Approximation of nonlinear filters for Markov systems with delayed observations *

Antonella Calzolari [†]   Patrick Florchinger [‡]   Giovanna Nappo [§]

Abstract

The aim of this paper is to give some approximation results for a class of nonlinear filtering problems with delay in the observation. First, we point out some general results on the approximation problem for the filter in nonlinear filtering. In particular we give a general procedure to obtain some upper bounds for the different approximations we consider. This procedure is then applied in the case of nonlinear filtering problems with delay (X, Y), which can be represented by means of a Markov system $(X, \hat{Y})$, in the sense that $Y_t = \hat{Y}_{a(t)}$. Finally these upper bounds are computed explicitly in the particular case of Markov jump process with counting observations.

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

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

1 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 $\sigma-$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 (see e.g. Pardoux [20] and the references therein).

Filtering problems involving jump-diffusion processes have been studied by many authors. In particular the jump system

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

$$y_t = \int_0^t \int_{D_1(x_{s^-}, y_{s^-})} \mathcal{N}(ds, d\zeta) \tag{3}$$

where

[†]Dipartimento di Matematica - Università di Roma "Tor Vergata", via della Ricerca Scientifica 1, I 00133 Roma, Italy
[‡]Département de Mathématiques, Université de Metz, 23 Allée des Oeillets F 57160 Moulins les Metz, France
[§]Dipartimento di Matematica - Università di Roma "La Sapienza", piazzale A. Moro 2, I 00185 Roma, Italy

- $\mathcal{N}(ds, d\zeta)$ is a Poisson measure on $\mathbb{R} \times \Sigma$ with mean measure $ds \times \nu(d\zeta)$,

- 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,

enters in the more general framework studied by Kliemann, Koch and Marchetti [15]. Under suitable hypotheses (see [15]), 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

$$
\begin{aligned}
\boldsymbol{L}f(x,y) = &\int_{D_0(x,y)} \left[f(x + K_0(x,y;\zeta), y) - f(x,y) \right] \nu(d\zeta) \\
&+ \int_{D_1(x,y)} \left[f(x + K_1(x,y;\zeta), y+1) - f(x,y) \right] \nu(d\zeta).
\end{aligned}
\tag{4}
$$

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

$$
B^y \varphi(x) = \int_{D_0(x,y)} \left[\varphi(x + K_0(x,y;\zeta)) - \varphi(x) \right] \nu(d\zeta)
\tag{5}
$$

and

$$
R^y \varphi(x) = \int_{D_1(x,y)} \left[\varphi(x + K_1(x,y;\zeta)) - \varphi(x) \right] \nu(d\zeta),
\tag{6}
$$

- the unnormalized filter $\hat{\rho}_t(dx|\boldsymbol{s})$ is given by the following self-contained procedure

$$
\hat{\rho}_t(\varphi|\boldsymbol{s}) = E\left[\varphi(X_{t-s_i}^i(\boldsymbol{s})) \exp\left\{ -\int_0^{t-s_i} \lambda_1(X_u^i(\boldsymbol{s}), i) du \right\} \right] \qquad s_i \leq t < s_{i+1},
\tag{7}
$$

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

$$
\hat{\rho}_{s_0}(\varphi|\boldsymbol{s}) = \mu_0^x(\varphi) \qquad \hat{\rho}_{s_{i+1}}(\varphi|\boldsymbol{s}) = \frac{\hat{\rho}_{s_{i+1}^-}(Q^i \varphi|\boldsymbol{s})}{\hat{\rho}_{s_{i+1}^-}(\lambda_1(\cdot, i)|\boldsymbol{s})}
\tag{8}
$$

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

Setting

$$
\hat{\Pi}_t(\varphi|\boldsymbol{s}) := \frac{\hat{\rho}_t(\varphi|\boldsymbol{s})}{\hat{\rho}_t(\boldsymbol{1}|\boldsymbol{s})},
\tag{9}
$$

we have

$$
\pi_t^x(\varphi) = E[\varphi(x_t)/\mathcal{F}_t^y] = \hat{\Pi}_t(\varphi|\boldsymbol{s})\Big|_{\boldsymbol{s} = \{T_i\}_{i \geq 0}}
\tag{10}
$$

where T_i are the observed jump times.

Different approximation procedures and numerical schemes have studied by many authors in the past years (see for example among others Kushner [17], Di Masi, Pratelli and Runggaldier [12], Le Gland

[18], Elliott and Glowinski [13], Lototsky, Mikulevicius and Rozovskii [19], Del Moral [11], Calzolari and Nappo [8] or Ceci, Gerardi and Tardelli [9] 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: $(\boldsymbol{X}, \hat{\boldsymbol{Y}})$ is a Markov process with generator $\boldsymbol{L}$, and the observation process $\boldsymbol{Y}$ satisfies

$$Y_t = \hat{Y}_{a(t)},$$

where $a(\cdot)$ is a nondecreasing continuous function, such that $a(0) = 0$, $a(t) \leq t$ for all $t \geq 0$.

For this system the following result has been proved by the authors in [4].

Theorem 1.1. *The filter π_t associated with the system $(\boldsymbol{X}, \boldsymbol{Y})$ described above can be represented as*

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

where $\phi(x, y) = \varphi(x)$.

Moreover, if the conditional law of X_s given $\mathcal{F}^{\hat{Y}}_s$ is known and denoted by $\hat{\pi}_s$, then

$$\pi_t(\varphi) = \hat{\pi}_{a(t)}\big((\exp\{\boldsymbol{L}(t - a(t))\}\phi)(\cdot, \hat{Y}_{a(t)})\big). \tag{11}$$

Remark 1.2. *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. In this case the generator $\boldsymbol{L}$ coincides with the one given in (4).*

Remark 1.3. *The case when the state process solves the stochastic differential equation*

$$X_t = X_0 + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dB_s, \qquad t \geq 0,$$

and the observation is given by

$$\begin{cases} Y_t = 0, & 0 \leq t \leq \tau; \\ Y_t = \int_\tau^t h(X_{s-\tau})ds + W_t - W_\tau, & t \geq \tau \end{cases} \tag{12}$$

corresponds to $a(t) = (t - \tau)^+$ and to the system $(\boldsymbol{X}, \hat{\boldsymbol{Y}})$ with the generator

$$\boldsymbol{L}f(x, y) = b(x)\nabla_x f(x, y) + h(x)\nabla_y f(x, y) + \frac{1}{2}tr\{\sigma(x)\sigma^*(x)\nabla^2_x f(x, y)\} + \frac{1}{2}\Delta_y f(x, y)$$

in the above theorem. On the other hand, by applying a shift in time of size τ to the system $(\boldsymbol{X}, \boldsymbol{Y})$, we get a stochastic delay system: indeed setting $X_{(\tau)}(s) := X_{s+\tau}$ and $Y_{(\tau)}(s) := Y_{s+\tau}$, we have

$$dX_{(\tau)}(s) = b\big(X_{(\tau)}(s)\big)ds + \sigma\big(X_{(\tau)}(s)\big)d\tilde{B}_s, \quad s \geq 0, \qquad X_{(\tau)}(s) = \xi(s), \quad -\tau \leq s \leq 0 \tag{13}$$

$$Y_{(\tau)}(t) = \int_0^t h\big(X_{(\tau)}(s - \tau)\big)ds + \hat{W}_t, \qquad t \geq 0, \tag{14}$$

where $\tilde{B}_s = B_{s+\tau}$, and $\xi(s)$ is the solution of the stochastic differential equation

$$d\xi(s) = b(\xi(s))ds + \sigma(\xi(s))d\tilde{B}_s, \text{ for } -\tau \leq s \leq 0, \qquad \text{with } \xi(-\tau) = X(0).$$

Clearly, in the above system the delay is present only in the observation.

