Approximation of nonlinear filters for Markov systems with delayed observations

Antonella Calzolari, Patrick Florchinger and Giovanna Nappo

***Abstract*— We obtain some approximation results for a class of nonlinear filtering problems with delay in the observation, i.e. systems (X, Y), which can be represented by means of a Markov system $(X, \hat{Y})$, in the sense that $Y_t = \hat{Y}_{a(t)}$. To this aim we give some general upper bounds which are computed explicitly in the particular case of Markov jump processes with counting observations.**

Key Words: Nonlinear Filtering, Jump Processes, Markov Processes, Approximation of Stochastic Processes

AMS Classification: 93E11, 60G35, 60J75, 65C50, 62M20

I. Introduction

Consider a partially observed stochastic system $(\boldsymbol{X}, \boldsymbol{Y}) = (X_t, Y_t)_{t\geq 0}$, that is, a *state process* $\boldsymbol{X} = (X_t)_{t\geq 0}$, which cannot be directly observed, and a completely observable process $\boldsymbol{Y} = (Y_t)_{t\geq 0}$, which is referred to as the *observation process*. The aim of stochastic nonlinear filtering is to compute the conditional law π_t of the state process at time t, given the observation process up to time t, i.e., the computation of the so-called filter

$$\pi_t(\varphi) = E\big[\varphi(X_t)/\mathcal{F}_t^Y\big], \tag{1}$$

for all functions φ belonging to a determining class, i.e., the best estimate of $\varphi(X_t)$ given the σ-algebra of the observations up to time t, $\mathcal{F}_t^Y = \sigma\{Y_s,\, s \leq t\}$.

A classical model of partially observed systems arises when both the state and the observation are diffusion processes. In this case it has been shown that the filter solves a stochastic partial differential equation known as the Kushner–Stratonovich equation (cf., e.g., Pardoux [17] and the references therein).

Filtering problems involving jump-diffusion processes have been studied by many authors. In particular, the jump system $(x_t, y_t)_{t\geq 0}$ defined below enters in the more general framework studied by Kliemann, Koch, and Marchetti [12]:

$$x_t = x_0 + \int_0^t \int_{D_0(x_{s^-}, y_{s^-})} K_0(x_{s^-}, y_{s^-}; \zeta) \mathcal{N}(ds, d\zeta)$$
$$+ \int_0^t \int_{D_1(x_{s^-}, y_{s^-})} K_1(x_{s^-}, y_{s^-}; \zeta) \mathcal{N}(ds, d\zeta), \tag{2}$$
$$y_t = \int_0^t \int_{D_1(x_{s^-}, y_{s^-})} \mathcal{N}(ds, d\zeta), \tag{3}$$

where $\mathcal{N}(ds, d\zeta)$ is a Poisson measure on $\mathbb{R} \times \Sigma$ with mean measure $ds \otimes \nu(d\zeta)$; the random variable x_0 has values in $\mathbb{R}^k$ and probability distribution μ_0^x; the random variable x_0 and the Poisson random measure $\mathcal{N}(ds, d\zeta)$ are independent; the sets $D_0(x, y)$ and $D_1(x, y)$ are disjoint. Under suitable hypotheses (cf. [12]), the above system has a unique solution $(x_t, y_t)_{t\geq 0}$, which is a Markov process with formal generator $\boldsymbol{L}$ given by

$$\boldsymbol{L} f(x, y) = \int_{D_0(x,y)} [f(x + K_0(x, y; \zeta), y) - f(x, y)]\, \nu(d\zeta)$$
$$+ \int_{D_1(x,y)} [f(x + K_1(x, y; \zeta), y + 1) - f(x, y)]\, \nu(d\zeta).$$

Then the filter $\pi_t^x(\varphi) = E[\varphi(x_t)/\mathcal{F}_t^y]$ can be obtained via the following normalization procedure: Let $\boldsymbol{s} = \{s_i\}_{i\geq 0}$ be an increasing sequence of times such that $s_0 = 0$; define

$$B^y \varphi(x) = \int_{D_0(x,y)} [\varphi(x + K_0(x, y; \zeta)) - \varphi(x)]\, \nu(d\zeta)$$

and

$$R^y \varphi(x) = \int_{D_1(x,y)} [\varphi(x + K_1(x, y; \zeta)) - \varphi(x)]\, \nu(d\zeta);$$

define $\hat{\rho}_t(dx|\boldsymbol{s})$ by the following self-contained procedure: for $s_i \leq t < s_{i+1}$

$$\hat{\rho}_t(\varphi|\boldsymbol{s}) = E\left[\varphi(X^i_{t-s_i}(\boldsymbol{s}))\, e^{-\int_0^{t-s_i} \lambda_1(X^i_u(\boldsymbol{s}), i) du}\right],$$

where $\lambda_1(x, i) = \nu(D_1(x, i))$, and $\{X^i_t(\boldsymbol{s}); t \geq 0\}$ is a Markov process with generator B^i and initial distribution $\hat{\rho}_{s_i}(dx|\boldsymbol{s})$ defined inductively by

$$\hat{\rho}_{s_0}(\varphi|\boldsymbol{s}) = \mu_0^x(\varphi), \qquad \hat{\rho}_{s_{i+1}}(\varphi|\boldsymbol{s}) = \frac{\hat{\rho}_{s^-_{i+1}}(Q^i\varphi|\boldsymbol{s})}{\hat{\rho}_{s^-_{i+1}}(\lambda_1(\cdot, i)|\boldsymbol{s})}$$

with $Q^i\varphi(x) = R^i\varphi(x) + \lambda_1(x, i)\varphi(x)$, or equivalently $Q^i\varphi(x) = \int_{D_1(x,i)} \varphi(x + K_1(x, i; \zeta))\, \nu(d\zeta)$.

Setting

$$\hat{\Pi}_t(\varphi|\boldsymbol{s}) := \frac{\hat{\rho}_t(\varphi|\boldsymbol{s})}{\hat{\rho}_t(\boldsymbol{1}|\boldsymbol{s})}, \tag{4}$$

This work was partially supported by MURST project ”*Metodi stocastici in finanza matematica” Anno 2004 - prot. 2004014572_007*

Antonella Calzolari is with Dipartimento di Matematica - Università di Roma ”Tor Vergata”, via della Ricerca Scientifica 1, I 00133 Roma, Italy calzolar@axp.mat.uniroma2.it

Patrick Florchinger is with Département de Mathématiques, Université de Metz, 23 Allée des Oeillets F 57160 Moulins les Metz, France florchin@math.univ-metz.fr

Giovanna Nappo is with Dipartimento di Matematica - Università di Roma ”La Sapienza”, piazzale A. Moro 2, I 00185 Roma, Italy nappo@mat.uniroma1.it

we have

$$\pi_t^x(\varphi) = E[\varphi(x_t)/\mathcal{F}_t^y] = \hat{\Pi}_t(\varphi|\boldsymbol{s})\Big|_{\boldsymbol{s}=\{T_i\}_{i\geq 0}}, \qquad (5)$$

where T_i are the observed jump times.

Different approximation procedures and numerical schemes have been studied by many authors in the past years (see, for example, Kushner [14], Di Masi, Pratelli, and Runggaldier [9], Le Gland [15], Elliott and Glowinski [10], Lototsky, Mikulevicius, and Rozovskii [16], Del Moral [8], Calzolari and Nappo [6], or Ceci, Gerardi, and Tardelli [7] and the references therein).

