Annales Mathematicae Silesianae **27** (2013), 39–58
Prace Naukowe Uniwersytetu Śląskiego nr 3106, Katowice

ERGODICITY OF FILTERING PROCESSES: THE HISTORY OF A MISTAKE AND ATTEMPTS TO CORRECT IT

Łukasz Stettner

2013 Annual Lecture dedicated to the memory of Professor Andrzej Lasota

Abstract. The paper describes briefly a history of filtering problems of Markov processes and then concentrates on ergodic properties of filtering process. A mistake in a famous Kunita paper on ergodicity of filtering processes is shown. Then the paper reviews various attempts trying to correct this mistake.

1. Introduction

Let (x_n), (y_n) be discrete time processes on a given probability space $(\Omega, \mathcal{F}, P)$ taking values in the Polish spaces E_0 and E respectively. The process (x_n), called the state process or hidden Markov process is characterized by the transition kernel $P(x, dx')$, while the process (y_n), called frequently the observation process has a transition kernel $P_1^{x'}(y, dy')$ parametrized by the current value x' of the hidden Markov process. We do not observe the state process (x_n), we observe the process (y_n), which can be considered

Received: 30.03.2013. Revised: 24.09.2013.
(2010) Mathematics Subject Classification: 93E11, 60G35.
Key words and phrases: Markov processes, partial observation, filtering process, invariant measures.
Research supported by NCN grant DEC-2012/07/B/ST1/03298.

as a noisy observation of (x_n) but also can have its own dynamics depending on the current value of the state process (x_n). To be more precise, for $X^n := \sigma\{x_0, x_1, \ldots, x_n\}$ and $Y^n := \sigma\{y_0, y_1, \ldots, y_n\}$, $n = 1, 2, \ldots$, we assume that P a.s.

$$P\{x_{n+1} \in A \mid X^n, Y^n\} = P(x_n, A), \tag{1.1}$$

$$P\left\{y_{n+1} \in B \mid X^{n+1}, Y^n\right\} = P_1^{x_{n+1}}(y_n, B), \tag{1.2}$$

where for fixed x, x', y, $P(x, \cdot)$ and $P_1^{x'}(y, \cdot)$ are probability measures on E_0 and E respectively and for fixed Borel subsets A of E_0 and B of E the mappings $x \mapsto P(x, A)$ and $(x', y) \mapsto P_1^{x'}(y, B)$ are $\mathcal{B}(E_0)$ and $\mathcal{B}(E_0 \times E)$ measurable, where $\mathcal{B}$ denote the suitable Borel σ-fields.

In what follows we shall assume the following particular form of the kernel P_1,

$$P_1^{x'}(y, B) = \int_B r(x', y, y')\, \eta(dy'), \tag{1.3}$$

where $\eta \in \mathcal{P}(E)$, with $\mathcal{P}(E)$ denoting the space of probability measures on E. This form is very important:the representation (1.3) means that there is a reference probability measure η which is independent on x' and y, i.e., on the initial state of the observation process (y_n) and the value of the state process in the next time step.

Notice that the above form of observation kernel covers models of the form

$$y_{n+1} = h(y_n, x_{n+1}, w_{n+1})$$

(in particular $y_{n+1} = h(y_n, x_{n+1}) + g(y_n, x_{n+1}) w_{n+1}$) with $h(y_n, x_{n+1}, \cdot)$ a C^1 diffeomorphism of R^d (in particular matrix $g(y_n, x_{n+1})$ being invertible) and w_{n+1} independent of X^{n+1}, Y^n and identically distributed with law $\eta(dy)$. Furthermore the form (1.3) is satisfied also in the case of denumerable observation space E, which is frequent in the practice.

Directly from (1.1) and (1.2) we have

LEMMA 1.1. *The pair* $\begin{pmatrix} x_n \\ y_n \end{pmatrix}$ *forms a Markov process with transition operator*

$$Tf(x, y) = \int_{E_0} \int_E f(x', y')\, P_1^{x'}(y, dy') P(x, dx') \tag{1.4}$$

for $f \in b\mathcal{B}(E_0 \times E)$, *namely the space of bounded Borel measurable functions on* $E_0 \times E$.

The process (x_n) is observable only by means of the observation process (y_n). In the case when dynamics of x_{n+1} and y_{n+1} is linear with respect to x_n with independent additive Gaussian noises and parameters dependent in nonlinear way on y_n we have so called conditional Gaussian model studied in [14], which means that conditional law of x_n given Y^n is Gaussian with conditional expected value and conditional variance, which can be derived recursively. In general nonlinear case we have nonlinear filtering problem. To describe its evolution we shall need the following family of indexed probability measures $(y, y' \in E,\ \nu \in \mathcal{P}(E_0),\ A \in \mathcal{B}(E_0))$

$$M(y, y', \nu)(A) = \frac{\int_A r(x', y, y') \int_{E_0} P(x, dx')\, \nu(dx)}{\int_{E_0} r(x', y, y') \int_{E_0} P(x, dx')\, \nu(dx)}.$$

In what follows we shall use the notation

$$P(\nu, dx') = \int_{E_0} P(x, dx')\, \nu(dx).$$

Moreover we shall assume that

$$M(y, y', \nu)(A) = 0 \quad \text{whenever} \quad \int_{E_0} r(x', y, y')\, P(\nu, dx') = 0.$$

Given the initial measure $\rho \in \mathcal{P}(E_0 \times E)$ we have the following regular conditional probability decomposition $\rho(dx, dy) = p_\rho(y, dx)\rho(E_0, dy)$ (see [8], Thm. I.3.1), where $p_\rho(y, dx)$ stands for conditional law of x given fixed y, when joint law of (x, y) is given by ρ. We define recursively the following measure valued process

$$\begin{aligned} \pi_0^\rho(A) &= p_\rho(y_0, A) \\ \pi_n^\rho(A) &= M\left(y_{n-1}, y_n, \pi_{n-1}^\rho\right)(A) \end{aligned} \tag{1.5}$$

for $A \in \mathcal{B}(E_0)$, $n = 1, 2, \ldots$, which we call *the filtering process.*