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 [21], Schweizer has given an example of information with delay for a financial model. From [21] one can easily deduce that when the market is incomplete the risk minimization criterion leads to a filtering problem with delayed observation. Risk minimization criterion corresponds to a quadratic loss function, and this is criticized as it implies

the minimization of both earnings and losses. This leads to use different loss functions and 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 [14], 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 [14], 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 not be any more 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 4. Another related situation fitting the same framework could arise if the interest were focused on a fixed strike lookback call option, that is

$$\left(\max_{t \in [0,T]} S_t - K \right)^+.$$

In this case, instead of the price, a natural process to observe could be the running maximum of the price $\max_{u \in [0,t]} S_u$, the observation of which is equivalent to the observation of one counting process.

This paper is divided into four sections and is organized as follows. In Section 2, we introduce different kinds of approximation for the filter that we will use in the sequel. These approximation schemes rely on the actually observed trajectories, and upper bounds for these approximations are obtained in Theorem 2.3. A relation between these upper bounds and weak convergence of the sequence of filters is presented in Theorem 2.6. In Section 3, we consider the approximation of filter for Markov models with delayed observations for which we obtain upper bounds depending on the actually observed trajectory (see Proposition 3.1). In Section 4, the approximation results obtained in Section 2 together with the representation for the filter in Theorem 1.1 will be applied to obtain some explicit upper bounds in the counting observation case as in Remark 1.2 (see Theorem 4.5).

Some approximation results can be obtained for different systems and by making use of different techniques. As an example we can consider the simple stochastic delay system of Remark 1.3. Taking into account this remark, the computation of the filter is equivalent to the computation of the filter of the stochastic delay system (13) and (14), for which the convergence result obtained in [5] can be applied. To our knowledge there are only two other papers dealing with the approximation of nonlinear filtering for delayed diffusion systems: Chang [10], and Twardowska, Marnik and Pasławaska-Połuniak [22].

2 General considerations on approximation for filters

2.1 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 arising 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. In this case one can furthermore 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 $\mathcal{P} = \mathcal{P}(\mathbb{R}^k)$ the space of probability measures on $\mathbb{R}^k$, such as the the total variation, i.e.

$$\|\nu_1 - \nu_2\|_{TV} = \sup \left\{ \frac{|\nu_1(\varphi) - \nu_2(\varphi)|}{\|\varphi\|}; \varphi \text{ bounded} \right\},$$

the Kantorovitch metric, i.e.

$$\boldsymbol{\kappa}(\nu_1, \nu_2) = \sup \left\{ \frac{|\nu_1(\varphi) - \nu_2(\varphi)|}{L_\varphi}; \varphi \text{ Lipschitz} \right\},$$

or the bounded-Lipschitz metric, i.e.

$$d_{BL}(\nu_1, \nu_2) = \sup \left\{ \frac{|\nu_1(\varphi) - \nu_2(\varphi)|}{\|\varphi\| \vee L_\varphi}; \varphi \text{ bounded and Lipschitz} \right\}$$

where $\|\varphi\|$ denotes the sup-norm, and L_φ is the Lipschitz constant of φ. We recall that the bounded-Lipschitz metric induces the weak convergence topology on the space of probability measures $\mathcal{P}$.

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 $U^n\big(t, \boldsymbol{y}\big) = U^n\big(t, y(\cdot \wedge t)\big)$ and $U\big(t, \boldsymbol{y}\big) = U\big(t, y(\cdot \wedge t)\big)$,

$$\pi_t^n(dx) = U^n\big(t, \boldsymbol{Y^n}; dx\big) \text{ and } \pi_t(dx) = U\big(t, \boldsymbol{Y}; dx\big),$$

then, in order to consider convergence in probability, one only needs 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 [16])). 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 that 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 like 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 (9), or like the classical diffusive model, for which the functionals U and U^n are also computable as exposed in [16], for example.

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

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 these situations

1.) the *true* filter is π_t, and it is natural to consider

$$\tilde{\pi}_t^n = U^n(t, \boldsymbol{Y}) \tag{15}$$

as an approximation of π_t, depending on the actually observed trajectory,

2.) the *true* filter is π_t^n, and it is natural to consider

$$\overline{\pi}_t^n = U(t, \boldsymbol{Y^n}) \tag{16}$$

as an approximation of π_t^n, depending on the actually observed trajectory.

The functional U^n is defined $P_{\boldsymbol{Y^n}}$-almost surely, therefore in order to define $\tilde{\boldsymbol{\pi}}^{\boldsymbol{n}} = \{\tilde{\pi}_t^n; t \geq 0\}$ almost surely it is natural to assume that $P_{\boldsymbol{Y}}$ is absolutely continuous with respect to $P_{\boldsymbol{Y^n}}$. Indeed when the functional U^n is defined $P_{\boldsymbol{Y^n}}$-almost surely, there exists a set $D^n \subset D_{\mathbb{R}^d}([0, T])$ with $P_{\boldsymbol{Y^n}}(D^n) = 1$, such that the filter π_t^n is almost surely equal to $\mathbf{1}_{D^n}(\boldsymbol{Y^n})U^n(t, \boldsymbol{Y^n}) + (1 - \mathbf{1}_{D^n}(\boldsymbol{Y^n}))\delta_{\{x_0\}}$, for any fixed point x_0, and one could always define

$$\tilde{\pi}_t^n = \mathbf{1}_{D^n}(\boldsymbol{Y})U^n(t, \boldsymbol{Y}) + (1 - \mathbf{1}_{D^n}(\boldsymbol{Y}))\delta_{\{x_0\}}.$$

Furthermore if $P_{\boldsymbol{Y}}(D^n) = 0$, as an extreme example, then the chosen approximation would be clearly useless: indeed, in this case $\tilde{\pi}_t^n = \delta_{\{x_0\}}$ almost surely, and this is clearly not a good approximation, while if $P_{\boldsymbol{Y}}(D^n) = 1$, then the approximation coincide almost surely with $U^n(t, \boldsymbol{Y})$.

In analogy, the functional U is defined $P_{\boldsymbol{Y}}$-almost surely, and to define $\overline{\boldsymbol{\pi}}^{\boldsymbol{n}} = \{\overline{\pi}_t^n; t \geq 0\}$ almost surely it is natural to assume that $P_{\boldsymbol{Y^n}}$ is absolutely continuous with respect to $P_{\boldsymbol{Y}}$.

