n)\big(\varrho_n-\varrho_{n-1}\big)\big|\mathscr{F}^Y_{n-1}\Big)\Big|\le C\bar{\mathsf{E}}|\varrho_n-\varrho_{n-1}|.
\end{aligned} \tag{2.3}$$

Since $(\varrho_n,\mathscr{F}^Y_n,\bar{\mathsf{P}})_{n\ge1}$ is a uniformly integrable martingale converging to $\varrho_\infty=\bar{\mathsf{E}}\Big(\dfrac{d\nu}{d\bar{\nu}}\big|\mathscr{F}^Y_\infty\Big)$, and $\bar{\mathsf{E}}|\varrho_n-\varrho_{n-1}|\le\bar{\mathsf{E}}|\varrho_n-\varrho_\infty|+\bar{\mathsf{E}}|\varrho_{n-1}-\varrho_\infty|$, the required result follows from $\lim_{n\to\infty}\bar{\mathsf{E}}|\varrho_n-\varrho_\infty|=0$ by the Scheffe theorem.

(ii) Set $g^C=gI_{\{|g|\le C\}}$, then by (**i**),

$$\lim_{n\to\infty}\mathsf{E}\big|\eta_{n|n-1}(g^C)-\bar{\eta}_{n|n-1}(g^C)\big|=0,\ \forall\ C>0.$$

and it is left to show that

$$\begin{aligned}
&\lim_{C\to\infty}\varlimsup_{n\to\infty}\mathsf{E}\big|\eta_{n|n-1}(g-g^C)\big|=0\\
&\lim_{C\to\infty}\varlimsup_{n\to\infty}\mathsf{E}\big|\bar{\eta}_{n|n-1}(g-g^C)\big|=0.
\end{aligned} \tag{2.4}$$

Let for definiteness $\dfrac{d\nu}{d\bar{\nu}}\le K$ and thus $\varrho_n\le K$, $\bar{\mathsf{P}}$-a.s. for all $n\ge1$. Then

$$\begin{aligned}
&\mathsf{E}\big|\eta_{n|n-1}(g-g^C)\big|\le\mathsf{E}|g(Y_n)|I_{\{|g(Y_n)|>C\}}\le K\bar{\mathsf{E}}|g(Y_n)|I\{|g(Y_n)|>C\}\\
&\mathsf{E}\big|\bar{\eta}_{n|n-1}(g-g^C)\big|=\bar{\mathsf{E}}\varrho_{n-1}\big|\bar{\eta}_{n|n-1}(g-g^C)\big|\le K\bar{\mathsf{E}}|g(Y_n)|I\{|g(Y_n)|>C\},
\end{aligned}$$

and (2.4) holds by the uniform integrability assumption from (**ii**).

(iii) By (2.3), it suffices to show that $\lim_{n\to\infty} \bar{\mathsf{E}}|g(Y_n)||\varrho_n - \varrho_{n-1}| = 0$. By the Hölder inequality we have

$$\bar{\mathsf{E}}|g(Y_n)||\varrho_n - \varrho_{n-1}| \le \Big(\bar{\mathsf{E}}|g(Y_n)|^q\Big)^{1/q}\Big(\bar{\mathsf{E}}|\varrho_n - \varrho_{n-1}|^p\Big)^{1/p}.$$

The $\bar{\mathsf{P}}$-uniform integrability of $(|g(Y_n)|^q)_{n\ge 0}$ provides $\sup_{n\ge 0} \bar{\mathsf{E}}|g(Y_n)|^q < \infty$. Since

$$\lim_{n\to\infty} \bar{\mathsf{E}}|\varrho_n - \varrho_{n-1}| = 0$$

it is left to check that the family $(|\varrho_n + \varrho_{n-1}|^p)_{n\ge 1}$ is $\bar{\mathsf{P}}$-uniformly integrable. This holds by the following upper bound

$$\bar{\mathsf{E}}|\varrho_n + \varrho_{n-1}|^p \le 2^{p-1}(\bar{\mathsf{E}}\varrho_n^p + \bar{\mathsf{E}}\varrho_{n-1}^p) \le 2^p \bar{\mathsf{E}}\Big(\frac{d\nu}{d\bar{\nu}}\Big)^p, \quad p \ge 1$$