We have that

LEMMA 1.2. *For $A \in \mathcal{B}(E_0)$ we have*

$$\pi_n^\rho(A) = P\{x_n \in A \mid Y^n\} \quad P \ a.s. \tag{1.6}$$

Proof. For $n=0$ the claim directly follows from the definition of regular conditional probability. For $n>0$ following the proof of Lemma 1.1 of [18] on the set

$$G := \{\omega : \int_{E_0} r(x', y_{n-1}, y_n) P(\pi^\rho_{n-1}, dx') > 0\}$$

or equivalently

$$G = \{\omega : F(y_0, y_1, \ldots, y_n)(\omega) > 0\}$$

with

$$F(y_0, y_1, \ldots, y_n) = \int_{E_0} r(x', y_{n-1}, y_n) P(\pi^\rho_{n-1}, dx')$$

we obtain that $\pi^\rho_n(A) = P\{x_n \in A \mid Y^n\}$ for all $A \in \mathcal{B}(E_0)$. On the other hand

$$\begin{aligned} P(G^c) &= E\left\{1_{F(y_0,y_1,\ldots,y_n)(\omega)=0} 1_{E_0}(x_n)\right\} \\ &= E\left\{\int_{E_0}\int_E 1_{F(y_0,y_1,\ldots,y)=0} P_1^{x'}(y_{n-1}, dy) P(\pi^\rho_n, dx')\right\} = 0 \end{aligned}$$

so that we finally have (1.6). □

It can be easily seen that we have

Lemma 1.3. *The pair* $\binom{\pi^\rho_n}{y_n}$ *forms a Markov process on* $\mathcal{P}(E_0) \times E$ *with transition operator*

$$\Pi F(\nu, y) = \int_{E_0}\int_E F\left(M\left(y, y', \nu\right), y'\right) P_1^x\left(y, dy'\right) P(\nu, dx) \tag{1.7}$$

for $F \in b\mathcal{B}(\mathcal{P}(E_0) \times E)$.

Limit behaviour of the functionals of Markov processes letting time n to ∞ is described using invariant measures. By an invariant measure for a Markov process we mean a measure such that, whenever Markov processes starts with such measure, at each time n its law is the same and coincides with that invariant measure. Although the existence of invariant measure is important the crucial fact is its uniqueness, since then limits of the functionals of this Markov process are uniquely defined. In what follows the existence of unique invariant measure we call ergodicity of Markov process. In this paper we are interested in the ergodicity of the pair $\binom{\pi^\rho_n}{y_n}$, that is in the uniqueness of invariant measures for the operator Π. Such problem has been extensively studied for partially observed Markov processes, when the function r in (1.3) does not depend on y. In this case the process (π_n) is itself a Markov process. In the famous paper [12] in the continuous-time setting, with compact state space, necessary and sufficient condition for the existence of unique invariant measure of the process (π_n) was formulated. Later on this result was extended to locally compact metric space and continuous and discrete time models in [20] and [13]. This result can be formulated as follows

THEOREM 1.4. *Suppose there is a unique invariant measure μ for the state process (x_n) on a locally compact metric space E. Then filtering process (π_n) admits exactly one invariant measure if and only if*

$$\limsup_{n\to\infty} \int |E_x\{f(x_n)\} - \mu(f)|\mu(dx) = 0.$$

Theorem 1.4 appeared to be not correct. In the paper [2] it has been pointed out a gap in [12] and a counterexample was formulated; a discrete-time version of it will be presented below (this counterexample in fact appeared first in [9]).

EXAMPLE. Let $E_0 = \{1,2,3,4\}$, $E = \{0,1\}$ and the transition matrix of (x_n) is given by

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{bmatrix}.$$

Let $E_{01} = \{1,3\}$, $E_{02} = \{2,4\}$ and assume that $r(x', y, y')$ does not depend on y and

$$r(x,1) = \begin{cases} 2 & \text{for } x \in E_{01} \\ 0 & \text{for } x \in E_{02} \end{cases} \qquad r(x,0) = \begin{cases} 2 & \text{for } x \in E_{02} \\ 0 & \text{for } x \in E_{01} \end{cases}$$

with $\eta(0) = \eta(1) = \frac{1}{2}$.

In other words the observation process can be described by

$$y_n = 1_{E_{01}}(x_n).$$

Notice that $\forall_{x \in E_0}\ P(x, E_{01}) = \frac{1}{2}$ and (y_n) is a sequence of i.i.d. random variables with $P\{y_n = 0\} = P\{y_n = 1\} = \frac{1}{2}$.

Then, for $\alpha \in (0,1)$ let

$$e_1^\alpha = \begin{bmatrix} \alpha \\ 0 \\ 1-\alpha \\ 0 \end{bmatrix}, \quad e_2^\alpha = \begin{bmatrix} 0 \\ \alpha \\ 0 \\ 1-\alpha \end{bmatrix}, \quad e_3^\alpha = \begin{bmatrix} 1-\alpha \\ 0 \\ \alpha \\ 0 \end{bmatrix}, \quad e_4^\alpha = \begin{bmatrix} 0 \\ 1-\alpha \\ 0 \\ \alpha \end{bmatrix}$$

and notice that starting from $\pi_0 = e_1^\alpha$ the process (π_n) will cyclically move through the points $e_2^\alpha, e_3^\alpha, e_4^\alpha, e_1^\alpha, \ldots$, changing its state at each time with probability $\frac{1}{2}$ (and remaining in the same state also with probability $\frac{1}{2}$). Consequently the set $\{e_1^\alpha, e_2^\alpha, e_3^\alpha, e_4^\alpha\}$ is invariant and the uniformly distributed measure on it is invariant for (π_n). Replacing α with β such that $\alpha \neq \beta$ and $\alpha \neq 1-\beta$ we obtain different invariant set and measure respectively. Therefore, in front of nice ergodic properties of the process (x_n) the filter (π_n) admits a continuum of invariant measures with disjoint supports. This behavior is possibly due to the singular structure of the observation process.