In this paper we are interested in the approximation of the filter when dealing with the following nonlinear filtering problem with delay in the observation: we consider the system $(\boldsymbol{X}, \boldsymbol{Y})$ such that $Y_t = \hat{Y}_{a(t)}$, where $(\boldsymbol{X}, \hat{\boldsymbol{Y}})$ is a Markov process with generator $\boldsymbol{L}$, and $a(\cdot)$ is a nondecreasing continuous function, with $a(0) = 0$, $a(t) \leq t$ for all $t \geq 0$. As an example we can take $(\boldsymbol{X}, \hat{\boldsymbol{Y}}) = (x_t, y_t)_{t\geq 0}$, with x_t and y_t defined as in (2) and (3), respectively.

Theorem I.1 (see [4]). *The filter π_t associated with the system $(\boldsymbol{X}, \boldsymbol{Y})$ described above can be represented as*

$$\pi_t(\varphi) = E\big[\big(e^{\boldsymbol{L}(t-a(t))}\phi\big)(X_{a(t)}, \hat{Y}_{a(t)})/\mathcal{F}_{a(t)}^{\hat{Y}}\big],$$

where $\phi(x, y) = \varphi(x)$. Moreover, if the conditional law of X_s given $\mathcal{F}_s^{\hat{Y}}$ is known and denoted by $\hat{\pi}_s$, then

$$\pi_t(\varphi) = \hat{\pi}_{a(t)}\big(\big(e^{\boldsymbol{L}(t-a(t))}\phi\big)(\cdot, \hat{Y}_{a(t)})\big).$$

The approximation problem we consider is interesting due to the fact that partially observed systems with delay in the observations appear in stochastic finance. For instance, in [18] Schweizer has given an example of information with delay for a financial model, using the the risk minimization criterion. This criterion corresponds to a quadratic loss function; the use of different loss functions leads to consider the hedging problem as an optimal control problem. Then, from a Bayesian point of view, filtering may appear in the case of partial observations, and in this direction diffusion-type models are mainly studied in the literature. More recently, in [11] Kirch and Runggaldier have studied a control problem with partial observation when the stock price evolves as

$$S_t = s_0\, e^{aN_t^+ - bN_t^-},$$

where a, $b > 0$ are constants, and N^+ and N^- are counting processes with random unobservable intensities Λ^+ and Λ^-.

The observation of the stock price is equivalent to the observation of the couple of counting processes N^+ and N^-.

In [11], N^+ and N^- are assumed to be conditionally independent Poisson processes given their random time constant intensities Λ^+ and Λ^-, which are independent and with prior distribution gamma. In this case the filter of (Λ^+, Λ^-) given N^+ and N^- up to time t can be reduced to a couple of filtering problems, each with one counting observation process, and can be computed explicitly. If the time constant assumption on the intensities were dropped, then the explicit computation of the filter would no longer be feasible, and a filtering approximation problem would naturally arise. Considering also a delay in the information would then lead to an example fitting the framework of section IV.

Some approximation results can be obtained for different systems and by making use of different techniques. As an example we can take the simple delayed diffusion model considered in [4], the filter of which can be approximated by the convergence result in [5], as explained in Remark 1.3 of [3], where we deal with the approximation of nonlinear filtering for systems solving stochastic delay differential equations (see [3] for some references on the last topic).

II. Approximation for general filters

A. *Different kinds of approximation*

Suppose that a sequence of stochastic systems $(\boldsymbol{X^n}, \boldsymbol{Y^n})$ with values in $\mathbb{R}^k \times \mathbb{R}^d$ converges to a system $(\boldsymbol{X}, \boldsymbol{Y})$. Then a natural question is whether the corresponding sequence of filters π^n converges to the filter of the limit system π.

Different kinds of convergence can be considered for both the systems and the filters. Convergence of their distributions is the first one, and, moreover, the only one that can be considered when the systems $(\boldsymbol{X^n}, \boldsymbol{Y^n})$ are defined on different probability spaces. The most frequently used is weak convergence of their distributions. Furthermore, in this case one can distinguish between convergence of π_t^n to π_t as random probability measures on $\mathbb{R}^k$, for each t, and convergence of the processes $\boldsymbol{\pi^n} = (\pi_t^n; t \geq 0)$ to $\boldsymbol{\pi} = (\pi_t; t \geq 0)$, as *càdlàg* measure-valued processes. Finally, one can consider different metrics $dist(\nu_1, \nu_2)$ on the space $\mathcal{P} = \mathcal{P}(\mathbb{R}^k)$, which is the space of probability measures on $\mathbb{R}^k$, such as the total variation $\|\nu_1 - \nu_2\|_{TV} = \sup\big\{|\nu_1(\varphi) - \nu_2(\varphi)|/\|\varphi\|; \varphi \text{ bounded}\big\}$, the Kantorovitch metric $\boldsymbol{\kappa}(\nu_1, \nu_2) = \sup\big\{|\nu_1(\varphi) - \nu_2(\varphi)|/L_\varphi; \varphi \text{ Lipschitz}\big\}$, or the bounded-Lipschitz metric $d_{BL}(\nu_1, \nu_2) = \sup\big\{|\nu_1(\varphi) - \nu_2(\varphi)|/(\|\varphi\| \vee L_\varphi); \varphi \text{ bounded and Lipschitz}\big\}$, where $\|\varphi\|$ denotes the sup-norm, and L_φ is the Lipschitz constant of φ. In the following we will consider $\mathcal{P}$ endowed with one of the above metrics.

When instead all the processes are defined on the same probability space $(\Omega, \mathcal{F}, P)$, then one can also consider other kinds of convergence such as convergence in probability, convergence in $L^p([0, T] \times \Omega,\ dt \otimes dP)$, and so on. This situation arises typically when one is interested in the limit system, and the sequence $(\boldsymbol{X^n}, \boldsymbol{Y^n})$ is constructed pathwise starting from the path $(\boldsymbol{X}, \boldsymbol{Y})$ such as, for instance, in the Euler approximations of diffusive systems.

When the filters $\boldsymbol{\pi^n}$ and $\boldsymbol{\pi}$ have a robust version, that is, when for suitable deterministic measure-valued functionals U^n and U, with paths in the Skorohod space $D_{\mathcal{P}}([0, T])$,

$$U^n,\, U :\ [0, T] \times D_{\mathbb{R}^d}([0, T]) \mapsto \mathcal{P}$$

such that $U^n(t, \boldsymbol{y}) = U^n(t, y(\cdot\wedge t))$, $U(t, \boldsymbol{y}) = U(t, y(\cdot\wedge t))$, $\pi_t^n(dx) = U^n(t, \boldsymbol{Y^n}; dx)$, and $\pi_t(dx) = U(t, \boldsymbol{Y}; dx)$; then, in order to consider convergence in probability, one has to assume that the processes $\boldsymbol{Y^n}$ and $\boldsymbol{Y}$ are defined on the same probability space. Such deterministic functionals U^n

and U satisfying the above properties always exist under very general conditions (see Kurtz and Ocone [13]). Note that the functionals U^n and U depend on the joint distribution of $(\boldsymbol{X^n}, \boldsymbol{Y^n})$ and $(\boldsymbol{X}, \boldsymbol{Y})$, respectively, and are defined in $D_{\mathbb{R}^d}([0,T])$ almost surely with respect to $P_{\boldsymbol{Y^n}}$, the law of $\boldsymbol{Y^n}$, and with respect to $P_{\boldsymbol{Y}}$, the law of $\boldsymbol{Y}$, respectively.