When using these kinds of approximations one is interested in evaluating

$$dist(\pi_t, \tilde{\pi}_t^n) = dist\left(U(t, \boldsymbol{Y}), U^n(t, \boldsymbol{Y})\right), \tag{17}$$

when we observe $\boldsymbol{Y}$ (situation **1**), or

$$dist(\pi_t^n, \overline{\pi}_t^n) = dist\left(U^n(t, \boldsymbol{Y^n}), U(t, \boldsymbol{Y^n})\right) \tag{18}$$

when we observe $\boldsymbol{Y^n}$ (situation **2**). Here the $dist(\nu_1, \nu_2)$ can be, for example, the total variation, the Kantorovitch metric or the bounded-Lipschitz metric.

Note that, in contrast to the approximations previously considered, neither $\tilde{\pi}_t^n$ nor $\overline{\pi}_t^n$ are conditional laws. Moreover, with this kind of approximation it is not even necessary that the sequence of processes $\{\boldsymbol{Y^n}\}$ and $\boldsymbol{Y}$ are defined on the same probability space, and both the almost sure convergence and the convergence in probability can be considered.

For instance, in situation **1**, the convergence in probability is implied by the convergence to zero of

$$E\left[dist(\pi_t, \tilde{\pi}_t^n)\right] = E\left[dist\left(U(t, \boldsymbol{Y}), U^n(t, \boldsymbol{Y})\right)\right] \tag{19}$$

$$= \int_{D_{\mathbb{R}^d}([0,T])} dist\left(U(t, \boldsymbol{y}), U^n(t, \boldsymbol{y})\right) dP_{\boldsymbol{Y}}(\boldsymbol{y}),$$

while, in situation **2**, the convergence in probability is implied by the convergence to zero of

$$E\left[dist(\pi_t^n, \overline{\pi}_t^n)\right] = E\left[dist\left(U^n(t, \boldsymbol{Y^n}), U(t, \boldsymbol{Y^n})\right)\right] \tag{20}$$

$$= \int_{D_{\mathbb{R}^d}([0,T])} dist\left(U^n(t, \boldsymbol{y}), U(t, \boldsymbol{y})\right) dP_{\boldsymbol{Y^n}}(\boldsymbol{y}).$$

Finally note that when the distance between π_t and $\tilde{\pi}_t^n$ is given in terms of one of the integral metrics quoted at the beginning of this section, then

$$E\big[dist(\pi_t, \tilde{\pi}_t^n)\big] = E\big[\sup_{\varphi \in \mathcal{K}} |\pi_t(\varphi) - \tilde{\pi}_t^n(\varphi)|\big] \tag{21}$$

and

$$E\big[dist(\pi_t^n, \overline{\pi}_t^n)\big] = E\big[\sup_{\varphi \in \mathcal{K}} |\pi_t^n(\varphi) - \overline{\pi}_t^n(\varphi)|\big] \tag{22}$$

where $\mathcal{K}$ is a suitable class of functions: the class $\mathcal{K}_{TV}$ of functions φ with $\|\varphi\| \leq 1$ for the total variation, the class $\mathcal{K}_{BL}$ of functions φ with $\|\varphi\| \vee L_\varphi \leq 1$ for the bounded Lipschitz metric, and the class $\mathcal{K}_K$ of functions φ with $L_\varphi \leq 1$ for the Kantorovitch metric.

2.2 General upperbounds for the approximations of the filters

In the following we study the approximation problem for the filter by using a general procedure leading to a partial result, which is summarized in Theorem 2.3. Moreover in the following we also consider the connections with weak convergence of the filters (cf. Theorem 2.6 and Remark 2.7). However, since as it will become clear in Section 3, these results are not sufficient to apply to the models with delayed observations, we conclude this section with two results (cf. Theorem 2.9 and Proposition 2.11). These results will be applied to the Markov jump models with delayed counting observations in Section 4.

Assume that it is possible to construct copies $(\tilde{\boldsymbol{X}}^{\boldsymbol{n}}, \tilde{\boldsymbol{Y}}^{\boldsymbol{n}})$ and $(\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{Y}})$ of the pairs $(\boldsymbol{X}^{\boldsymbol{n}}, \boldsymbol{Y}^{\boldsymbol{n}})$ 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 $\tilde{\boldsymbol{Y}}^{\boldsymbol{n}} = \tilde{\boldsymbol{Y}}$, i.e.

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

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

Remark 2.1. *In particular, when the systems $(\boldsymbol{X}, \boldsymbol{Y})$ and $(\boldsymbol{X}^{\boldsymbol{n}}, \boldsymbol{Y}^{\boldsymbol{n}})$ are Markovian, condition $(\boldsymbol{a})$ means that under $\mathbb{P}$ the pair $(\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{Y}})$ has the generator $\boldsymbol{L}$ of $(\boldsymbol{X}, \boldsymbol{Y})$, and the initial condition $\mu = \mu_0^{X,Y}$, while condition $(\boldsymbol{a_n})$ means that under $\mathbb{P}^n$ the pair $(\tilde{\boldsymbol{X}}^{\boldsymbol{n}}, \tilde{\boldsymbol{Y}})$ has the generator $\boldsymbol{L}^{\boldsymbol{n}}$ of $(\boldsymbol{X}^{\boldsymbol{n}}, \boldsymbol{Y}^{\boldsymbol{n}})$, and the initial condition $\mu^n = \mu_0^{X^n, Y^n}$. Moreover if the systems are Markov with generators $\boldsymbol{L}$ and $\boldsymbol{L}^{\boldsymbol{n}}$ and initial conditions $\mu_0^{X,Y} = \delta_{\{x,y\}}$ and $\mu_0^{X^n,Y^n} = \delta_{\{x,y\}}$, then the semigroups can be represented respectively as*

$$\exp\{\boldsymbol{L}t\}\phi(x,y) = \mathbb{E}\big[\phi(\tilde{X}_t, \tilde{Y}_t)\big] \quad and \quad \exp\{\boldsymbol{L^n}t\}\phi(x,y) = \mathbb{E}^n\big[\phi(\tilde{X}_t^n, \tilde{Y}_t)\big].$$

As it will become clear in the application to the jump models in Section 4, 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

$(\boldsymbol{b1})$ $\mathbb{P}$ and $\mathbb{P}^n$ are absolutely continuous w.r.t. $\mathbb{Q}$ 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 w.r.t. $\mathbb{Q}$ on $\mathcal{F}_t^{\tilde{Y}}$. For technical reasons it is more convenient to assume the further condition