where the Jensen inequality has been used. ■

Corollary 2.2 *Let f be a measurable function and assume that there is a function g solving* (1.7) *and satisfying the assumptions of Theorem 2.1. Then*

$$\lim_{n\to\infty} \mathsf{E}|\pi_n(f) - \bar{\pi}_n(f)| = 0.$$

Proof. Since $\pi_{n-1}(f) = \eta_{n|n-1}(g)$ and $\bar{\pi}_{n-1}(f) = \bar{\eta}_{n|n-1}(g)$, the claim is nothing but (2.1). ■

3 Examples

3.1 Hidden Markov Chains

Let X be a Markov chain taking values in a finite alphabet $\mathbb{S} = \{a_1, ..., a_d\}$ and the observation

$$Y_n = \sum_{j=1}^d \xi_n(j) I_{\{X_{n-1}=a_j\}},$$

where $\xi_n(j)$, $j = 1, \ldots, d$, are independent entries of the random vectors ξ_n, which form an i.i.d. sequence independent of X.
This variant of *Hidden Markov Model* is popular in various applications (see e.g. [12]) and its stability analysis has been carried out by several authors (see e.g. [3], [18], [4]) mainly for ergodic chain X. The nonlinear filter (1.1) is finite dimensional, namely, the conditional distribution $\pi_n(dx)$ is just the vector of conditional probabilities $\pi_n(i) = \mathsf{P}(X_n = a_i|\mathscr{F}_n^Y)$, $i = 1, ..., d$ and

$$\|\pi_n - \bar{\pi}_n\|_{\mathsf{tv}} = \sum_{i=1}^d |\pi_n(i) - \bar{\pi}_n(i)|.$$

The following holds regardless of the ergodic properties of X:

Proposition 3.1 *Assume*

a1. all atoms of $\bar\nu$ *are positive*

a2. $\mathsf{E}\big|\xi_1(j)\big|^i<\infty$, $i,j=1,\dots,d$

a3. the $d\times d$ *matrix* B *with the entries* $B_{ij}=\mathsf{E}\big(\xi_1(j)\big)^i$ *is nonsingular*

Then,

$$\lim_{n\to\infty}\mathsf{E}\big\|\pi_n-\bar\pi_n\big\|_{\mathsf{tv}}=0.$$

Proof. The condition (**ii**) of Theorem 2.1 is satisfied for any $g_i(y)=y^i$, $i=1,\dots,d$. Indeed, (a1) and (a2) imply $\dfrac{d\nu}{d\bar\nu}\le$ const. and the uniform integrability of $g_i(Y_n)$ for any i since $\bar{\mathsf{E}}\big|g_i(Y_n)\big|\le\sum_{j=1}^d\mathsf{E}\big|\xi_1(j)\big|^i<\infty$. Finally,

$$\eta_{n|n-1}(g_i)=\mathsf{E}\big((Y_n)^i|\mathscr{F}^Y_{n-1}\big)=\sum_{j=1}^d\pi_{n-1}(j)\mathsf{E}\big(\xi_1(j)\big)^i=\sum_{j=1}^d\pi_{n-1}(j)B_{ij}.$$

and, then, by Theorem 2.1,

$$\mathsf{E}\big|\eta_{n|n-1}(g_i)-\bar\eta_{n|n-1}(g_i)\big|=\mathsf{E}\Big|\sum_{j=1}^d\big(\pi_{n-1}(j)-\bar\pi_{n-1}(j)\big)B_{ij}\Big|\xrightarrow[n\to\infty]{}0.$$