As an example one can consider approximating and limit models such as the jump model with counting observations, as given by (2) and (3), for which the functionals U and U^n can be computed as in (4), or such as the classical diffusive model, for which the functionals U and U^n are also computable as shown in [13], for example.

A different approach to the problem of the approximation of the filter takes into account that a realistic approximation depends on the trajectory actually observed and uses the robust representation described above. Two different types of situations can arise as follows:

1 The *true* model is $(\boldsymbol{X}, \boldsymbol{Y})$, and therefore we observe $\boldsymbol{Y}$, while the models $(\boldsymbol{X^n}, \boldsymbol{Y^n})$ are *more manageable* approximations.

2 The *true* model is $(\boldsymbol{X^n}, \boldsymbol{Y^n})$, depending on a large parameter n, and therefore we observe $\boldsymbol{Y^n}$, while $(\boldsymbol{X}, \boldsymbol{Y})$ is a *more manageable* limit model.

In situation 1 the *true* filter is π_t, and it is natural to consider

$$\tilde{\pi}_t^n = U^n(t, \boldsymbol{Y}) \tag{6}$$

as an approximation of π_t, depending on the trajectory actually observed. The functional U^n is defined $P_{\boldsymbol{Y^n}}$-almost surely; therefore, in order to define $\boldsymbol{\tilde{\pi}^n} = \{\tilde{\pi}_t^n; t \geq 0\}$ almost surely it is natural to assume that $P_{\boldsymbol{Y}}$ is absolutely continuous with respect to $P_{\boldsymbol{Y^n}}$. One is interested in evaluating

$$dist(\pi_t, \tilde{\pi}_t^n) = dist(U(t, \boldsymbol{Y}), U^n(t, \boldsymbol{Y})), \tag{7}$$

where $dist$ can be one of the integral metric considered above. In situation 2 the *true* filter is π_t^n, and it is natural to consider $\overline{\pi}_t^n = U(t, \boldsymbol{Y^n})$ as an approximation of π_t^n, depending on the trajectory actually observed. In this case one is interested in $dist(\pi_t^n, \overline{\pi}_t^n)$.

In this presentation we consider only situation 1. However the results here obtained can be easily rephrased in situation 2 (see [3]).

Note that, $\tilde{\pi}_t^n$ is not a conditional law. Moreover, with this kind of approximation it is not even necessary that the sequence of processes $\{\boldsymbol{Y^n}\}$ and $\boldsymbol{Y}$ be defined on the same probability space, and both the almost sure convergence and the convergence to zero in probability of (7) can be considered. For instance, the convergence in probability is implied by the convergence to zero of

$$E[dist(\pi_t, \tilde{\pi}_t^n)] = E[dist(U(t, \boldsymbol{Y}), U^n(t, \boldsymbol{Y}))].$$

Finally, note that when the distance between π_t and $\tilde{\pi}_t^n$ is given in terms of the total variation or the bounded-Lipschitz metrics, then

$$E[dist(\pi_t, \tilde{\pi}_t^n)] = E\left[\sup_{\varphi \in \mathcal{K}} |\pi_t(\varphi) - \tilde{\pi}_t^n(\varphi)|\right] \tag{8}$$

where $\mathcal{K}$ is a suitable class of functions: for the total variation, the class $\mathcal{K}_{TV} = \mathcal{K}_{TV}(1)$, where $\mathcal{K}_{TV}(\alpha)$ is the class of measurable functions φ with $\|\varphi\| \leq \alpha$; for the bounded-Lipschitz metric, the class $\mathcal{K}_{BL} = \mathcal{K}_{BL}(\alpha, \Lambda)$, where $\mathcal{K}_{BL}(\alpha, \Lambda) \subseteq \mathcal{K}_{TV}(\alpha)$ is the subclass of Lipschitz functions, with Lipschitz constant $L_\varphi \leq \Lambda$.

B. General upper bounds

In the following we assume that it is possible to construct copies $(\boldsymbol{\tilde{X}^n}, \boldsymbol{\tilde{Y}^n})$ and $(\boldsymbol{\tilde{X}}, \boldsymbol{\tilde{Y}})$ of the pairs $(\boldsymbol{X^n}, \boldsymbol{Y^n})$ and $(\boldsymbol{X}, \boldsymbol{Y})$ on the same measurable space $(\Omega, \mathcal{F})$, equipped with different probability measures $\mathbb{P}$ and $\mathbb{P}^n$ and with the property that $\boldsymbol{\tilde{Y}^n} = \boldsymbol{\tilde{Y}}$, i.e.,

(a) on $(\Omega, \mathcal{F}, \mathbb{P})$ the model $(\boldsymbol{\tilde{X}}, \boldsymbol{\tilde{Y}})$ has the same law as the model $(\boldsymbol{X}, \boldsymbol{Y})$,

$(\mathrm{a_n})$ on $(\Omega, \mathcal{F}, \mathbb{P}^n)$ the model $(\boldsymbol{\tilde{X}^n}, \boldsymbol{\tilde{Y}^n}) = (\boldsymbol{\tilde{X}^n}, \boldsymbol{\tilde{Y}})$ has the same law as the model $(\boldsymbol{X^n}, \boldsymbol{Y^n})$.

Remark II.1. When the systems $(\boldsymbol{X}, \boldsymbol{Y})$ and $(\boldsymbol{X^n}, \boldsymbol{Y^n})$ are Markovian, condition (a) means that under $\mathbb{P}$ the pair $(\boldsymbol{\tilde{X}}, \boldsymbol{\tilde{Y}})$ has the generator $\boldsymbol{L}$ of $(\boldsymbol{X}, \boldsymbol{Y})$, and the initial distribution $\mu = \mu_0^{X,Y}$, while condition $(\mathrm{a_n})$ means that under $\mathbb{P}^n$ the pair $(\boldsymbol{\tilde{X}^n}, \boldsymbol{\tilde{Y}})$ has the generator $\boldsymbol{L}^n$ of $(\boldsymbol{X^n}, \boldsymbol{Y^n})$, and the initial distribution $\mu^n = \mu_0^{X^n, Y^n}$.

As will become clear in the application to the jump models in section IV, it is natural to construct the probability spaces $(\Omega, \mathcal{F}, \mathbb{P})$ and $(\Omega, \mathcal{F}, \mathbb{P}^n)$ starting from a given probability space $(\Omega, \mathcal{F}, \mathbb{Q})$ in such a way that

(b1) $\mathbb{P}$ and $\mathbb{P}^n$ are absolutely continuous with respect to $\mathbb{Q}$ on $\tilde{\mathcal{F}}_t = \mathcal{F}_t^{\tilde{X}, \tilde{X}^n, \tilde{Y}}$ for all $t \geq 0$.

The above condition implies that $\mathbb{P}$ and $\mathbb{P}^n$ are absolutely continuous with respect to $\mathbb{Q}$ on $\mathcal{F}_t^{\tilde{Y}}$. For technical reasons, it is more convenient to assume the further condition that

(b2) the probability measures $\mathbb{Q}$, $\mathbb{P}$, and $\mathbb{P}^n$ are equivalent on $\mathcal{F}_t^{\tilde{Y}}$ for all $t \geq 0$.