Where was the gap? In the paper stationary flows were studied and we had the following σ fields $Y_{-\infty}^n = \sigma\{\ldots, y_{-n}, y_{-n+1}, \ldots, y_0, y_1, \ldots, y_n\}$, $X_{-\infty}^m = \sigma\{\ldots, x_{-n}, x_{-n+1}, \ldots, \ldots, x_m\}$ and the following limit was studied for a bounded measurable function ϕ

$$\lim_{m\to-\infty} E\left\{\phi(x_n)|Y_{-\infty}^n \vee X_{-\infty}^m\right\} \tag{1.8}$$

with $X_{-\infty}^{-\infty} = \{\emptyset, \Omega\}$. By Levy theorem (see Theorem 6.23 in [11]) the limit in (1.8) exists. However the limit σ field is different than $Y_{-\infty}^n$. Consequently

$$\lim_{m\to-\infty} E\left\{\phi(x_n)|Y_{-\infty}^n \vee X_{-\infty}^m\right\} \neq E\left\{\phi(x_n)|Y_{-\infty}^n\right\}.$$

On the other hand, when we change in our example the observation structure to $y_n = 1_{E_{01}}(x_n) + w_n$ with (w_n) a sequence of i.i.d. standard normal random variables, then one can show that there a unique invariant measure for (π_n). This led to the following conjecture formulated in [5]:

Conjecture. Existence of a unique invariant measure for (x_n) (of the pair (x_n, y_n) in general case) plus equivalence of observation transition operators $\int r(x,y)\eta(dy)$ $(P_1^x(y,\cdot)$ in general case) for $x \in E_0$ implies the existence of a unique invariant measure for (π_n) ($\begin{pmatrix} \pi_n^\rho \\ y_n \end{pmatrix}$ in general case).

As we shall see later this conjecture appeared to be false. In the paper we shall review a number of necessary and sufficient results for the existence of unique invariant measure to the pair $\begin{pmatrix} \pi_n^\rho \\ y_n \end{pmatrix}$. Sufficient results will be based on so called Hilbert metric and are presented in [5] or on vanishing discount approach from the paper [6]. A number of equivalent results will be based on [5] and references therein. In the final section we present recent results obtained by R. Van Handel as well as formulate further open problems.

2. Asymptotic stability of filtering processes using Hilbert metric

The filtering process (π_n^ρ) defined in (1.5) is a process of conditional expected values whenever we know the initial conditional law $p_\rho(y_0, \cdot)$. Usually we don't know it. Given any law $\rho' \in \mathcal{P}(E_0 \times E)$, which is not necessarily the real initial law ρ, we can construct recursively a sequence of measures using the operator M, namely for $A \in \mathcal{B}(E_0)$ we define

$$\pi_0^{\rho\rho'}(A) = p_{\rho'}(y_0, A)$$
$$\pi_{n+1}^{\rho\rho'}(A) = M\left(y_n, y_{n+1}, \pi_n^{\rho\rho'}\right)(A)$$

where $\rho'(dx, dy) = p_{\rho'}(y, dx)\rho'(E_0, dy)$. We call the process $\pi_n^{\rho\rho'}$ *approximate filtering process*. Clearly, it does not have conditional law representation (1.6). We say that we have *asymptotical stability* of approximate filtering processes whenever for any $\rho, \rho_1, \rho_2 \in \mathcal{P}(E_0 \times E)$ we have

$$\pi_n^{\rho\rho_1}(f) - \pi_n^{\rho\rho_2}(f) \to 0$$

in P_ρ probability, as $n \to \infty$, for $f \in C(E_0)$ - the space of continuous bounded functions on E_0, where P_ρ is conditional probability given initial law ρ of the process (x_n, y_n) and where

$$\pi_n^{\rho\rho_1}(f) := \int_{E_0} f(y)\pi_n^{\rho\rho_1}(dy).$$

Notice that for $\rho_1 = \rho$ the approximate filtering process $(\pi_n^{\rho\rho})$ coincides with real filtering process (π_n^ρ), so that the asymptotical stability means that no matter what is the initial law for large time we are close to the real filtering process. In this section we shall formulate sufficient conditions for asymptotical stability using so called Hilbert norm.

Let $\mathcal{M}(E_0)$ be the space of finite measures on E_0. For $\mu, \nu \in \mathcal{M}(E_0)$ define

$$h(\mu, \nu) := \sup_{A,B \in \mathcal{B}(E_0), \mu(B), \nu(A) > 0} \ln \frac{\mu(A)\nu(B)}{\mu(B)\nu(A)}.$$

One can notice that (see [5] and references therein) h is a pseudonorm in $\mathcal{M}(E_0)$ and has the following equivalent representation

$$h(\mu, \nu) = \ln \frac{\alpha(\mu, \nu)}{\beta(\mu, \nu)}$$

with

$$\alpha(\mu, \nu) = \inf \{a : a\mu \geq \nu\}, \quad \beta(\mu, \nu) = \sup \{b : b\mu \leq \nu\}.$$

Furthermore if L is a linear transformation preserving order in $\mathcal{M}(E_0)$ then (see Theorem 1.1 of [15])

$$h(L\mu, L\nu) \leq \tanh\left(\frac{\Delta}{4}\right) h(\mu, \nu) \tag{2.1}$$

with $\Delta = \sup_{\mu,\nu} h(L\mu, L\nu)$.