The latter and the nonsingularity of B proves the claim. ■

3.2 Observations with multiplicative white noise

This example is borrowed from [13]. The signal process is defined by the linear recursive equation

$$X_n=aX_{n-1}+\theta_n,$$

where $|a|<1$ and $(\theta_n)_{n\ge1}$ is $(0,b^2)$-Gaussian white noise independent of X_0, that is, the signal process is ergodic. The distribution function ν has density $q(x)$ relative to dx from the *Serial Gaussian* (SG) family:

$$q(x)=\Big(\sum_{i=0}^\infty\alpha_i\frac{x^{2i}}{\sigma^{2i}C_{2i}}\Big)\frac{1}{\sigma\sqrt{2\pi}}\exp\Big(-\frac{x^2}{2\sigma^2}\Big),$$

where σ is the scaling parameter, α_i's are nonnegative weight coefficients, $\sum_{i\ge0}\alpha_i=1$ and C_{2i} are the normalizing constants. The observation sequence is given by

$$Y_n=X_{n-1}\xi_n,$$

where ξ_n is a sequence of i.i.d. random variables. The distribution function of ξ_1 is assumed to have the following density relative to dx:

$$p(x)=\frac{\rho}{|x|^3}\exp\Big(-\frac{\rho}{x^2}\Big),\quad p(0)=0,\tag{3.1}$$

where ρ is a positive constant. This filtering model is motivated by financial applications when $|X|$ is interpreted as the stochastic volatility parameter of an asset price.

As proved in [13], the filter (1.1) admits a finite dimensional realization provided that $\alpha_j \equiv 0$, $j > N$ for some integer $N \geq 1$, namely for any time $n \geq 1$ the filtering distribution $\pi_n(dx)$ has a density of SG type with the scaling parameter σ_n and the weights a_{in}, which are propagated by a finite (growing with n) set of recursive equations driven by the observations. Thus, the evolution of $\pi_n(dx)$ is completely determined via σ_n and α_{in}. Some stability analysis for the sequence $\big(\sigma_n, (\alpha_{in})_{i\geq 1}\big)_{n\geq 1}$ has been done in [14].
Assume that the density $q(x)$ of ν belongs to the SG family, but its parameters are unknown. If the filter is started from the Gaussian density with zero mean and variance $\bar{\sigma}^2$, the filtering equation remains finite dimensional and the density

$$\frac{d\nu}{d\bar{\nu}}(x) = \frac{q(x)}{\bar{q}(x)} = \Big(\sum_{i=0}^{N} \alpha_i \frac{x^{2i}}{\sigma^{2i} C_{2i}}\Big) \frac{\bar{\sigma}}{\sigma} \exp\Big(- \frac{x^2}{2}\Big(\frac{1}{\sigma^2} - \frac{1}{\bar{\sigma}^2}\Big)\Big)$$

is bounded, if $\bar{\sigma} > \sigma$.
In terms of the setting under consideration $\gamma(x,y) = \dfrac{1}{|x|} p(y/x)$, where $p(\cdot)$ is defined in (3.1) and $\varphi(dy) = dy$. For $f(x) := |x|$

$$\eta_{n|n-1}(f) = \mathsf{E}\big(f(Y_n)|\mathscr{F}^Y_{n-1}\big) = \pi_{n-1}(f)\mathsf{E}|\xi_1|,$$

where $\mathsf{E}|\xi_1| > 0$ and hence $g(y) = |y|/\mathsf{E}|\xi_1|$ solves (1.7). Finally, $\big(g(Y_n)\big)_{n\geq 1}$ is $\bar{\mathsf{P}}$-uniformly integrable family since

$$\bar{\mathsf{E}}|X_{n-1}\xi_n| I_{\{|X_{n-1}\xi_n|>C\}} \leq \frac{\bar{\mathsf{E}}|X_{n-1}\xi_n|^{1+\varepsilon}}{C^\varepsilon} = \frac{\bar{\mathsf{E}}|X_{n-1}|^{1+\varepsilon}\mathsf{E}|\xi_1|^{1+\varepsilon}}{C^\varepsilon},$$

where $\mathsf{E}|\xi_1|^{1+\varepsilon} < \infty$ if $\varepsilon \in [0,1)$ and $\sup_{n\geq 0} \bar{\mathsf{E}}|X_{n-1}|^{1+\varepsilon} < \infty$ is implied by $|a| < 1$ and $\int_{\mathbb{R}} |x|\bar{\nu}(dx) < \infty$. Thus, by Corollary 2.2

$$\lim_{n\to\infty} \mathsf{E}\Big| \int_{\mathbb{R}} |x|\pi_n(dx) - \int_{\mathbb{R}} |x|\bar{\pi}_n(dx)\Big| = 0.$$