Indeed, under conditions (a) and $(\mathrm{a_n})$, then $U(t, \boldsymbol{\tilde{Y}}; \varphi) = \mathbb{E}[\varphi(\tilde{X}_t)/\mathcal{F}_t^{\tilde{Y}}]$, $\mathbb{P}$-$a.s.$, and analogously $U^n(t, \boldsymbol{\tilde{Y}}; \varphi) = \mathbb{E}^n[\varphi(\tilde{X}_t^n)/\mathcal{F}_t^{\tilde{Y}}]$, $\mathbb{P}^n$-$a.s.$, and therefore, if condition (b2) holds, then the previous relations hold almost surely under $\mathbb{Q}$, $\mathbb{P}$, and $\mathbb{P}^n$. Therefore, by (6) and (8)

$$\begin{aligned}
E[dist(\pi_t, \tilde{\pi}_t^n)] &= E\left[\sup_{\varphi \in \mathcal{K}} |U(t, \boldsymbol{Y}; \varphi) - U^n(t, \boldsymbol{Y}; \varphi)|\right] \\
&= \mathbb{E}\left[\sup_{\varphi \in \mathcal{K}} |U(t, \boldsymbol{\tilde{Y}}; \varphi) - U^n(t, \boldsymbol{\tilde{Y}}; \varphi)|\right] \\
&= \mathbb{E}\left[\sup_{\varphi \in \mathcal{K}} |\mathbb{E}[\varphi(\tilde{X}_t)/\mathcal{F}_t^{\tilde{Y}}] - \mathbb{E}^n[\varphi(\tilde{X}_t^n)/\mathcal{F}_t^{\tilde{Y}}]|\right],
\end{aligned}$$

where we stress that, under $\mathbb{P}$, the law of $U^n(t, \boldsymbol{\tilde{Y}}; \varphi) = \mathbb{E}^n[\varphi(\tilde{X}_t^n)/\mathcal{F}_t^{\tilde{Y}}]$ is the same as the law of $\tilde{\pi}_t^n(\varphi)$, and therefore not the law of $\pi_t^n(\varphi)$, the filter of the approximating system, evaluated at φ. Then it is natural to start by looking for $\mathbb{Q}$-a.s. upper bounds of

$$\sup_{\varphi \in \mathcal{K}} \left|\mathbb{E}\left[\varphi(\tilde{X}_t)/\mathcal{F}_t^{\tilde{Y}}\right] - \mathbb{E}^n\left[\varphi(\tilde{X}_t^n)/\mathcal{F}_t^{\tilde{Y}}\right]\right|, \tag{9}$$

and then to take the expectation with respect to $\mathbb{E}$. Furthermore, if condition (b1) holds, it can be shown (e.g., by Lemma 4.1 of [6]) that, for any bounded function φ,

$$\big|\mathbb{E}\big[\varphi(\tilde{X}_t)/\mathcal{F}_t^{\tilde{Y}}\big] - \mathbb{E}^n\big[\varphi(\tilde{X}_t^n)/\mathcal{F}_t^{\tilde{Y}}\big]\big| \qquad (10)$$

is bounded above by

$$2\|\varphi\|\,\mathcal{Z}_t + \mathbb{E}\big[|\varphi(\tilde{X}_t) - \varphi(\tilde{X}_t^n)|/\mathcal{F}_t^{\tilde{Y}}\big],$$

$\mathbb{Q}$-a.s., and therefore also $\mathbb{P}$-a.s. and $\mathbb{P}^n$-a.s., where

$$\mathcal{Z}_t = \frac{\mathbb{E}_{\mathbb{Q}}\big[|(d\mathbb{P}^n/d\mathbb{Q})|_{\tilde{\mathcal{F}}_t} - (d\mathbb{P}/d\mathbb{Q})|_{\tilde{\mathcal{F}}_t}|/\mathcal{F}_t^{\tilde{Y}}\big]}{\mathbb{E}_{\mathbb{Q}}\big[(d\mathbb{P}/d\mathbb{Q})|_{\tilde{\mathcal{F}}_t}/\mathcal{F}_t^{\tilde{Y}}\big]}. \qquad (11)$$

Taking into account that

$$|\varphi(\tilde{X}_t) - \varphi(\tilde{X}_t^n)| \leq 2\|\varphi\|\,\mathbb{I}_{\{\tilde{X}_t \neq \tilde{X}_t^n\}},$$

or that, if φ is also a Lipschitz function,

$$|\varphi(\tilde{X}_t) - \varphi(\tilde{X}_t^n)| \leq L_\varphi\,\big|\tilde{X}_t - \tilde{X}_t^n\big|,$$

the previous observations can be used to get upper bounds for $E[dist(\pi_t, \tilde{\pi}_t^n)]$ when using the total variation or the bounded-Lipschitz metric. Indeed one has the following upper bounds for (9)

$$2\,\mathcal{Z}_t + 2\,\mathbb{P}\big(\{\tilde{X}_t \neq \tilde{X}_t^n\}/\mathcal{F}_t^{\tilde{Y}}\big), \qquad \text{when } \mathcal{K} = \mathcal{K}_{TV},$$

$$2\,\mathcal{Z}_t + \mathbb{E}\big[\big|\tilde{X}_t - \tilde{X}_t^n\big|/\mathcal{F}_t^{\tilde{Y}}\big], \qquad \text{when } \mathcal{K} = \mathcal{K}_{BL}.$$

Moreover, an easy computation gives

$$\mathbb{E}\big[\mathcal{Z}_t\big] = \mathbb{E}_{\mathbb{Q}}\big[\big|(d\mathbb{P}^n/d\mathbb{Q})|_{\tilde{\mathcal{F}}_t} - (d\mathbb{P}/d\mathbb{Q})|_{\tilde{\mathcal{F}}_t}\big|\big] =: \zeta_t^n. \quad (12)$$

As a consequence of the above analysis we get the following upper bounds.

Theorem II.2. *Under conditions* (a), (a$_n$), (b1), *and* (b2),

$$E\big[\|\pi_t - \tilde{\pi}_t^n\|_{TV}\big] \leq 2\,\zeta_t^n + 2\mathbb{P}\big(\{\tilde{X}_t \neq \tilde{X}_t^n\}\big) \qquad (13)$$

and

$$E\big[d_{BL}(\pi_t, \tilde{\pi}_t^n)\big] \leq 2\,\zeta_t^n + \mathbb{E}\big[\big|\tilde{X}_t - \tilde{X}_t^n\big|\big] \qquad (14)$$

Remark II.3. *Under* (a), (a$_n$), (b1) *and* (b2) *one can get sufficient conditions for the weak convergence of π_t^n to π_t (see Theorem 2.6 and Remark 2.7 in [3], see also [1] and [2]).*

C. Further upper bounds

In this section we give a slight generalization of Theorem II.2, which we use in the sequel to treat the delayed observations. To this end we introduce $\mathcal{K}'_{TV}(\alpha)$, the class of measurable functions $\phi(x, y)$ bounded above by α, and $\mathcal{K}'_{BL}(\alpha, \Lambda) \subseteq \mathcal{K}'_{TV}(\alpha)$ is the subclass of functions $\phi(x, y)$, such that for all x, x', and y, $|\phi(x, y) - \phi(x', y)| \leq \Lambda|x - x'|$. Note that, with a slight abuse of notation, we can write $\mathcal{K}_{TV}(\alpha) \subseteq \mathcal{K}'_{TV}(\alpha)$, and that $\mathcal{K}_{BL}(\alpha, \Lambda) \subseteq \mathcal{K}'_{BL}(\alpha, \Lambda)$.