When $\mu, \nu \in \mathcal{P}(E_0)$ then (see Theorem 2.2 of [3])

$$\|\mu - \nu\|_{var} \leq \frac{2}{\ln 2} h(\mu, \nu) \tag{2.2}$$

where $\|\cdot\|_{var}$ stands for total variation norm in $\mathcal{P}(E_0)$. We have (for the proof see Theorem 1 of [5])

Theorem 2.1. *If for $k = 1$ we have*

(A1) $\sup_{x,x' \in E_0} h\big(P^k(x, \cdot), P^k(x', \cdot)\big) < \infty$

and for $k > 1$ we have (A1) and

(A2) *there exist continuous functions $\underline{r}(y, y')$, $\overline{r}(y, y')$, $\overline{\overline{r}}(y')$ such that for all $x \in E_0$, $y, y' \in E$, we have $0 < \underline{r}(y, y') \leq r(x, y, y') \leq \overline{r}(y, y') \leq \overline{\overline{r}}(y')$, and $\sum_{i=1}^{k-1} E_\rho \left\{\ln \frac{\overline{r}(y_{i-1}, y_i)}{\underline{r}(y_{i-1}, y_i)}\right\} < \infty$ and*

$$\int_E \int_E \cdots \int_E \int_E \underline{r}(y(k-2), y(k-1))\, \eta\big(dy(k-1)\big)$$
$$\overline{r}\big(y(k-3), y(k-2)\big)\eta\big(dy(k-2)\big) \ldots \overline{r}\big(y(0), y(1)\big)$$
$$\eta\big(dy(1)\big)\overline{\overline{r}}\big(y(0)\big)\eta\big(dy(0)\big) < \infty.$$

then for any $\rho_1, \rho_2 \in \mathcal{P}(E_0 \times E)$ *we have*

$$E_\rho \left[h\left(\pi_n^{\rho\rho_1}, \pi_n^{\rho\rho_2}\right)\right] \to 0 \tag{2.3}$$

as $n \to \infty$.

By (2.2) we see that the convergence (2.3) holds even in variation norm so that asymptotically approximate filtering processes coincide with real filtering processes, which is a desired property in practise. The assumption (A1) is rather strong and it says that k-th iterations of transition probabilities of the state process (x_n) with different initial states should be mutually equivalent with transition densities bounded from above. Practically the conditions of the Theorem are frequently satisfied whenever the state space E_0 is compact. The proof of Theorem 2.1 is based on (2.1) with operator L defined by the numerator of the operator M. We can relax assumptions of Theorem 2.1 if we require more about transition operator T (defined in (1.4)). We have (see Proposition 1 of [5])

Proposition 2.2. *If for* $k = 1$ *we have* (B1) *and* (B2) *or for* $k > 1$ *we have* (B1)–(B3), *where*

(B1) $\exists_{k \in N}$ *such that*

$$\sup_{x,x' \in E_0} \sup_{y,y' \in E} h\left(T^k(x, y, \cdot), T^k\left(x', y', \cdot\right)\right) < \infty,$$

(B2) *there exist continuous functions* $\underline{r}(y, y')$, $\overline{r}(y, y')$ *such that for each* $x \in E_0$

$$0 < \underline{r}\left(y, y'\right) \le r\left(x, y, y'\right) \le \overline{r}\left(y, y'\right)$$

and for k *as in* (B1) *we have* $\sum_{i=1}^{k-1} E_\rho \left\{\ln \frac{\overline{r}(y_{i-1}, y_i)}{\underline{r}(y_{i-1}, y_i)}\right\} < \infty,$

(B3) *for* $f_1 \in C(E_0)$, $f_2 \in C(E)$ *the mappings*

$$x \mapsto P f_1(x) \quad \text{and} \quad (x, y) \mapsto P_1^x f_2(y)$$

are continuous,

then for any measures $\rho_1, \rho_2 \in \mathcal{P}(E_0 \times E)$, *we have*

$$E_\rho \left[h\left(\pi_n^{\rho\rho_1}, \pi_n^{\rho\rho_2}\right)\right] \to 0,$$

whenever $n \to \infty$.

3. Ergodic properties of approximate filtering processes

Approximate filtering process $(\pi_n^{\rho,\rho'})$ is not a Markov process. To have Markov property we have to consider the triple $(x_n, y_n, \pi_n^{\rho\rho'})$, which is Markov with transition operator S defined for $F \in b\mathcal{B}(E_0 \times E \times \mathcal{P}(E_0))$ by the formula $SF(x,y,\nu) = \int_{E_0} \int_E F(x', y', M(y, y', \nu)) P_1^{x'}(y, dy') P(x, dx')$. Denote by $\pi^{\mu_0\eta_0\nu}$ or $\pi^{xy\nu}$ the approximate filtering processes $\pi^{\rho\rho'}$ with $\rho = \mu_0 \times \eta_0$ and $\rho' = \nu \times \eta_0$ or $\rho = \delta_x \times \delta_y$ and $\rho' = \nu \times \delta_y$, which mean that we know initial law η_0 or initial value y of (y_n) but don't know the real initial law μ_0 or initial value x of (x_n). We say that approximate filtering process $\pi^{\mu_0\eta_0\nu}$ is *asymptotically stable in probability* at (μ_0, η_0) if for any $\nu_1, \nu_2 \in \mathcal{P}(E_0)$ and $\varphi \in C(E_0)$ we have $\pi_n^{\mu_0\eta_0\nu_1}(\varphi) - \pi_n^{\mu_0\eta_0\nu_2}(\varphi) \to 0$, in $P_{\mu_0\eta_0}$ probability, as $n \to \infty$.

Ergodic behaviour of the triple $(x_n, y_n, \pi_n^{\rho\rho'})$, or the transition operator S is important to characterize the ergodic properties of the transition operator Π of the pair $\begin{pmatrix} \pi_n^{\rho} \\ y_n \end{pmatrix}$. We have the following (see Theorem 2 of [5])

THEOREM 3.1. *Assume that there exists a unique invariant measure $\zeta(dx, dy)$ for the transition operator T and that the approximate filtering processes $(\pi_n^{xy\nu})$ are asymptotically stable in probability at (x, y) for ζ almost all (x, y). Then there is at most one invariant measure for the transition operator S.*

Whenever transition operator transforms the class of continuous bounded functions into itself we say that it has *Feller property.* This property is frequently required when we want to use weak convergence technics to study ergodic properties of transition operators. Invariant measures for the operator S exist in many situations, as one can see in the following (see Proposition 2 and Corollary 2 of [5])

PROPOSITION 3.2. *If the operator S is Feller and there is an invariant measure ζ of the operator T, then there exists an invariant measure for S. Moreover if the operator Π is Feller and there is an invariant measure ζ of the operator T, then there exists an invariant measure for Π.*