$(\boldsymbol{b2})$ 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, if the previous conditions $(\boldsymbol{a})$, $(\boldsymbol{a_n})$, and $(\boldsymbol{b2})$ hold, then

$$U(t, \tilde{\boldsymbol{Y}}; \varphi) = \mathbb{E}\big[\varphi(\tilde{X}_t)/\mathcal{F}_t^{\tilde{Y}}\big], \ \mathbb{P} - a.s. \quad (\text{and therefore also } \mathbb{Q} - a.s. \text{ and } \mathbb{P}^n - a.s.),$$

and analogously

$$U^n(t, \tilde{\boldsymbol{Y}}; \varphi) = \mathbb{E}^n\big[\varphi(\tilde{X}_t^n)/\mathcal{F}_t^{\tilde{Y}}\big], \ \mathbb{P}^n - a.s. \quad (\text{and therefore also } \mathbb{Q} - a.s. \text{ and } \mathbb{P} - a.s.).$$

Therefore

$$E\big[dist(\pi_t, \tilde{\pi}_t^n)\big] = E\big[\sup_{\varphi \in \mathcal{K}} |\pi_t(\varphi) - \tilde{\pi}_t^n(\varphi)|\big] \tag{23}$$

$$= E\big[\sup_{\varphi \in \mathcal{K}} |U(t, \boldsymbol{Y}; \varphi) - U^n(t, \boldsymbol{Y}; \varphi)|\big] \tag{24}$$

$$= \mathbb{E}\big[\sup_{\varphi \in \mathcal{K}} |U(t, \tilde{\boldsymbol{Y}}; \varphi) - U^n(t, \tilde{\boldsymbol{Y}}; \varphi)|\big]$$

$$= \mathbb{E}\big[\sup_{\varphi \in \mathcal{K}} |\mathbb{E}\big[\varphi(\tilde{X}_t)/\mathcal{F}_t^{\tilde{Y}}\big] - \mathbb{E}^n\big[\varphi(\tilde{X}_t^n)/\mathcal{F}_t^{\tilde{Y}}\big]|\big]$$

where we stress that, under $\mathbb{P}$, the law of $U^n(t, \tilde{\boldsymbol{Y}}; \varphi) = \mathbb{E}^n\big[\varphi(\tilde{X}_t^n)/\mathcal{F}_t^{\tilde{Y}}\big]$ 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. Analogously

$$E\big[dist(\pi_t^n, \overline{\pi}_t^n)\big] = E\big[\sup_{\varphi \in \mathcal{K}} |\pi_t^n(\varphi) - \overline{\pi}_t^n(\varphi)|\big] \tag{25}$$

$$= E\big[\sup_{\varphi \in \mathcal{K}} |U^n(t, \boldsymbol{Y^n}; \varphi) - U(t, \boldsymbol{Y^n}; \varphi)|\big]$$

$$= \mathbb{E}^n\big[\sup_{\varphi \in \mathcal{K}} |U^n(t, \tilde{\boldsymbol{Y}}; \varphi) - U(t, \tilde{\boldsymbol{Y}}; \varphi)|\big]$$

$$= \mathbb{E}^n\big[\sup_{\varphi \in \mathcal{K}} |\mathbb{E}^n\big[\varphi(\tilde{X}_t^n)/\mathcal{F}_t^{\tilde{Y}}\big] - \mathbb{E}\big[\varphi(\tilde{X}_t)/\mathcal{F}_t^{\tilde{Y}}\big]|\big].$$

So our aim is equivalent to prove the convergence to zero of either $\mathbb{E}\big[dist\big(U^n(t, \tilde{\boldsymbol{Y}}), U(t, \tilde{\boldsymbol{Y}})\big)\big]$, or $\mathbb{E}^n\big[dist\big(U^n(t, \tilde{\boldsymbol{Y}}), U(t, \tilde{\boldsymbol{Y}})\big)\big]$, depending on which metric and situation we are interested in. In any case it is natural to start by looking for $\mathbb{Q}$-a.s. upper bounds of

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

with $\mathcal{K}$ depending on the metric, and then to take the expectation with respect to either $\mathbb{E}$ or $\mathbb{E}^n$, depending on the situation we are interested in.

Furthermore, if condition **(b1)** holds, it can be shown (e.g. by Lemma 4.1 of [8]) that, for any bounded function φ,

$$|\mathbb{E}\big[\varphi(\tilde{X}_t)/\mathcal{F}_t^{\tilde{Y}}\big] - \mathbb{E}^n\big[\varphi(\tilde{X}_t^n)/\mathcal{F}_t^{\tilde{Y}}\big]| \tag{26}$$

is bounded above by either

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

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

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

$\mathbb{Q} - a.s.$, and therefore also $\mathbb{P} - a.s.$ and $\mathbb{P}^n - a.s..$ Again we use the first or the second upper bound, depending on which situation we are interested in.

Taking into account that

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

or that, if φ is also a Lipschitz function,

$$|\varphi(\tilde{X}_t) - \varphi(\tilde{X}_t^n)| \le L_\varphi \left| \tilde{X}_t - \tilde{X}_t^n \right|,$$

the previous observations can be used to get upper bounds for $E\big[dist(\pi_t, \tilde{\pi}_t^n)\big]$ and for $E\big[dist(\pi_t^n, \overline{\pi}_t^n)\big]$, when using the total variation or the bounded Lipschitz metric. Indeed, if one is interested in total variation one has the following upper bounds depending on the actually observed trajectories: either

$$\sup_{\varphi \in \mathcal{K}_{TV}} |\mathbb{E}\big[\varphi(\tilde{X}_t)/\mathcal{F}_t^{\tilde{Y}}\big] - \mathbb{E}^n\big[\varphi(\tilde{X}_t^n)/\mathcal{F}_t^{\tilde{Y}}\big]| \le 2\, \mathcal{Z}_t^n + 2\, \mathbb{P}^n\big(\{\tilde{X}_t \ne \tilde{X}_t^n\}/\mathcal{F}_t^{\tilde{Y}}\big),$$

or

$$\sup_{\varphi \in \mathcal{K}_{TV}} |\mathbb{E}\big[\varphi(\tilde{X}_t)/\mathcal{F}_t^{\tilde{Y}}\big] - \mathbb{E}^n\big[\varphi(\tilde{X}_t^n)/\mathcal{F}_t^{\tilde{Y}}\big]| \le 2\, \mathcal{Z}_t + 2\, \mathbb{P}\big(\{\tilde{X}_t \ne \tilde{X}_t^n\}/\mathcal{F}_t^{\tilde{Y}}\big),$$

where, as in the rest of this paper, the following notations are used

$$\mathcal{Z}_t^n = \frac{\mathbb{E}_{\mathbb{Q}}\big[|(d\mathbb{P}^n/d\mathbb{Q})|_{\tilde{\mathcal{F}}_t} - (d\mathbb{P}/d\mathbb{Q})|_{\tilde{\mathcal{F}}_t}|/\mathcal{F}_t^{\tilde{Y}}\big]}{\mathbb{E}_{\mathbb{Q}}\big[(d\mathbb{P}^n/d\mathbb{Q})|_{\tilde{\mathcal{F}}_t}/\mathcal{F}_t^{\tilde{Y}}\big]} \tag{27}$$

and

$$\mathcal{Z}_t = \frac{\mathbb{E}_{\mathbb{Q}}\big[|(d\mathbb{P}^n/d\mathbb{Q})|_{\tilde{\mathcal{F}}_t} - (d\mathbb{P}/d\mathbb{Q})|_{\tilde{\mathcal{F}}_t}|/\mathcal{F}_t^{\tilde{Y}}\big]}{\mathbb{E}_{\mathbb{Q}}\big[(d\mathbb{P}/d\mathbb{Q})|_{\tilde{\mathcal{F}}_t}/\mathcal{F}_t^{\tilde{Y}}\big]}. \tag{28}$$