Theorem II.4. *If conditions* (a), (a$_n$), (b1), *and* (b2) *hold, then* $\mathbb{E}\big[\sup_{\phi\in\mathcal{K}} |U(t, \tilde{\boldsymbol{Y}}; \phi(\cdot, \tilde{Y}_t)) - U^n(t, \tilde{\boldsymbol{Y}}; \phi(\cdot, \tilde{Y}_t))|\big]$ *is bounded above by*

$$C\,\alpha\,\zeta_t^n + 2\alpha\,\mathbb{P}\big(\{\tilde{X}_t \neq \tilde{X}_t^n\}\big), \quad \textit{when } \mathcal{K} = \mathcal{K}'_{TV}(\alpha)$$

$$C\,\alpha\,\zeta_t^n + \Lambda\,\mathbb{E}\big[\big|\tilde{X}_t - \tilde{X}_t^n\big|\big], \quad \textit{when } \mathcal{K} = \mathcal{K}_{BL}(\alpha, \Lambda)$$

with $C = 2$. The same upper bounds hold for $\sup_{\phi\in\mathcal{K}} \big|\mathbb{E}\big[\phi(\tilde{X}_t, \tilde{Y}_t)\big] - \mathbb{E}^n\big[\phi(\tilde{X}_t^n, \tilde{Y}_t)\big]\big|$*, but with $C = 1$.*

III. Approximation in filtering for Markov models with delayed observations

Let $(\boldsymbol{X^n}, \hat{\boldsymbol{Y}}^{\boldsymbol{n}})$ and $(\boldsymbol{X}, \hat{\boldsymbol{Y}})$ be Markov systems with generators $\boldsymbol{L^n}$ and $\boldsymbol{L}$ respectively. In this section we consider the systems $(\boldsymbol{X^n}, \boldsymbol{Y^n})$ and $(\boldsymbol{X}, \boldsymbol{Y})$ such that $Y_t^n = \hat{Y}_{a(t)}^n$, and $Y_t = \hat{Y}_{a(t)}$, where $a(\cdot)$ has the same properties as in Theorem I.1. We are interested in the approximation (6) of the filter, and therefore we need a representation formula for the functionals U^n and U. Thanks to Theorem I.1, these functionals can be expressed in terms of the corresponding functionals $\hat{U}^n$ and $\hat{U}$ of the underlying Markov systems, i.e., the functionals such that $\hat{\pi}_t^n = \hat{U}^n\big(t, \hat{\boldsymbol{Y}}^{\boldsymbol{n}}\big)$ and $\hat{\pi}_t = \hat{U}\big(t, \hat{\boldsymbol{Y}}\big)$ (note that the functionals $\hat{U}^n$ and $\hat{U}$ depend on the initial distributions and the generators of the corresponding Markov systems). Indeed

$$\begin{aligned}\pi_t^n(\varphi) &= \int_{\mathbb{R}^k} \hat{U}^n\big(a(t), \hat{\boldsymbol{Y}}^{\boldsymbol{n}}; dx\big) e^{\boldsymbol{L^n}(t-a(t))}\phi(x, Y_{a(t)}^n) \\ &= \hat{U}^n\big(r, \boldsymbol{y}; e^{\boldsymbol{L^n}(t-r)}\phi(\cdot, y_r)\big)\big|_{r=a(t), \boldsymbol{y}=\hat{\boldsymbol{Y}}^{\boldsymbol{n}}}\end{aligned}$$

and an analogous representation holds for $\pi_t(\varphi)$ by dropping the index n in the above expression. Therefore

$$|\pi_t(\varphi) - \tilde{\pi}_t^n(\varphi)| = \big|U\big(t, \boldsymbol{Y}; \varphi\big) - U^n\big(t, \boldsymbol{Y}; \varphi\big)\big|$$

can be written as

$$|\hat{U}\big(r, \boldsymbol{y}; e^{\boldsymbol{L}(t-r)}\phi(\cdot, y_r)\big) - \hat{U}^n\big(r, \boldsymbol{y}; e^{\boldsymbol{L^n}(t-r)}\phi(\cdot, y_r)\big)|_{\substack{r=a(t)\\ \boldsymbol{y}=\hat{\boldsymbol{Y}}}}$$

and is is bounded above by the sum of

$$|\hat{U}\big(r, \boldsymbol{y}; e^{\boldsymbol{L}(t-r)}\phi(\cdot, y_r)\big) - \hat{U}^n\big(r, \boldsymbol{y}; e^{\boldsymbol{L}(t-r)}\phi(\cdot, y_r)\big)|$$

and

$$|\hat{U}^n\big(r, \boldsymbol{y}; e^{\boldsymbol{L}(t-r)}\phi(\cdot, y_r)\big) - \hat{U}^n\big(r, \boldsymbol{y}; e^{\boldsymbol{L^n}(t-r)}\phi(\cdot, y_r)\big)|,$$

both evaluated in $r = a(t)$ and $\boldsymbol{y} = \hat{\boldsymbol{Y}}$. Moreover, natural upper bounds for the second addend are

$$\begin{aligned}&\sup_x \big|e^{\boldsymbol{L}(t-r)}\phi(x, y_r) - e^{\boldsymbol{L^n}(t-r)}\phi(x, y_r)\big| \\ &\leq \sup_{x,y} \big|e^{\boldsymbol{L}(t-r)}\phi(x, y) - e^{\boldsymbol{L^n}(t-r)}\phi(x, y)\big|.\end{aligned}$$

As a consequence the distance between the filter and the approximation can be bounded above by means of the distance between $\hat{U}^n$ and $\hat{U}$ and the distance between the semigroups $e^{\boldsymbol{L^n}s}$ and $e^{\boldsymbol{L}s}$:

Proposition III.1. *For the system with delayed observations described at the beginning of this section we have that* $\sup_{\varphi\in\mathcal{K}}|\pi_t(\varphi) - \tilde{\pi}_t^n(\varphi)|$ *is bounded above by the sum of*

$$\sup_{\psi\in\mathcal{K}_1} |\hat{U}(r, \boldsymbol{y}; \psi(\cdot, y_r)) - \hat{U}^n(r, \boldsymbol{y}; \psi(\cdot, y_r))|_{r=a(t), \boldsymbol{y}=\hat{\boldsymbol{Y}}}$$

and

$$\sup_{\phi\in\mathcal{K}_2} \sup_{x,y} |e^{\boldsymbol{L}(t-r)}\phi(x, y) - e^{\boldsymbol{L^n}(t-r)}\phi(x, y)|_{r=a(t)},$$

with $\mathcal{K}_1 = \mathcal{K}_2 = \mathcal{K}'_{TV}(\alpha)$, when $\mathcal{K} = \mathcal{K}_{TV}(\alpha)$.