Consequently if we have a unique invariant measure for the operator S then we also have a unique invariant measure for the operator Π, which is a restriction of the operator S to the second and third variables with approximate filtering processes replaced by real filtering process. By Theorem 3.1 we have at most one invariant measure for S. Under Feller property of the operator S and assumptions of Theorem 3.1, using Proposition 3.2 there exists exactly one invariant measure for the operator S and consequently for the operator Π. For the purpose of further analysis of ergodicity it will be important to introduce the notion of the barycenter of a measure. Given a measure $\Phi \in \mathcal{P}(\mathcal{P}(E_0) \times E)$ we define its *barycenter* $b\Phi$, for $A \in \mathcal{B}(E_0)$ and $B \in \mathcal{B}(E)$, as

$$b\Phi(A \times B) = \int_{\mathcal{P}(E_0)} \nu(A)\Phi(d\nu, B).$$

Clearly $b\Phi \in \mathcal{P}(E_0 \times E)$. Moreover we have (see Lemma 5 and Theorem 3 of [5])

LEMMA 3.3. *If Φ invariant for Π, then $b\Phi$ invariant for T.*

and

PROPOSITION 3.4. *If S is Feller and does not admit more than one invariant measure, then the operator Π has at most one invariant measure.*

We can now summarize Theorem 3.1, Propositions 3.2 and 3.4

COROLLARY 3.5. *If S is Feller, there exists a unique invariant measure ζ for T and for ζ almost an (x, y) the approximate filters $(\pi_n^{xy\rho'})$ are asymptotically stable in probability at (x, y), then there exist unique invariant measures for the operators S and Π.*

Now we introduce an order within the class of probability measures on $\mathcal{P}(E_0) \times E$ which contains potential invariant measures of the operator Π. Let $C_c(\mathcal{P}(E_0) \times E)$ be the family of functions $\mathcal{P}(E_0) \times E \ni (\nu, y) \mapsto F(\nu, y)$ which are continuous and bounded and convex with respect to ν for fixed $y \in E$. Ordering on $\mathcal{P}(\mathcal{P}(E_0) \times E)$ is defined as follows

$$q_1 \prec q_2 \quad \text{if and only if} \quad \forall_{f \in C_c(\mathcal{P}(E_0)\times E)} \quad q_1(f) \leq q_2(f).$$

One can consider following [12] the following two filtering processes:

$$\tag{3.1} \begin{aligned} \tilde{\pi}_n^\rho(A) &= P_\rho \left\{ x_n \in A | x_0 \vee Y^n \right\}, \\ \pi_n^\rho(A) &= P_\rho \left\{ x_n \in A | Y^n \right\} \end{aligned}$$

defined for $A \in \mathcal{B}(E_0)$. The first process $\tilde{\pi}_n^\rho$ corresponds to the situation when we know exactly the initial value x_0 of the process (x_n), although later we observe only the process (y_n). The second process π_n^ρ corresponds to the model in which we don't know x_0, and observe only y_0 at time 0. It appears that the pairs $\begin{pmatrix}\tilde{\pi}_n^\rho \\ y_n\end{pmatrix}$ and $\begin{pmatrix}\pi_n^\rho \\ y_n\end{pmatrix}$ form Markov processes on $\mathcal{P}(E_0) \times E$ with the same transition operator Π.

Define the following family of measures for $F \in b\mathcal{B}(\mathcal{P}(E_0) \times E)$:

$$m_n^\rho(F) = E_\rho \left\{F(\pi_n^\rho, y_n)\right\}, \quad M_n^\rho(F) = E_\rho \left\{F(\tilde{\pi}_n^\rho, y_n)\right\}.$$

The following result summarizes Lemmas 9, 10 and Proposition 3 of [5]

THEOREM 3.6. *If ρ is an invariant measure for the operator T then*

$$m_n^\rho \prec m_{n+1}^\rho \prec M_{n+1}^\rho \prec M_n^\rho.$$

Furthermore, there exist measures m and M such that $m_n^\rho \Rightarrow m$ and $M_n^\rho \to M$, as $n \to \infty$, where $\Rightarrow$ denotes weak convergence of probability measures. The measures m and M are invariant for the operator Π and for any invariant measure Φ of the operator Π with barycenter ρ we have $m \prec \Phi \prec M$.

Consequently the measures m and M are minimal and maximal invariant measures of the operator Π with respect to the ordering $\prec$.

Consider now again the filtering processes $\tilde{\pi}_n^\rho$ and π_n^ρ. Both processes have the same initial law of the pair $\begin{pmatrix}x_n \\ y_n\end{pmatrix}$. The difference is that by the formula (3.1) we observe random x_0 in the first process case. In fact the law of x_0 is $\rho(\cdot \times E)$, while the law of y_0 given x_0 is given by the conditional law $q_\rho(x_0, \cdot)$, where using conditional law decomposition (Theorem I.3.1 of [8]) $\rho(dx, dy) = q_\rho(x, dy)\rho(dx \times E)$. We say that filtering processes $\tilde{\pi}_n^\rho$ and π_n^ρ are *asymptotically stable in probability* whenever for any $f \in C(E_0)$ we have

$$\tilde{\pi}_n^\rho(f) - \pi_n^\rho(f) \to 0$$

in P_ρ probability as $n \to \infty$. We have (see Theorem 4 of [5])

THEOREM 3.7. *Assume that ρ in an invariant measure for the operator T and the operator Π is Feller. Then there is a unique invariant measure for the operator Π with barycenter ρ if and only if the filtering processes $(\tilde{\pi}_n)$ and (π_n) are asymptotically stable in P_ρ probability.*

In the Theorem 3.7 above we have equivalence of asymptotical stability of the processes $(\tilde{\pi}_n)$ and (π_n) and existence of unique invariant measure for the operator Π. It appears that we have a similar situation in the case of asymptotical stability of approximate filtering processes. Namely, as is shown in Theorem 5 of [5], following [17] the existence of a unique invariant measure for the operator Π under nonrestrictive assumptions implies asymptotical stability of approximate filtering processes. Consequently taking into account Theorem 3.1 and Corollary 3.5 the asymptotical stability of approximate filtering processes is almost equivalent to the existence of a unique invariant measure for the operator Π.