If one is interested in bounded Lipschitz metric one has either

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

or

$$\sup_{\varphi \in \mathcal{K}_{BL}} |\mathbb{E}\big[\varphi(\tilde{X}_t)/\mathcal{F}_t^{\tilde{Y}}\big] - \mathbb{E}^n\big[\varphi(\tilde{X}_t^n)/\mathcal{F}_t^{\tilde{Y}}\big]| \le 2\, \mathcal{Z}_t + \mathbb{E}\big[|\tilde{X}_t - \tilde{X}_t^n|/\mathcal{F}_t^{\tilde{Y}}\big].$$

Remark 2.2. *Considering that*

$$|\varphi(\tilde{X}_t) - \varphi(\tilde{X}_t^n)| \le \max(2\|\varphi\|, L_\varphi)\left(|\tilde{X}_t - \tilde{X}_t^n| \wedge 1\right) \le 2\max(\|\varphi\|, L_\varphi)\left(|\tilde{X}_t - \tilde{X}_t^n| \wedge 1\right),$$

one can also have the bounds with either

$$\sup_{\varphi \in \mathcal{K}_{BL}} |\mathbb{E}\big[\varphi(\tilde{X}_t)/\mathcal{F}_t^{\tilde{Y}}\big] - \mathbb{E}^n\big[\varphi(\tilde{X}_t^n)/\mathcal{F}_t^{\tilde{Y}}\big]| \le 2\, \mathcal{Z}_t^n + 2\, \mathbb{E}^n\big[|\tilde{X}_t - \tilde{X}_t^n| \wedge 1/\mathcal{F}_t^{\tilde{Y}}\big],$$

or

$$\sup_{\varphi \in \mathcal{K}_{BL}} |\mathbb{E}\big[\varphi(\tilde{X}_t)/\mathcal{F}_t^{\tilde{Y}}\big] - \mathbb{E}^n\big[\varphi(\tilde{X}_t^n)/\mathcal{F}_t^{\tilde{Y}}\big]| \le 2\, \mathcal{Z}_t + 2\, \mathbb{E}\big[|\tilde{X}_t - \tilde{X}_t^n| \wedge 1/\mathcal{F}_t^{\tilde{Y}}\big].$$

Moreover, an easy computation gives

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

Indeed

$$\begin{aligned}
\mathbb{E}^n\big[\mathcal{Z}_t^n\big] &= \mathbb{E}_{\mathbb{Q}}\Big[(d\mathbb{P}^n/d\mathbb{Q})|_{\mathcal{F}_t^{\tilde{Y}}} \frac{\mathbb{E}_{\mathbb{Q}}\big[|(d\mathbb{P}^n/d\mathbb{Q})|_{\tilde{\mathcal{F}}_t} - (d\mathbb{P}/d\mathbb{Q})|_{\tilde{\mathcal{F}}_t}|/\mathcal{F}_t^{\tilde{Y}}\big]}{\mathbb{E}_{\mathbb{Q}}\big[(d\mathbb{P}^n/d\mathbb{Q})|_{\tilde{\mathcal{F}}_t}/\mathcal{F}_t^{\tilde{Y}}\big]}\Big] \\
&= \mathbb{E}_{\mathbb{Q}}\Big[\mathbb{I}_{\big\{(d\mathbb{P}^n/d\mathbb{Q})|_{\mathcal{F}_t^{\tilde{Y}}} \ne 0\big\}} \mathbb{E}_{\mathbb{Q}}\big[|(d\mathbb{P}^n/d\mathbb{Q})|_{\tilde{\mathcal{F}}_t} - (d\mathbb{P}/d\mathbb{Q})|_{\tilde{\mathcal{F}}_t}|/\mathcal{F}_t^{\tilde{Y}}\big]\Big] \\
&= \mathbb{E}_{\mathbb{Q}}\Big[|(d\mathbb{P}^n/d\mathbb{Q})|_{\tilde{\mathcal{F}}_t} - (d\mathbb{P}/d\mathbb{Q})|_{\tilde{\mathcal{F}}_t}|\Big]
\end{aligned} \tag{29}$$

where in the last equality we have used condition **(b2)**, and analogously we get

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

As a consequence of the above analysis, by taking (23) and (25) into account, we get the following upper bounds.

Theorem 2.3. *Under conditions* $(\boldsymbol{a})$, $(\boldsymbol{a_n})$, $(\boldsymbol{b1})$, *and* $(\boldsymbol{b2})$,

$$E\big[\|\pi_t - \tilde{\pi}_t^n\|_{TV}\big] \leq 2\mathbb{E}_{\mathbb{Q}}\Big[\big|(d\mathbb{P}^n/d\mathbb{Q})|_{\widetilde{\mathcal{F}}_t} - (d\mathbb{P}/d\mathbb{Q})|_{\widetilde{\mathcal{F}}_t}\big|\Big] + 2\mathbb{P}\big(\{\tilde{X}_t \neq \tilde{X}_t^n\}\big), \tag{30}$$

or

$$E\big[\|\pi_t^n - \overline{\pi}_t^n\|_{TV}\big] \leq 2\mathbb{E}_{\mathbb{Q}}\Big[\big|(d\mathbb{P}^n/d\mathbb{Q})|_{\widetilde{\mathcal{F}}_t} - (d\mathbb{P}/d\mathbb{Q})|_{\widetilde{\mathcal{F}}_t}\big|\Big] + 2\mathbb{P}^n\big(\{\tilde{X}_t \neq \tilde{X}_t^n\}\big), \tag{31}$$

and

$$E\big[d_{BL}(\pi_t, \tilde{\pi}_t^n)\big] \leq 2\mathbb{E}_{\mathbb{Q}}\Big[\big|(d\mathbb{P}^n/d\mathbb{Q})|_{\widetilde{\mathcal{F}}_t} - (d\mathbb{P}/d\mathbb{Q})|_{\widetilde{\mathcal{F}}_t}\big|\Big] + \mathbb{E}\big[\big|\tilde{X}_t - \tilde{X}_t^n\big|\big], \tag{32}$$

or

$$E\big[d_{BL}(\pi_t^n, \overline{\pi}_t^n)\big] \leq 2\mathbb{E}_{\mathbb{Q}}\Big[\big|(d\mathbb{P}^n/d\mathbb{Q})|_{\widetilde{\mathcal{F}}_t} - (d\mathbb{P}/d\mathbb{Q})|_{\widetilde{\mathcal{F}}_t}\big|\Big] + \mathbb{E}^n\big[\big|\tilde{X}_t - \tilde{X}_t^n\big|\big]. \tag{33}$$