3.3 Additive observation noise

Suppose

$$Y_n = h(X_{n-1}) + \xi_n, \tag{3.2}$$

where h is a fixed measurable function, $\xi = (\xi_n)_{n\geq 1}$ is an i.i.d. sequence of random variables independent of X. Since

$$\mathsf{E}\big(g(Y_n)|\mathscr{F}^Y_{n-1}\big) = \pi_{n-1}(h) + \mathsf{E}\xi_1$$

and if one of the integrability conditions in Theorem 2.1 is satisfied for $g(y) := y$, (1.2) holds true for h:

$$\lim_{n\to\infty} \mathsf{E}\big|\pi_n(h) - \bar{\pi}_n(h)\big| = 0. \tag{3.3}$$

Remark 3.2 (3.3) *resembles the result of J.M.C. Clark et al [9] in the continuous time setting: for a general Markov signal $X = (X_t)_{t\geq 0}$ and the observations $Y = (Y_t)_{t\geq 0}$ of the form* (1.3)*,*

$$\mathsf{E}\int_0^\infty \big(\pi_t(h) - \bar{\pi}_t(h)\big)^2 dt \leq 2\int_{\mathbb{R}} \log \frac{d\nu}{d\bar{\nu}}(x)\nu(dx).$$

This bound is verified information theoretical arguments.

3.3.1 Linear observations $\mathbf{h}(\mathbf{x}) \equiv \mathbf{x}$

Consider the linear observation model (3.2) with $h(x) = x$:

$$Y_n = X_{n-1} + \xi_n.$$

Proposition 3.3 *Assume*

A1. $\dfrac{d\nu}{d\bar{\nu}} \le c < \infty$

A2. X_n^p *is* $\bar{\mathsf{P}}$*-uniformly integrable for some* $p \ge 1$

A3. $\left|\mathsf{E}e^{i\xi_1 t}\right| > 0$ *for all* $t \in \mathbb{R}$

Then for any continuous function $f(x), x \in \mathbb{R}$, *growing not faster than a polynomial of order* p,

$$\lim_{n\to\infty} \mathsf{E}|\pi_n(f) - \bar{\pi}_n(f)| = 0.$$

Proof. If f is an unbounded function, it can be approximated by a sequence of bounded functions f_ℓ, $\ell \ge 1$ with $f_\ell(x) = g_\ell(f(x))$, where $g_\ell(x) = \begin{cases} x, & |x| \le \ell \\ \ell \operatorname{sign}(x), & |x| > \ell. \end{cases}$
Further, for $k = 1, 2, \ldots$, set

$$f_{\ell,k}(x) = \begin{cases} f_\ell(x), & |x| \le k-1 \\ \tilde{f}_{\ell,k}(x), & k-1 < |x| \le k \\ 0, & |x| > k, \end{cases}$$

where $\tilde{f}_{\ell,k}(x)$ is chosen so that the function $f_{\ell,k}(x)$ is continuous and

$$|f_{\ell,k}(x)| \le |f_\ell(x)|.$$

By the second Weierstrass approximating theorem (see e.g. [21]) one can choose a trigonometrical polynomial $P_{m,\ell,k}(x)$ such that for any positive number m,

$$\max_{x\in[-k,k]} \left|f_{\ell,k}(x) - P_{m,\ell,k}(x)\right| \le \frac{1}{m}.$$

Since $P_{m,\ell,k}(x)$ is a periodic function,

$$|P_{m,\ell,k}(x)| \le \frac{1}{m} + \max_{|y|\le k} |f_{\ell,k}(y)| \le \frac{1}{m} + \ell, \quad \text{for any } |x| > k.$$