4. Ergodicity by vanishing discount approach

In this section we present an alternative method to show uniqueness of the invariant measure for the operator Π based on an ideas from [19] adapted to the problem in [6]. Given a nonnegative bounded Borel measurable function $F\colon \mathcal{P}(E_0) \times E \mapsto R$ define

$$g_F := \inf_{\mu \in \mathcal{P}(E_0), y \in E} \limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \Pi^i F(\mu, y)$$

and

$$w_F^\beta(\mu, y) := \sum_{i=0}^{\infty} \beta^i \Pi^i F(\mu, y)$$

with $0 < \beta < 1$. Let

$$m_F^\beta := \inf_{\mu \in \mathcal{P}(E_0), y \in E} w_F^\beta(\mu, y),$$

$$\bar{g}_F := \limsup_{\beta\to 1} (1-\beta) m_F^\beta,$$

and

$$\underline{g}_F := \liminf_{\beta\to 1} (1-\beta) m_F^\beta.$$

By the Tauberian theorem (see Lemma 1.2 of [19]) we have

$$0 \le \underline{g}_F \le \bar{g}_F \le g_F. \tag{4.1}$$

Our approach is based on the following Lemma (see Lemma 1 of [6])

Lemma 4.1. *If there is a nonnegative Borel measurable function w_F such that for $\mu \in \mathcal{P}(E_0)$ and $y \in E$*

$$w_F(\mu, y) + \underline{g}_F \geq F(\mu, y) + \Pi w_F(\mu, y) \tag{4.2}$$

then for each $\mu \in \mathcal{P}(E_0)$ and $y \in E$

$$\underline{g}_F = \limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \Pi^i F(\mu, y). \tag{4.3}$$

Proof. Iterating (4.2) we obtain

$$w_F(\mu, y) + n\underline{g}_F \geq \sum_{i=0}^{n-1} \Pi^i F(\mu, y) + \Pi^n w_F(\mu, y) \geq \sum_{i=0}^{n-1} \Pi^i F(\mu, y)$$

and therefore

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \Pi^i F(\mu, y) \leq \underline{g}_F. \tag{4.4}$$

Since by (4.1) we have $\underline{g}_F \leq \limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \Pi^i F(\mu, y)$, taking into account (4.4) we obtain (4.3). □

Let

$$h_F^\beta(\mu, y) := w_F^\beta(\mu, y) - m_F^\beta.$$

Clearly $h_F^\beta(\mu, y) \geq 0$. Our main assumption is

(A_F) $\sup_{\beta<1} h_F^\beta(\mu, y) < \infty$ for all $(\mu, y) \in \mathcal{P}(E_0) \times E$.

One can easily show that w_F^β is a solution to the following equation

$$w_F^\beta(\mu, y) = F(\mu, y) + \beta\Pi w_F^\beta(\mu, y). \tag{4.5}$$

Consequently we have the following equation for h_F^β

$$h_F^\beta(\mu, y) = F(\mu, y) - (1 - \beta)m_F^\beta + \beta\Pi h_F^\beta(\mu, y). \tag{4.6}$$

We would like to let $\beta \to 1$ in (4.6). We have

LEMMA 4.2. *Under* (A_F) *there is a nonnegative Borel measurable function* w_F *such that for* $\mu \in \mathcal{P}(E_0)$ *and* $y \in E$ *inequality* (4.2) *is satisfied.*

PROOF. We first choose a subsequence $\beta_m \to 1$ such that $(1-\beta_m)m_F^{\beta_m} \to \underline{g}_F$ as $m \to \infty$ and define $w_F(\mu, y) := \liminf_{m\to\infty} h_F^{\beta_m}(\mu, y)$. Then using the Fatou Lemma in (4.6) we obtain (4.2). □

Consequently under (A_F) we have (4.3) and by (4.1) $\underline{g}_F$ coincides with g_F. We have the following main result (see Proposition 1 of [6])

THEOREM 4.3. *Assume that the assumption* (A_F) *is satisfied for the bounded functions* F *from the class which determines measures on* $\mathcal{P}(E_0)\times E$*. Then there is at most one invariant measure* Φ *for the pair* $\begin{pmatrix}\pi_n^\rho \\ y_n\end{pmatrix}$.

PROOF. By Kakutani theorem (see [10] and [25] Section XIII.2) if Φ is an invariant measure for the pair $\begin{pmatrix}\pi_n^\rho \\ y_n\end{pmatrix}$ then $\frac{1}{n}\sum_{i=0}^{n-1}\Pi^i F(\mu, y)$ converges for Φ almost all $(\mu, y) \in \mathcal{P}(E_0) \times E$. By Lemma 4.1 and Lemma 4.2 this limit is defined in a unique way as $\underline{g}_F$. The limit is therefore the same for each $(\mu, y) \in \mathcal{P}(E_0) \times E$. By the individual ergodic theorem (see Theorem XIII.2.6 of [25]) the integral of the limit of $\frac{1}{n}\sum_{i=0}^{n-1}\Pi^i F(\mu, y)$ as $n \to \infty$ with respect to Φ (equal to $\underline{g}_F$) coincides with the value of $\int_{\mathcal{P}(E_0)\times E} F(\mu, y)\Phi(d\mu, dy)$. Therefore

$$\underline{g}_F = \int_{\mathcal{P}(E_0)\times E} F(\mu, y)\Phi(d\mu, dy).$$

Since this is true for functions F which determine measures we have at most one invariant measure for the pair $\begin{pmatrix}\pi_n^\rho \\ y_n\end{pmatrix}$. □