Remark 2.4. *It is clear from the proof that we can rewrite the above inequalities (30)−(33) by substituting* $\mathbb{E}^n\big[\mathcal{Z}_t^n\big]$ *or* $\mathbb{E}\big[\mathcal{Z}_t\big]$ *to the first addend. Moreover when in addition* $\mathbb{P}$ *is absolutely continuous w.r.t.* $\mathbb{P}^n$ *on* $\tilde{\mathcal{F}}_t$ *we can substitute* $\mathbb{E}\Big[\big|\frac{d\mathbb{P}^n}{d\mathbb{P}}|_{\widetilde{\mathcal{F}}_t} - 1\big|\Big]$ *to the first addend, since*

$$\mathbb{E}^n\Big[\big|1 - \frac{d\mathbb{P}}{d\mathbb{P}^n}|_{\widetilde{\mathcal{F}}_t}\big|\Big] = \mathbb{E}_{\mathbb{Q}}\Big[\big|(d\mathbb{P}^n/d\mathbb{Q})|_{\widetilde{\mathcal{F}}_t} - (d\mathbb{P}/d\mathbb{Q})|_{\widetilde{\mathcal{F}}_t}\big|\Big].$$

An analogous substitution can be done when in addition $\mathbb{P}$ *is absolutely continuous w.r.t.* $\mathbb{P}^n$ *on* $\tilde{\mathcal{F}}_t$, *since*

$$\mathbb{E}_{\mathbb{Q}}\big[\big|(d\mathbb{P}^n/d\mathbb{Q})|_{\widetilde{\mathcal{F}}_t} - (d\mathbb{P}/d\mathbb{Q})|_{\widetilde{\mathcal{F}}_t}\big|\big] = \mathbb{E}\Big[\big|\frac{d\mathbb{P}^n}{d\mathbb{P}}|_{\widetilde{\mathcal{F}}_t} - 1\big|\Big].$$

Remark 2.5. *By taking into account the estimates obtained in Remark 2.2 one can substitute in (32) and (33) the expectations of* $\big|\tilde{X}_t - \tilde{X}_t^n\big|$ *with the corresponding expectations of* $2\big|\tilde{X}_t - \tilde{X}_t^n\big| \wedge 1$.

In the following theorem and remark we highlight the connections of the above considerations with the weak convergence of the sequence of filters π_t^n to π_t (see [1] and [2] for similar considerations).

Theorem 2.6. *Assume that*

1. $(\boldsymbol{a})$, $(\boldsymbol{a_n})$ *and* $(\boldsymbol{b2})$ *are satisfied,*

2. *the sequence* $U^n(t, \tilde{\boldsymbol{Y}})$ *converges in* $\mathbb{Q}$-*probability to* $U(t, \tilde{\boldsymbol{Y}})$, *as random variables with values in* $\mathcal{P}$, *the space of probability measures endowed with the metric dist, that is, for all* $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{Q}\big(A_\varepsilon^n\big) = 0, \quad \text{where } A_\varepsilon^n := \Big\{dist\big(U^n(t, \tilde{\boldsymbol{Y}}), U(t, \tilde{\boldsymbol{Y}})\big) \geq \varepsilon\Big\},$$

3. $\displaystyle\lim_{n\to\infty} \mathbb{E}_{\mathbb{Q}}\Big[\big|(d\mathbb{P}/d\mathbb{Q})|_{\mathcal{F}_t^{\tilde{Y}}} - (d\mathbb{P}^n/d\mathbb{Q})|_{\mathcal{F}_t^{\tilde{Y}}}\big|\Big] = 0.$

Then, the filters converge in distribution for each t as random variables with values in $\mathcal{P}$, *i.e.*

$$\pi_t^n \Rightarrow \pi_t.$$

Remark 2.7. *I)* *It is important to stress the connection with the convergence in probability of (17) and (18). Indeed conditions 2 and (b2) imply that $\mathbb{P}\big(A^n_\varepsilon\big)$ converge to zero, so that by condition (a)*

$$P\Big(dist\big(\tilde{\pi}^n_t;\pi_t\big)\Big) \geq \varepsilon\Big) = P\Big(dist\big(U^n(t,\boldsymbol{Y}),U(t,\boldsymbol{Y})\big) \geq \varepsilon\Big) = \mathbb{P}\big(A^n_\varepsilon\big) \to 0.$$

Similarly, by condition (aₙ),

$$P\Big(dist\big(\pi^n_t;\overline{\pi}^n_t\big)\Big) \geq \varepsilon\Big) = P\Big(dist\big(U^n(t,\boldsymbol{Y^n}),U(t,\boldsymbol{Y^n})\big) \geq \varepsilon\Big) = \mathbb{P}^n\big(A^n_\varepsilon\big),$$

and, by conditions 2, 3 and (b2), $\mathbb{P}^n\big(A^n_\varepsilon\big)$ converge to zero: the set $A^n_\varepsilon = \big\{dist\big(U^n(t,\tilde{\boldsymbol{Y}}),U(t,\tilde{\boldsymbol{Y}})\big) \geq \varepsilon\big\}$ is $\mathcal{F}^{\tilde{Y}}_t$−measurable, and by condition (b2)

$$\mathbb{P}^n(A^n_\varepsilon) = \mathbb{E}_{\mathbb{Q}}\Big[\,(d\mathbb{P}^n/d\mathbb{Q})|_{\mathcal{F}^{\tilde{Y}}_t}\,\mathbb{I}_{A^n_\varepsilon}\Big]$$

$$= \mathbb{E}_{\mathbb{Q}}\Big[\,(d\mathbb{P}/d\mathbb{Q})|_{\mathcal{F}^{\tilde{Y}}_t}\,\mathbb{I}_{A^n_\varepsilon}\Big] + \mathbb{E}_{\mathbb{Q}}\Big[\,\big[(d\mathbb{P}^n/d\mathbb{Q})|_{\mathcal{F}^{\tilde{Y}}_t} - (d\mathbb{P}/d\mathbb{Q})|_{\mathcal{F}^{\tilde{Y}}_t}\big]\,\mathbb{I}_{A^n_\varepsilon}\Big]$$

$$\leq \mathbb{E}_{\mathbb{Q}}\Big[\,(d\mathbb{P}/d\mathbb{Q})|_{\mathcal{F}^{\tilde{Y}}_t}\,\mathbb{I}_{A^n_\varepsilon}\Big] + \mathbb{E}_{\mathbb{Q}}\Big[\,\big|(d\mathbb{P}^n/d\mathbb{Q})|_{\mathcal{F}^{\tilde{Y}}_t} - (d\mathbb{P}/d\mathbb{Q})|_{\mathcal{F}^{\tilde{Y}}_t}\big|\Big].$$

The first addend in the last line, i.e. $\mathbb{E}_{\mathbb{Q}}\Big[\,(d\mathbb{P}/d\mathbb{Q})|_{\mathcal{F}^{\tilde{Y}}_t}\,\mathbb{I}_{A^n_\varepsilon}\Big] = \mathbb{P}(A^n_\varepsilon)$ converges to zero, by conditions 2 and (b2), while condition 3 asserts that the second addend converges to zero.