Furthermore, if for all $u \geq 0$ there exists a constant $\Lambda'(u, \Lambda)$ such that $e^{\boldsymbol{L}u}\big(\mathcal{K}'_{BL}(\alpha,\Lambda)\big) \subseteq \mathcal{K}'_{BL}(\alpha, \Lambda'(u,\Lambda))$, then, the previous upper bound holds with $\mathcal{K}_1 = \mathcal{K}'_{BL}(\alpha, \Lambda'(t-r,\Lambda))$, and $\mathcal{K}_2 = \mathcal{K}'_{BL}(\alpha,\Lambda)$, when $\mathcal{K} = \mathcal{K}_{BL}(\alpha,\Lambda)$.

The addends in the above upper bounds will be evaluated when studying filtering systems with counting observations in the following section. In particular, upper bounds for their expectations will be obtained by a slight modification of the arguments used in Theorem II.4.

IV. Counting processes observations

In this section we assume that the pairs $(\boldsymbol{X}, \hat{\boldsymbol{Y}})$ and $(\boldsymbol{X^n}, \hat{\boldsymbol{Y}}^{\boldsymbol{n}})$ are Markov systems in $\mathbb{R}\times\mathbb{N}$, with respective initial distributions $\mu_0^X \otimes \delta_{\{y\}}$ and $\mu_0^{X^n} \otimes \delta_{\{y\}}$, and with respective generators $\boldsymbol{L}$ and $\boldsymbol{L^n}$, where

$$\boldsymbol{L}\phi(x,y) = \lambda_0(x,y)\int \big(\phi(x',y) - \phi(x,y)\big)\mu_0(x,y;dx') + \lambda_1(x,y)\int\big(\phi(x',y+1) - \phi(x,y)\big)\mu_1(x,y;dx')$$

and $\boldsymbol{L^n}$ has an analogous definition, but with $\lambda_i^n(x,y)$ and $\mu_i^n(x,y;\cdot)$ instead of $\lambda_i(x,y)$ and $\mu_i(x,y;\cdot)$, for $i=0,1$. Thus the predictable intensities of $\hat{\boldsymbol{Y}}$ and $\hat{\boldsymbol{Y}}^{\boldsymbol{n}}$ are $\hat\lambda_1(t) := \lambda_1(X_{t^-}, \hat{Y}_{t^-})$ and $\hat\lambda_1^n(t) := \lambda_1^n(X^n_{t^-}, \hat{Y}^n_{t^-})$, respectively. Furthermore we assume:

(H_0) $0 \leq \underline{\lambda}_i \leq \lambda_i(x,y), \lambda_i^n(x,y) \leq \overline{\lambda}_i, \quad i = 0,1;$
(H_1) $0 < \lambda_1(x,y), \lambda_1^n(x,y);$
(H_2) $\Delta_i^n := \sup_{x,y}\boldsymbol{\kappa}(\mu_i(x,y;\cdot), \mu_i^n(x,y;\cdot)) < \infty$, $i = 0,1$.

Note that the operator $\boldsymbol{L}$ given above can be expressed in the same form as the one defined in the introduction and vice versa. From now on we will say that the operator $\boldsymbol{L}$ is *bL-regular* when it satisfies the following conditions: (R_0) For every z the function $x \mapsto \lambda_i(x,z)$ is bounded-Lipschitz continuous and the Lipschitz constant is bounded from above by L_{λ_i}, $i=0,1$,; (R_1) $\sup_y \boldsymbol{\kappa}(\mu_i(x,y;\cdot), \mu_i(x',y;\cdot)) \leq \Gamma_{\mu_i}|x-x'|$, $i = 0,1$; (R_2) $\sup_{x,y}\int|z-x|\mu_0(x,y;dz) \leq b_0$, and $\int|z-x|\mu_1(x,y;dz) \leq a_1(|x|+y)+b_1$.

It is important to stress that, if $\boldsymbol{L}$ is bL-regular, then, for a suitable constant M (given below in Proposition IV.1), $e^{\boldsymbol{L}u}(\mathcal{K}'_{BL}(\alpha,\Lambda)) \subseteq \mathcal{K}'_{BL}(\alpha, \Lambda e^{Mu})$ (for the proof see Corollary 4.9 in [3]), and thus Proposition III.1 holds with $\Lambda'(u,\Lambda) = \Lambda e^{Mu}$.

The central tool for the proofs of our bounds (see Theorem IV.3 below) is a particular construction of the pairs $(\boldsymbol{X^n}, \hat{\boldsymbol{Y}}^{\boldsymbol{n}})$ and $(\boldsymbol{X}, \hat{\boldsymbol{Y}})$ on the same measurable space $(\Omega, \mathcal{F})$, equipped with three different probability measures $\mathbb{Q}$, $\mathbb{P}$, and $\mathbb{P}^n$, in a similar way to that used in section II, and such that: $(\hat{\mathrm{a}})$ under $\mathbb{P}$ the pair $(\boldsymbol{X}, \hat{\boldsymbol{Y}})$ has generator $\boldsymbol{L}$ and initial distribution $\mu = \mu_0^X \otimes \delta_{\{y\}}$; $(\hat{\mathrm{a}}_\mathrm{n})$ under $\mathbb{P}^n$ the pair $(\boldsymbol{X^n}, \hat{\boldsymbol{Y}})$ has generator $\boldsymbol{L}^n$ and initial distribution $\mu^n = \mu_0^{X^n}\otimes\delta_{\{y\}}$; $(\hat{\mathrm{b}})$ the probability measures $\mathbb{Q}$, $\mathbb{P}$, and $\mathbb{P}^n$ are equivalent on $\widehat{\mathcal{F}}^n_t = \mathcal{F}_t^{X,X^n,\hat{Y}}$ (note that, in order to avoid the notation $\hat{Y}$, we have used the notation $(\boldsymbol{X}, \hat{\boldsymbol{Y}})$ instead of $(\tilde{\boldsymbol{X}}, \tilde{\hat{\boldsymbol{Y}}})$). Then, in this framework, Theorem II.4 applies with $(\boldsymbol{X}, \boldsymbol{X^n}, \hat{\boldsymbol{Y}})$ instead of $(\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{X}}^{\boldsymbol{n}}, \tilde{\boldsymbol{Y}})$ and with $\hat{U}$ and $\hat{U}^n$ instead of U and U^n, respectively. Furthermore (see (4) and (5)) $\hat{U}(t,\boldsymbol{y};\varphi)$ is defined as $\hat\Pi_t(\varphi|\boldsymbol{s})$ when $y(t) = \sum_{i=1}^\infty \mathbb{I}_{[0,s_i]}(t)$, and arbitrarily defined otherwise. An analogous representation holds for $\hat{U}^n$. The construction of the triplet $(\boldsymbol{X}, \boldsymbol{X^n}, \hat{\boldsymbol{Y}})$ is possible under the assumptions (H_0) and (H_1). The space $(\Omega, \mathcal{F}, \mathbb{Q})$ is a probability space on which two independent Poisson random measures are defined: $\mathcal{N}_0(ds,d\zeta)$ on $[0,T]\times[0,\bar\lambda_0]$, with intensity measure $ds\otimes d\zeta$, and $\mathcal{N}_1(ds,d\zeta)$ on $[0,T]\times[0,1]$, with intensity measure $ds\otimes d\zeta$. Then for suitable functions K_0, K_1, K_0^n, K_1^n (see Remark 4.1 in [3] and Lemma 3.7 in [6]), the triplet $(\boldsymbol{X}, \boldsymbol{X^n}, \hat{\boldsymbol{Y}})$ is