To use Theorem 4.3 need the assumption (A_F) to be satisfied for the bounded functions F from the class which determines measures on $\mathcal{P}(E_0)\times E$. The family $C_c(\mathcal{P}(E_0) \times E)$ of functions $\mathcal{P}(E_0) \times E \ni (\nu, y) \mapsto F(\nu, y)$, which are continuous and bounded and convex with respect to ν for fixed $y \in E$ is in particular such class. Let the operator Z that transforms the functions on $\mathcal{P}(E_0) \times E$ into $\mathcal{M}(E_0) \times E$ be given by the formula

$$ZF(\zeta, y) = \zeta(E_0)F\Big(\frac{\zeta}{\zeta(E_0)}, y\Big).$$

Then one can notice that the operator Π defined in (1.7) is also of the form

$$\Pi F(\mu, y) = \int_E ZF(N(y, y', \mu), y')\eta(dy')$$

where for $A \in \mathcal{B}(E_0)$ we define $N(y, y', \mu)(A) := \int_A r\,(x', y, y')\,P\,(\nu, dx')$. We also have (see Lemma 2 in [7])

LEMMA 4.4. *If $F\colon \mathcal{P}(E_0) \times E \mapsto R$ is concave with respect to the first coordinate then $ZF\colon M^+(E_0) \times E \mapsto R$ is also concave with respect to the first coordinate.*

Let $A \in \mathcal{B}(E_0)$

$$N_0(y, \mu)(A) := \mu(A),$$
$$N_1(y, y', \mu)(A) := N(y, y', \mu)(A),$$

and by induction

$$N_n(y_0, y_1, \ldots, y_n, \mu)(A) := N(y_{n-1}, y_n, N_{n-1}(y_0, y_1, \ldots, y_{n-1}, \mu))(A),$$
$$M_n(y_0, y_1, \ldots, y_n, \mu)(A) := \frac{N_n(y_0, y_1, \ldots, y_n, \mu)(A)}{N_n(y_0, y_1, \ldots, y_n, \mu)(E_0)}.$$

If the initial law of (x_n) is μ and $y_0 = y$ then one can show that

$$\pi_n^{\mu y}(A) = M_n(y, y_1, \ldots, y_n, \mu)(A)$$

P a.e.. We have the following important interpretation for iterations of the operator Π (for details see Lemma 4 of [6])

LEMMA 4.5.

$$\begin{aligned} \Pi^n F(\mu, y) &= E_{\mu y}\left\{F(\pi_n, y_n)\right\} \\ &= E^0\left\{ZF(N_n(y, \tilde{y}_1, \ldots, \tilde{y}_n, \mu), \tilde{y}_n)\right\}, \end{aligned} \tag{4.7}$$

where E^0 is with respect to P^0 under which random variables $\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_n$ are i.i.d. with common law η.

Also

LEMMA 4.6. *For bounded Borel measurable function* $F\colon \mathcal{P}(E_0)\times E \mapsto R$ *and* $\beta\in(0,1)$ *there is a unique solution* w_F^β *of* (4.5). *If* F *is continuous and* Π *is Feller then* w_F^β *is also continuous. If* F *concave with respect to the first coordinate then* w_F^β *is also concave with respect to the first coordinate.*

For given $\mu,\nu\in\mathcal{P}(E_0)$, $\epsilon\in(0,1)$, and a positive integer m define

$$D_{\epsilon,m}^{\nu,\mu}(y,y') := \Big\{\omega : N_m(y,\tilde{y}_1,\ldots,\tilde{y}_m,\nu) \geq \epsilon N_m(y',\tilde{y}_1,\ldots,\tilde{y}_m,\mu)\Big\}.$$

We introduce the following assumption

(C1) $\exists_m\exists_{\epsilon>0,\delta>0}$ such that $\forall_{y,y'\in E}$

$$E^0\left\{1_{D_{\epsilon,m}^{\nu,\mu}(y,y')}(\omega)N_m(y',\tilde{y}_1,\ldots,\tilde{y}_m,\mu)(E)\right\}\geq\delta.$$

Iterating (4.5) and taking into account (4.7) we have

$$\begin{aligned} h_F^\beta(\mu,y) = \sum_{i=0}^{m-1}\beta^i E^0\left\{ZF(N_i(y,\tilde{y}_1,\ldots,\tilde{y}_i,\mu),\tilde{y}_i)\right\} \\ -(1-\beta)m_F^\beta\sum_{i=0}^{m-1}\beta^i+\beta^m E^0\left\{Zh_F^\beta(N_m(y,\tilde{y}_1,\ldots,\tilde{y}_m,\mu),\tilde{y}_m\right\} \end{aligned}$$

from which using (C1) and concavity of Zh_F^β we obtain (see Proposition 1 of [6])

PROPOSITION 4.7. *Under* (C1) *for continuous bounded concave with respect to the first argument function* F

$$\|h_F^\beta\|\leq\frac{m\|F\|}{\epsilon\delta},$$

where $\|\cdot\|$ *stands for the supremum norm.*

Notice that we have much more than is required in (A_F), on the other hand we need in (C1) to have δ uniform for all $y,y'\in E$, which is quite restrictive in the case when E is a not compact.

5. Recent results and open problems

The main result concerning ergodicity of filtering processes was formulated in the paper [23], where the case of observation independent of previous observation i.e. the case when r in (1.1) was of the form $r(x', y')$ was studied. We shall present below a further result from [22] in which the observation structure (1.3) was considered. We introduce the following assumption

(H) there is a probability measure $\phi \in \mathcal{P}(E_0 \times E)$ such that

$$P\left\{(x_n, y_n) \in \cdot\right\} \to \phi(\cdot),$$

in variation norm as $n \to \infty$.

This assumption is clearly satisfied for a wide family of ergodic processes called positive aperiodic Harris processes (see Theorem 13.3.1 of [16]). We have from [22]

THEOREM 5.1. *Under* (H), *whenever the kernels* $P_1^{x'}(y, \cdot)$ *are equivalent for* $x' \in E_0$ *and* $y \in E$, *there is a unique invariant measure* Φ *for the pair* (π_n, y_n) *and* Π^n *converges in variation norm to* Φ, *as* $n \to \infty$.