II) *Note that if condition (b1) also holds then condition 3 is implied by the stronger condition*

3' $\displaystyle \lim_{n\to\infty} \mathbb{E}_{\mathbb{Q}}\big[|\,(d\mathbb{P}/d\mathbb{Q})|_{\tilde{\mathcal{F}}_t} - (d\mathbb{P}^n/d\mathbb{Q})|_{\tilde{\mathcal{F}}_t}\,|\big] = 0,$

that is to say the convergence to zero of the first addend in all the upper bounds (30)−(33).

Therefore, if we can prove that the second addends in the upper bounds (30) or (32), depending on the chosen metric, also converge to zero, we get the convergence to zero of $\mathbb{P}\big(A^n_\varepsilon\big)$. As a consequence condition 2 is automatically satisfied and, provided condition 1 is enforced, we get the weak convergence of the filters. The convergence to zero in condition 3' can also be used to get the upper bounds for the second addends, as we do in Section 4 for the counting observation case.

An analogous result holds for the convergence to zero of $\mathbb{P}^n\big(A^n_\varepsilon\big)$, when assuming that the second addends in the upper bounds (31) or (33), depending on the chosen metric, also converge to zero, since one can prove that $\mathbb{P}\big(A^n_\varepsilon\big)$ converge to zero.

Proof. **(of Theorem 2.6)**
For any bounded function $G : \mathcal{P} \to \mathbb{R}$, continuous w.r.t. *dist*,

$$E[G(\pi^n_t)] = \mathbb{E}^n\big[G\big(U^n(t,\tilde{Y})\big)\big] = \mathbb{E}_{\mathbb{Q}}\big[\,(d\mathbb{P}^n/d\mathbb{Q})|_{\mathcal{F}^{\tilde{Y}}_t}\,G\big(U^n(t,\tilde{Y})\big)\big]$$

$$\to \mathbb{E}_{\mathbb{Q}}\big[\,(d\mathbb{P}/d\mathbb{Q})|_{\mathcal{F}^{\tilde{Y}}_t}\,G\big(U(t,\tilde{Y})\big)\big] = \mathbb{E}\big[G\big(U(t,\tilde{Y})\big)\big] = E[G(\pi_t)]$$

All the equalities in the above relations are obvious, we only need to prove the convergence result.
We can restrict to G uniformly continuous, with modulus of continuity $\boldsymbol{w}_G(\varepsilon)$

$$\Big|\mathbb{E}_{\mathbb{Q}}\big[\,(d\mathbb{P}^n/d\mathbb{Q})|_{\mathcal{F}^{\tilde{Y}}_t}\,G\big(U^n(t,\tilde{Y})\big)\big] - \mathbb{E}_{\mathbb{Q}}\big[\,(d\mathbb{P}/d\mathbb{Q})|_{\mathcal{F}^{\tilde{Y}}_t}\,G\big(U(t,\tilde{Y})\big)\big]\Big|$$

$$\leq \mathbb{E}_{\mathbb{Q}}\big[|\,(d\mathbb{P}^n/d\mathbb{Q})|_{\mathcal{F}^{\tilde{Y}}_t} - (d\mathbb{P}/d\mathbb{Q})|_{\mathcal{F}^{\tilde{Y}}_t}\,|\,|G\big(U^n(t,\tilde{Y})\big)|\big]$$

$$+ \mathbb{E}_{\mathbb{Q}}\big[\,(d\mathbb{P}/d\mathbb{Q})|_{\mathcal{F}^{\tilde{Y}}_t}\,|G\big(U^n(t,\tilde{Y})\big) - G\big(U(t,\tilde{Y})\big)|\big]$$

$$\leq \|G\|\,\mathbb{E}_{\mathbb{Q}}\big[|\,(d\mathbb{P}^n/d\mathbb{Q})|_{\mathcal{F}^{\tilde{Y}}_t} - (d\mathbb{P}/d\mathbb{Q})|_{\mathcal{F}^{\tilde{Y}}_t}\,|\big]$$

$$+ \mathbb{E}_{\mathbb{Q}}\big[\,(d\mathbb{P}/d\mathbb{Q})|_{\mathcal{F}^{\tilde{Y}}_t}\,\big(2\|G\|\,\mathbf{1}_{\{dist\big(U^n(t,\tilde{Y}),U(t,\tilde{Y})\big)\geq\varepsilon\}} + \boldsymbol{w}_G(\varepsilon)\big)\big]$$

11

The first addend converges to zero by condition **3**. The second addend is equal to

$$2\|G\|\mathbb{P}(A_\varepsilon^n) + \boldsymbol{w}_G(\varepsilon)$$

and converges to zero by conditions **2** and **(b2)**, and the arbitrariness of ε. $\qquad\square$

This result has been obtained by using the same kind of techniques as in [1]. In the latter paper the context is more general in the sense that the state space $\mathcal{S}$ is a complete separable metric space, and the filters converge weakly as measure valued processes in $D_{\mathcal{P}(\mathcal{S})}([0,T])$. However in [1], the space $\mathcal{P}(\mathcal{S})$ is always endowed with the Prohorov topology while in this paper we consider different kinds of metrics.

Remark 2.8. *If one is interested in weak convergence of the filters $\boldsymbol{\pi^n}$ to $\boldsymbol{\pi}$ as random processes in $\mathcal{D}_{\mathcal{P}}[0,T]$, then condition **2** can be replaced by*

2' $\displaystyle\lim_{n\to\infty} \mathbb{Q}\Big(\sup_{t\in[0,T]} dist\big(U^n(t,\tilde{\boldsymbol{Y}}), U(t,\tilde{\boldsymbol{Y}})\big) \geq \varepsilon \Big) = 0,$
*and condition **3** must be satisfied for $t = T$:*

$$\lim_{n\to\infty} \mathbb{E}_\mathbb{Q}\big[|\,(d\mathbb{P}/d\mathbb{Q})|_{\mathcal{F}_T^{\tilde{Y}}} - (d\mathbb{P}^n/d\mathbb{Q})|_{\mathcal{F}_T^{\tilde{Y}}}\,|\big] = 0.$$

*Indeed, if for any bounded function $G : \mathcal{D}_{\mathcal{P}}[0,T] \to \mathbb{R}$, continuous w.r.t. **Dist**, the Skorohod metric,*

$$E[G(\boldsymbol{\pi^n})] = \mathbb{E}^n\big[G\big(U^n(\cdot,\tilde{Y})\big)\big] = \mathbb{E}_\mathbb{Q}\big[\,(d\mathbb{P}^n/d\mathbb{Q})|_{\mathcal{F}_T^{\tilde{Y}}}\, G\big(U^n(\cdot,\tilde{Y})\big)\big]$$
$$\to \mathbb{E}_\mathbb{Q}\big[\,(d\mathbb{P}/d\mathbb{Q})|_{\mathcal{F}_T^{\tilde{Y}}}\, G\big(U(\cdot,\tilde{Y})\big)\big] = \mathbb{E}\big[G\big(U(\cdot,\tilde{Y})\big)\big] = E[G(\boldsymbol{\pi})]$$