$$X_t = X_0 + \int_0^t\int_0^{\bar\lambda_0} K_0(X_{s^-}, X^n_{s^-}, \hat{Y}_{s^-};\zeta)\mathcal{N}_0(ds,d\zeta) + \int_0^t\int_0^1 K_1(X_{s^-}, X^n_{s^-}, \hat{Y}_{s^-};\zeta)\mathcal{N}_1(ds,d\zeta),$$

$$X^n_t = X^n_0 + \int_0^t\int_0^{\bar\lambda_0} K^n_0(X_{s^-}, X^n_{s^-}, \hat{Y}_{s^-};\zeta)\mathcal{N}_0(ds,d\zeta) + \int_0^t\int_0^1 K^n_1(X_{s^-}, X^n_{s^-}, \hat{Y}_{s^-};\zeta)\mathcal{N}_1(ds,d\zeta),$$

$$\hat{Y}_t = y + \int_0^t\int_0^1 \mathcal{N}_1(ds,d\zeta) = y + \mathcal{N}_1((0,t]\times[0,1]),$$

with initial conditions (X_0, X_0^n) independent of $\mathcal{N}_0$ and $\mathcal{N}_1$, $d\mathbb{P}/d\mathbb{Q}\big|_{\widehat{\mathcal{F}}^n_t} = \exp\big\{\int_0^t \log\big(\hat\lambda_1(s)\big)d\hat{Y}_s - \int_0^t\big(\hat\lambda_1(s)-1\big)ds\big\}$, and $d\mathbb{P}^n/d\mathbb{Q}\big|_{\widehat{\mathcal{F}}^n_t}$ is defined analogously. In this framework two preliminary propositions hold.

Proposition IV.1. *Under the hypotheses* (H_0) *and* (H_1)*, if $\mathcal{K} = \mathcal{K}'_{TV}(\alpha)$, then*

$$\mathbb{E}\big[\sup_{\psi\in\mathcal{K}}\big|\hat{U}\big(r,\boldsymbol{y};\psi(\cdot,y_r)\big) - \hat{U}^n\big(r,\boldsymbol{y};\psi(\cdot,y_r)\big)\big|_{\boldsymbol{y}=\hat{\boldsymbol{Y}}}\big] \quad (15)$$

is bounded above by

$$\alpha\big[4\|\lambda_1 - \lambda_1^n\|r + e^{(\bar\lambda_1 - \underline{\lambda}_1)\,r}\,C^n\big],$$

where

$$C^n = \|\mu_0^X - \mu_0^{X^n}\|_{TV}\big(\frac{2\bar\lambda_1}{\bar\lambda_1 - \underline{\lambda}_1} + 1\big) + (\bar\lambda_0 + \bar\lambda_1)J^n\,A(r),$$

$$A(r) = 2r + 4\bar\lambda_1\frac{e^{-(\bar\lambda_1 - \underline{\lambda}_1)\,r} + (\bar\lambda_1 - \underline{\lambda}_1)r - 1}{(\bar\lambda_1 - \underline{\lambda}_1)^2},$$

$$J^n = \max(\|p_0 - p_0^n\|_\infty, \|\mu_1 - \mu_1^n\|_\infty)/2,$$

with $\|\nu\|_\infty = \sup_{x,y}\|\nu(x,y,\cdot)\|_{TV}$, $p_0(x,y;\cdot)$ defined by $\big(1 - (\lambda_0(x,y)/\bar\lambda_0)\big)\delta_x(\cdot) + \big(\lambda_0(x,y)/\bar\lambda_0\big)\ \mu_0(x,y;\cdot)$, and $p_0^n(x,y;\cdot)$ defined by a similar expression, but involving λ_0^n and μ_0^n.

If furthermore the operator $\boldsymbol{L}$ is bL-regular, assumption (H_2) *holds, and the initial distributions μ_0^X and $\mu_0^{X^n}$ have finite first moments, and if $\mathcal{K} = \mathcal{K}'_{BL}(\alpha,\Lambda')$, then* (15) *is bounded above by*

$$4\alpha\big[\|\lambda_1 - \lambda_1^n\|r + L_{\lambda_1}H^n\,\boldsymbol{\mathcal{E}}_2^M(r)\big] + \Lambda' H^n\,\boldsymbol{\mathcal{E}}_1^M(r) + \boldsymbol{\kappa}\big(\mu_0^X, \mu_0^{X^n}\big)\big(4\alpha L_{\lambda_1}\boldsymbol{\mathcal{E}}_1^M(r) + \Lambda'\boldsymbol{\mathcal{E}}_0^M(r)\big),$$

with $H^n = \bar\lambda_0 \Delta_0^n + b_0 \, \|\lambda_0 - \lambda_0^n\| + \bar\lambda_1 \Delta_1^n$,

$$M = \sum_{i=0}^{1} \big[\bar\lambda_i (\Gamma_{\mu_i} - 1)^+ - \underline{\lambda}_i (1 - \Gamma_{\mu_i})^+\big] + b_0 L_{\lambda_0},$$

$$\boldsymbol{\mathcal{E}}_k^M(t) := \frac{1}{M^k} \sum_{h=k}^{\infty} \frac{(tM)^h}{h!} = \frac{1}{M^k}\left(e^{tM} - \sum_{h=0}^{k-1} \frac{(tM)^h}{h!}\right).$$

Proposition IV.2. *Under the assumptions* $(\mathrm{H}_0), (\mathrm{H}_1)$ *and the notation of Proposition* $IV.1$, *if* $\mathcal{K} = \mathcal{K}'_{TV}(\alpha)$, *the*

$$\sup\nolimits_{\phi\in\mathcal{K}} \sup\nolimits_{x,y} \left| e^{\boldsymbol{L} t}\phi(x,y) - e^{\boldsymbol{L^n} t}\phi(x,y)\right| \tag{16}$$

is bounded above by

$$\alpha\big(2\|\lambda_1 - \lambda_1^n\|\, t + e^{(\bar\lambda_1 - \underline{\lambda}_1)t} (\bar\lambda_0 + \bar\lambda_1)\, J^n\, B(t)\big),$$

where $B(t) = 2t + A(t)/2$.

If furthermore the operator $\boldsymbol{L}$ *is bL-regular, assumption* (H_2) *holds, and the initial distributions* μ_0^X *and* $\mu_0^{X^n}$ *have finite first moments, and if* $\mathcal{K} = \mathcal{K}'_{BL}(\alpha, \Lambda)$, *then* (16) *is bounded above by*

$$2\alpha\big[\|\lambda_1 - \lambda_1^n\| t + L_{\lambda_1} H^n\, \boldsymbol{\mathcal{E}}_2^M(t)\big] + \Lambda H^n\, \boldsymbol{\mathcal{E}}_1^M(t).$$