Using the triangular inequality for

$$f = P_{m,\ell,k} + \big(f_{\ell,k} - P_{m,\ell,k}\big) + \big(f_\ell - f_{\ell,k}\big) + \big(f - f_\ell\big),$$

and the estimates

$$|f_{\ell,k} - P_{m,\ell,k}| \le \tfrac{1}{m} I_{\{|x|\le k\}} + \Big(\tfrac{1}{m} + \ell\Big) I_{\{|x|>k\}}$$

$$|f_\ell - f_{\ell,k}| \le \ell I_{\{|x|>k\}}$$

$$|f-f_\ell| \le C(1+|x|^p)I_{\{C(1+|x|^p)>\ell\}}, \text{ for some constant } C>0$$

we find the following upper bound

$$|f-P_{m,\ell,k}| \le \frac{1}{m}I_{\{|x|\le k\}} + \Big(\frac{1}{m}+2\ell\Big)I_{\{|x|>k\}} + C(1+|x|^p)I_{\{C(1+|x|^p)>\ell\}},$$

implying

$$\begin{aligned}\mathsf{E}\big|\pi_n(f)-\bar\pi_n(f)\big| \le \mathsf{E}\big|\pi_n(P_{m,\ell,k})-\bar\pi_n(P_{m,\ell,k})\big| + \frac{2}{m} + 2\ell\mathsf{E}\int_{\{|x|>k\}}[\pi_n(dx)+\bar\pi_n(dx)] \\ + C\mathsf{E}\int_{\{C(1+|x|^p)\ge\ell\}}(1+|x|^p)\pi_n(dx) + C\mathsf{E}\int_{\{C(1+|x|^p)\ge\ell\}}(1+|x|^p)\bar\pi_n(dx).\end{aligned}$$

Thus, the desired result holds by arbitrariness of m provided that

(1). $\lim\limits_{n\to\infty}\mathsf{E}\big|\pi_n(P_{m,\ell,k})-\bar\pi_n(P_{m,\ell,k})\big|=0,\ \forall\ m,\ell,k;$

(2). $\lim\limits_{k\to\infty}\overline{\lim\limits_{n\to\infty}}\ 2\ell\mathsf{E}\int_{\{|x|>k\}}[\pi_n(dx)+\bar\pi_n(dx)]=0,\ \forall\ \ell;$

(3). $\lim\limits_{\ell\to\infty}\overline{\lim\limits_{n\to\infty}}\ \mathsf{E}\int_{\{C(1+|x|^p)\ge\ell\}}(1+|x|^p)[\pi_n(dx)+\bar\pi_n(dx)]=0;$

(1) holds due to $\mathsf{E}\big(e^{itY_n}\big|\mathscr{F}^Y_{n-1}\big)=\mathsf{E}\big(e^{iX_{n-1}t}\big|\mathscr{F}^Y_{n-1}\big)\mathsf{E}e^{it\xi_1}=\pi_{n-1}(e^{itx})\mathsf{E}e^{it\xi_1}$ and the assumption (A3) since, by Theorem 2.1,

$$\lim_{n\to\infty}\mathsf{E}\big|\pi_n\big(e^{itx}\big)-\bar\pi_n\big(e^{itx}\big)\big|=0, \quad \forall t\in\mathbb{R}.$$

(2) is implied by the Chebyshev inequality

$$\mathsf{E}\int_{\{|x|>k\}}[\pi_n(dx)+\bar\pi_n(dx)] \le \frac{1}{k}\bar{\mathsf{E}}\Big(1+\frac{d\nu}{d\bar\nu}(X_0)\Big)|X_n|,$$

and the assumptions (A1) and (A2).

(3) follows from

$$\mathsf{E}\int_{\{C(1+|x|^p)\ge\ell\}}(1+|x|^p)[\pi_n(dx)+\bar\pi_n(dx)] = \bar{\mathsf{E}}I_{\{C(1+|X_n|^p)\ge\ell\}}\Big(1+\frac{d\nu}{d\bar\nu}(X_0)\Big)\big(1+|X_n|^p\big)$$

and the assumptions (A1) and (A2). ■

Acknowledgement.

We are grateful to Valentine Genon-Catalot for bringing [13], [14] to our attention.