If the convergence in (H) is replaced by convergence in weak topology and the kernels $P_1^{x'}(y, \cdot)$ are equivalent, then we may have more invariant measures, as is shown in [24], for the pair $\begin{pmatrix} \pi_n^\rho \\ y_n \end{pmatrix}$. Therefore the conjecture formulated in the introduction is false. On the other hand a long standing problem consisting in filling out the gap in the famous paper [9] has been then partially solved by [23] and [22]. It is still an open problem to clarify what we should add to the weak convergence of $P\left\{(x_n, y_n) \in \cdot\right\} \to \phi(\cdot)$, as $n \to \infty$ to get a unique invariant measure Φ for the pair $\begin{pmatrix} \pi_n^\rho \\ y_n \end{pmatrix}$.

From the application to system theory point of view ergodicity of filtering processes is only the first step required to study partially observed control problems with average cost per unit time functionals. We would like to study the case when the state process is controlled using at time n control v_n values in a compact set U, which is adapted to Y^n and the processes (x_n) and (y_n) have controlled transition kernels $P^{v_n}(x_n, dx)$ and $P_1^{x_{n+1}, v_n}(y_n, dy)$ respectively. Denote by V the class of such admissible controls (v_n). We would like to minimize the cost functional

$$J(V) = \limsup_{n\to\infty} \frac{1}{n} E^V \left\{ \sum_{i=0}^{n-1} c(x_i, y_i) \right\}$$

and study the ergodicity of the pair $\begin{pmatrix} \pi_n^\rho \\ y_n \end{pmatrix}$, where now π_n^ρ is a controlled filtering process. A good candidate for the value function of the cost functional $J(V)$ is $\underline{g}$, which comes from so called Bellman equation

$$w(\mu, y) + \underline{g} = \inf_{a \in U} [c(\mu, y) + \Pi^a w(\mu, y)]. \tag{5.1}$$

Although under certain assumptions we are able to show the existence of solutions to (5.1) (see [21] or [4]), the question of ergodicity of the pair remains still open.

References

[1] Atar R., Zeitouni O., *Lyapunov exponents for finite-state nonlinear filtering*, SIAM J. Control Optim. **35** (1997), 36–55.

[2] Baxendale P., Chigansky P., Liptser R., *Asymptotic stability of the Wonham filter: ergodic and nonergodic signals*, Preprint 2002.

[3] Borkar V.S., *Ergodic control of partially observed Markov chains*, Systems Control Lett. **34** (1998), 185–189.

[4] Borkar V.S., Budhiraja A., *A further remark on dynamic programming for partially observed Markov processes*, Stochastic Process. Appl. **112** (2004), 79–93.

[5] Di Masi G., Stettner L., *Ergodicity of Hidden Markov Models*, Math. Control Signals Systems **17** (2005), 269–296.

[6] Di Masi G., Stettner L., *Ergodicity of filtering process by vanishing discount approach*, Systems Control Lett. **57** (2008), 150–157.

[7] Di Masi G., Stettner L., *Risk sensitive control of discrete time partially observed Markov processes with infinite horizon*, Stochastics and Stochastics Rep. **67** (1999), 309–322.

[8] Ikeda N., Watanabe Sh., *Stochastic Differential Equations and Diffusion Processes*, North-Holland, Amsterdam, 1981.

[9] Kaijser T., *A limit theorem for partially observed Markov chains*, Ann. Probab. **3** (1975), 677–696.

[10] Kakutani S., *Ergodic theorems and the Markoff processes with a stable distribution*, Proc. Imp. Acad. Tokyo **16** (1940), 49–54.

[11] Kallenberg O., *Foundations of Modern Probability*, Springer-Verlag, New York–Berlin, 1997.

[12] Kunita H., *Asymptotic behaviour of the nonlinear filtering errors of Markov process*, J. Multivariate Anal. **1** (1971), 365–393.

[13] Kunita H., *Ergodic properties of nonlinear filtering processes*, in: *Spatial Stochastic Processes*, ed. by Aleksander K.C., Watkins J.C., Progr. Probab. **19**, Birkhauser, Boston, 1991, pp. 233–256.

[14] Liptser R., Shiryaev A.N., *Statistics of Random Processes II. Applications*, Second Edition, Springer-Verlag, New York–Berlin, 2001.

[15] Liverani C., *Decay of correlations*, Ann. of Math. **142** (1995), 239–301.

[16] Meyn S.P., Tweedie R.L., *Markov Chains and Stochastics Stability*, Springer-Verlag, New York–Berlin, 1993.

[17] Ocone D., Pardoux E., *Asymptotic stability of the optimal filter with respect to its initial conditions*, SIAM J. Control Optim. **34** (1996), 226–243.
[18] Runggaldier W., Stettner L., *Approximations of Discrete Time Partially Observed Control Problems*, Applied Mathematics Monographs CNR, Giardini Editori, Pisa, 1994.
[19] Schäl M., *Average Optimality in Dymanic Programming with General State Space*, Math. Oper. Res. **18** (1993), 163–172.
[20] Stettner L., *On Invariant Measures of Filtering Processes*, in: *Proc. 4th Bad Honnef Conf. on Stochastic Differential Systems*, ed. by Christopeit N., Helmes K., Kohlmann M., Lect. Notes in Control Inf. Sci. 126, Springer-Verlag, New York–Berlin, 1989, pp. 279–292.
[21] Stettner L., *Ergodic control of partially observed Markov processes with equivalent transition probabilities*, Appl. Math. (Warsaw) **22** (1993), 25–38.
[22] Tong X.T., van Handel R., *Ergodicity and stability of the conditional distributions of nondegenerate Markov chains*, Ann. Appl. Probab. **22** (2012), 1495–1540.
[23] van Handel R., *The stability of conditional Markov processes and Markov chains in random environments*, Ann. Probab. **37** (2009), 1876–1925.
[24] van Handel R., *A nasty filtering problem*, Preprint 2010, arXiv:1009.0507.
[25] Yosida K., *Functional Analysis*, Springer-Verlag, New York–Berlin, 1978.

Institute of Mathematics
Polish Academy of Sciences
Sniadeckich 8
00-956 Warsaw
also Vistula University
Poland
e-mail: stettner@impan.pl