The proof of the above results (see [3]) is based on Theorem II.4 with $(\boldsymbol{X}, \boldsymbol{X^n}, \hat{\boldsymbol{Y}})$ instead of $(\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{X}}^{\boldsymbol{n}}, \tilde{\boldsymbol{Y}})$, some upperbounds for $\mathbb{P}\big(\{X_t \neq X_t^n\}\big)$ and $\mathbb{E}\big(|X_t - X_t^n|\big)$ (see Propositions 4.6 and 4.8 in [3]), and the observation that the semigroups can be represented, respectively, as $e^{\boldsymbol{L}t}\phi(x,y) = \mathbb{E}\big[\phi(X_t, \hat{Y}_t)\big]$ and $e^{\boldsymbol{L^n}t}\phi(x,y) = \mathbb{E}^n\big[\phi(X_t^n, \hat{Y}_t)\big]$, when $\mu_0^X = \delta_{\{x\}}$ and $\mu_0^{X^n} = \delta_{\{x\}}$. Finally, using the results and the notation of Propositions III.1, with $\Lambda'(u, \Lambda) = \Lambda\, e^{M(t-r)}$, $\alpha = 1$, and $\Lambda = 1$, Proposition IV.1 with $\Lambda' = \Lambda'(t-r, 1)$, and Proposition IV.2, we get immediately the main result of this section:

Theorem IV.3. *Under the assumptions* (H_0) *and* (H_1), *and the notation introduced in Propositions* $IV.1$ *and* $IV.2$, $E\big[\|\pi_t - \tilde\pi_t^n\|_{TV}\big]$ *is bounded above by*

$$\begin{aligned} &2\|\lambda_1 - \lambda_1^n\|(t+r) + \|\mu_0^X - \mu_0^{X^n}\|_{TV}\, e^{(\bar\lambda_1 - \underline{\lambda}_1)\, r}\, \frac{3\bar\lambda_1 - \underline{\lambda}_1}{\bar\lambda_1 - \underline{\lambda}_1} \\ &+ (\bar\lambda_0 + \bar\lambda_1) J^n \big(e^{(\bar\lambda_1 - \underline{\lambda}_1)\, r} A(r) + e^{(\bar\lambda_1 - \underline{\lambda}_1)\,(t-r)} B(t-r)\big) \end{aligned}$$

evaluated in $r = a(t)$. *Furthermore assume* (H_2), *the bL-regularity of* $\boldsymbol{L}$, *and that the initial distributions* μ_0^X *and* $\mu_0^{X^n}$ *have finite first moments. Then* $E\big[d_{BL}(\pi_t, \tilde\pi_t^n)\big]$ *is bounded above by*

$$\begin{aligned} &4\big[\|\lambda_1 - \lambda_1^n\| r + L_{\lambda_1} H^n\, \boldsymbol{\mathcal{E}}_2^M(r)\big] + e^{M(t-r)} H^n\, \boldsymbol{\mathcal{E}}_1^M(r) \\ &+ \boldsymbol{\kappa}\big(\mu_0^X, \mu_0^{X^n}\big)\big(4 L_{\lambda_1} \boldsymbol{\mathcal{E}}_1^M(r) + \boldsymbol{\mathcal{E}}_0^M(t)\big) + H^n\, \boldsymbol{\mathcal{E}}_1^M(t-r) \\ &+ 2\big(\|\lambda_1 - \lambda_1^n\|(t-r) + L_{\lambda_1} H^n\, \boldsymbol{\mathcal{E}}_2^M(t-r)\big) \end{aligned}$$

evaluated in $r = a(t)$.

When the measures $\mu_i(x, y, \cdot)$ are Gaussian, then the constants involved in the previous upper bounds can be computed explicitely (see the *Example* in section 4 of [3]). Furthermore we observe that M may be negative, and then, when $\boldsymbol{L^n} = \boldsymbol{L}$, the above theorem is a kind of stability property for the filter with respect to the initial conditions (see the the related discussion in section 4 of [3]).

References

[1] A. G. Bhatt, G. Kallianpur, and R. L. Karandikar, *Robustness of the nonlinear filter*, Stochastic Process. Appl., 81 (1999), pp. 247–254.

[2] A. G. Bhatt and R. L. Karandikar, *Robustness of the nonlinear filter: The correlated case*, Stochastic Process. Appl., 97 (2002), pp. 41–58.

[3] A. Calzolari, P. Florchinger, and G. Nappo, *Approximation of nonlinear filters for Markov systems with delayed observations*, SIAM J. Control Optim., Vol. 45, No. 2, pp. 599633.

[4] A. Calzolari, P. Florchinger, and G. Nappo, *Nonlinear Filtering for Markov Systems with Delayed Observations*, preprint, Department of Mathematics, University of Rome "La Sapienza," Rome, Italy, 2004. Available online at http://www.mat.uniroma1.it/people/nappo/papers.pdf/Delay42.pdf.

[5] A. Calzolari, P. Florchinger, and G. Nappo, *Convergence in Nonlinear Filtering for Stochastic Delay Systems*, preprint, Department of Mathematics, University of Rome "La Sapienza," Rome Italy, 2005. Available online at http://www.mat.uniroma1.it/people/nappo/papers.pdf/CFN-conv-delay-rate.pdf.

[6] A. Calzolari and G. Nappo, *Counting observations: A note on state estimation sensitivity with an L^1-bound*, Appl. Math. Optim., 44 (2001), pp. 177–201.

[7] C. Ceci, A. Gerardi, and P. Tardelli, *An estimate of the approximation error in the filtering of a discrete jump process*, Math. Models Methods Appl. Sci., 11 (2001), pp. 181–198.

[8] P. Del Moral, *Measure-valued processes and interacting particle systems. Application to nonlinear filtering problems*, Ann. Appl. Probab., 8 (1998), pp. 438–495.

[9] G. B. Di Masi, M. Pratelli, and W. J. Runggaldier, *An approximation for the nonlinear filtering problem, with error bound*, Stochastics, 14 (1985), pp. 247–271.

[10] R. J. Elliott and R. Glowinski, *Approximations to solutions of the Zakai filtering equation*, Stochastic Anal. Appl., 7 (1989), pp. 145–168.

[11] M. Kirch and W. J. Runggaldier, *Efficient hedging when asset prices follow a geometric Poisson process with unknown intensities*, SIAM J. Control Optim., 43 (2004), pp. 1174–1195.

[12] W. Kliemann, G. Koch, and F. Marchetti, *On the unnormalized solution of the filtering problem with counting observations*, IEEE Trans. Inform. Theory, 316 (1990), pp. 1415–1425.

[13] T. G. Kurtz and D. L. Ocone, *Unique characterization of conditional distributions in nonlinear filtering*, Ann. Probab., 16 (1988), pp. 80–107.

[14] H. J. Kushner, *Probability Methods for Approximations in Stochastic Control and for Elliptic Equations*, Math. Sci. Engrg. 129, Academic Press (Harcourt Brace Jovanovich, Publishers), New York, London, 1977.

[15] F. Le Gland, *Time discretization of nonlinear filtering equations*, in Proceedings of the 28th IEEE Conference on Decision and Control (Tampa, FL), Vol. 3, 1989, pp. 2601–2606.

[16] S. Lototsky, R. Mikulevicius, and B. L. Rozovskii, *Nonlinear filtering revisited: A spectral approach*, SIAM J. Control Optim., 35 (1997), pp. 435–461.

[17] É. Pardoux, *Filtrage non linéaire et équations aux dérivées partielles stochastiques associées*, in École d'Été de Probabilités de Saint-Flour XIX—1989, Lecture Notes in Math. 1464, Springer, Berlin, 1991, pp. 67–163.

[18] M. Schweizer, *Risk-minimizing hedging strategies under restricted information*, Math. Finance, 4 (1994), pp. 327–342.