References

[1] R. Atar, Exponential stability for nonlinear filtering of diffusion processes in a noncompact domain, Ann. Probab. **26** (1998), no. 4, 1552–1574. MR1675039 (99k:93088)

[2] R. Atar and O. Zeitouni, Exponential stability for nonlinear filtering, Ann. Inst. H. Poincaré Probab. Statist. **33** (1997), no. 6, 697–725. MR1484538 (98i:60070)

[3] R. Atar and O. Zeitouni, Lyapunov exponents for finite state nonlinear filtering, SIAM J. Control Optim. **35** (1997), no. 1, 36–55. MR1430282 (97k:93065)

[4] P. Baxendale, P. Chigansky and R. Liptser, Asymptotic stability of the Wonham filter: ergodic and nonergodic signals, SIAM J. Control Optim. **43** (2004), no. 2, 643–669 (electronic). MR2086177 (2005e:93152)

[5] D. Blackwell, The entropy of functions of finite-state Markov chains, in *Transactions of the first Prague conference on information theory, Statistical decision functions, random processes held at Liblice near Prague from November 28 to 30, 1956*, 13–20, Publ. House Czech. Acad. Sci., Prague. MR0100297 (20 #6730)

[6] A. Budhiraja and D. Ocone, Exponential stability in discrete-time filtering for non-ergodic signals, Stochastic Process. Appl. **82** (1999), no. 2, 245–257. MR1700008 (2000d:94010)

[7] A. Budhiraja and D. Ocone, Exponential stability of discrete-time filters for bounded observation noise, Systems Control Lett. **30** (1997), no. 4, 185–193. MR1455877 (98c:93110)

[8] P. Chigansky and R. Liptser, Stability of nonlinear filters in nonmixing case, Ann. Appl. Probab. **14** (2004), no. 4, 2038–2056. MR2099662 (2005h:62265)

[9] J. M. C. Clark, D. L. Ocone and C. Coumarbatch, Relative entropy and error bounds for filtering of Markov processes, Math. Control Signals Systems **12** (1999), no. 4, 346–360. MR1728373 (2001m:60095)

[10] B. Delyon and O. Zeitouni, Lyapunov exponents for filtering problems, in *Applied stochastic analysis (London, 1989)*, 511–521, Gordon and Breach, New York. MR1108433 (92i:93092)

[11] P. Del Moral and A. Guionnet, On the stability of interacting processes with applications to filtering and genetic algorithms, Ann. Inst. H. Poincaré Probab. Statist. **37** (2001), no. 2, 155–194. MR1819122 (2002k:60013)

[12] Y. Ephraim and N. Merhav, Hidden Markov processes, IEEE Trans. Inform. Theory **48** (2002), no. 6, 1518–1569. MR1909472 (2003f:94024)

[13] V. Genon-Catalot, A non-linear explicit filter, Statist. Probab. Lett. **61** (2003), no. 2, 145–154. MR1950665 (2004a:60085)

[14] V. Genon-Catalot and M. Kessler, Random scale perturbation of an AR(1) process and its properties as a nonlinear explicit filter, Bernoulli **10** (2004), no. 4, 701–720. MR2076070 (2005g:60111)

[15] T. Kaijser, A limit theorem for partially observed Markov chains, Ann. Probability **3** (1975), no. 4, 677–696. MR0383536 (52 #4417)

[16] H. Kunita, Asymptotic behavior of the nonlinear filtering errors of Markov processes, J. Multivariate Anal. **1** (1971), 365–393. MR0301812 (46 #967)

[17] F. LeGland and N. Oudjane, A robustification approach to stability and to uniform particle approximation of nonlinear filters: the example of pseudo-mixing signals, Stochastic Process. Appl. **106** (2003), no. 2, 279–316. MR1989630 (2004i:93184)

[18] F. Le Gland and L. Mevel, Exponential forgetting and geometric ergodicity in hidden Markov models, Math. Control Signals Systems **13** (2000), no. 1, 63–93. MR1742140 (2001b:93075)

[19] R. S. Liptser and A. N. Shiryaev, *Statistics of random processes. II*, Translated from the 1974 Russian original by A. B. Aries, Second, revised and expanded edition, Springer, Berlin, 2001. MR1800858 (2001k:60001b)

[20] D. Ocone and E. Pardoux, Asymptotic stability of the optimal filter with respect to its initial condition, SIAM J. Control Optim. **34** (1996), no. 1, 226–243. MR1372912 (97e:60073)

[21] G. Szegő, *Orthogonal polynomials*, Fourth edition, Amer. Math. Soc., Providence, R.I., 1975. MR0372517 (51 #8724)