All the equalities in the above relations are obvious, we only need to prove the convergence result. We can restrict to G uniformly continuous, with modulus of continuity $\boldsymbol{w}_G^{\mathcal{S}}(\varepsilon)$ w.r.t. Skorohod metric

$$\Big|\mathbb{E}_\mathbb{Q}\big[\,(d\mathbb{P}^n/d\mathbb{Q})|_{\mathcal{F}_T^{\tilde{Y}}}\, G\big(U^n(\cdot,\tilde{Y})\big)\big] - \mathbb{E}_\mathbb{Q}\big[\,(d\mathbb{P}/d\mathbb{Q})|_{\mathcal{F}_T^{\tilde{Y}}}\, G\big(U(\cdot,\tilde{Y})\big)\big]\Big|$$
$$\leq \mathbb{E}_\mathbb{Q}\big[|\,(d\mathbb{P}^n/d\mathbb{Q})|_{\mathcal{F}_T^{\tilde{Y}}} - (d\mathbb{P}/d\mathbb{Q})|_{\mathcal{F}_T^{\tilde{Y}}}\,|\,\big|G\big(U^n(\cdot,\tilde{Y})\big)|\big]$$
$$+ \mathbb{E}_\mathbb{Q}\big[\,(d\mathbb{P}/d\mathbb{Q})|_{\mathcal{F}_T^{\tilde{Y}}}\,\big|G\big(U^n(\cdot,\tilde{Y})\big) - G\big(U(\cdot,\tilde{Y})\big)|\big]$$
$$\leq \|G\|\,\mathbb{E}_\mathbb{Q}\big[|\,(d\mathbb{P}^n/d\mathbb{Q})|_{\mathcal{F}_T^{\tilde{Y}}} - (d\mathbb{P}/d\mathbb{Q})|_{\mathcal{F}_T^{\tilde{Y}}}\,|\big]$$
$$+ \mathbb{E}_\mathbb{Q}\big[\,(d\mathbb{P}/d\mathbb{Q})|_{\mathcal{F}_T^{\tilde{Y}}}\,\big(2\|G\|\,\mathbf{1}_{\{\boldsymbol{Dist}\big(U^n(\cdot,\tilde{Y}),U(\cdot,\tilde{Y})\big)\geq\varepsilon\}} + \boldsymbol{w}_G^{\mathcal{S}}(\varepsilon)\big)\big]$$

*The first addend converges to zero by condition **3** for $t = T$. The second addend is equal to*

$$2\|G\|\mathbb{P}(\boldsymbol{Dist}\big(U^n(\cdot,\tilde{Y}),U(\cdot,\tilde{Y})\big) \geq \varepsilon) + \boldsymbol{w}_G^{\mathcal{S}}(\varepsilon)$$
$$\leq 2\|G\|\mathbb{P}(\sup_{[0,T]} dist\big(U^n(t,\tilde{Y}),U(t,\tilde{Y})\big) \geq \varepsilon) + \boldsymbol{w}_G^{\mathcal{S}}(\varepsilon)$$

*and converges to zero by conditions **2'** and **(b2)**, and the arbitrariness of ε.*

2.3 Further upperbounds

In the next section we will use a slight modification of the above procedure, taking into account the delay structure of the model we are dealing with in order to prove our approximation results. Indeed for technical reasons we will need the following obvious observation

$$\sup_{\varphi\in\mathcal{K}} |\mathbb{E}\big[\varphi(\tilde{X}_t)/\mathcal{F}_t^{\tilde{Y}}\big] - \mathbb{E}^n\big[\varphi(\tilde{X}_t^n)/\mathcal{F}_t^{\tilde{Y}}\big]| \leq \sup_{\phi\in\mathcal{K}'} |\mathbb{E}\big[\phi(\tilde{X}_t,\tilde{Y}_t)/\mathcal{F}_t^{\tilde{Y}}\big] - \mathbb{E}^n\big[\phi(\tilde{X}_t^n,\tilde{Y}_t)/\mathcal{F}_t^{\tilde{Y}}\big]|$$

whenever $\mathcal{K}'$ contains all functions $\phi(x,y)$ such that $\phi(x,y) = \varphi(x)$, with $\varphi \in \mathcal{K}$. The r.h.s. of the last inequality being equal to

$$\sup_{\phi \in \mathcal{K}'} |U(t, \tilde{\boldsymbol{Y}}; \phi(\cdot, \tilde{Y}_t)) - U^n(t, \tilde{\boldsymbol{Y}}; \phi(\cdot, \tilde{Y}_t))|,$$

which coincides with

$$\sup_{\phi \in \mathcal{K}'} |\pi_t(\phi(\cdot, \tilde{Y}_t)) - \tilde{\pi}_t^n(\phi(\cdot, \tilde{Y}_t))|, \qquad \mathbb{P} - a.s.,$$

and with

$$\sup_{\phi \in \mathcal{K}'} |\bar{\pi}_t^n(\phi(\cdot, \tilde{Y}_t)) - \pi_t^n(\phi(\cdot, \tilde{Y}_t))|, \qquad \mathbb{P}^n - a.s.,$$

our aim will be accomplished once we prove the convergence to zero of the expectation of the later quantity for a suitable choice of $\mathcal{K}'$, the expectation being taken w.r.t. $\mathbb{P}$ or $\mathbb{P}^n$ depending on which situation we are interested in. With this aim, denote by $\mathcal{K}_{TV}(\alpha)$ the class of measurable functions $\varphi(x)$ bounded above by α, and $\mathcal{K}_{BL}(\alpha, \Lambda)$ the class of measurable functions $\varphi(x)$ bounded above by α, and such that for all x, x'

$$|\varphi(x) - \varphi(x')| \le \Lambda |x - x'|,$$

so that $\mathcal{K}_{TV} = \mathcal{K}_{TV}(1)$, and $\mathcal{K}_{BL} = \mathcal{K}_{BL}(1,1)$. $\mathcal{K}'_{TV}(\alpha)$ is the class of measurable functions $\phi(x,y)$ bounded above by α, and $\mathcal{K}'_{BL}(\alpha, \Lambda)$ is the class of measurable functions $\phi(x,y)$ bounded above by α, and such that for all x, x', and y

$$|\phi(x,y) - \phi(x',y)| \le \Lambda |x - x'|.$$

Note that $\mathcal{K}'_{TV}(\alpha)$ contains all functions $\phi(x,y) = \varphi(x)$, with $\varphi \in \mathcal{K}_{TV}(\alpha)$, and that analogously $\mathcal{K}'_{BL}(\alpha, \Lambda)$ contains all functions $\phi(x,y) = \varphi(x)$, with $\varphi \in \mathcal{K}_{BL}(\alpha, \Lambda)$.

Theorem 2.9. *Assume that conditions $(\boldsymbol{a})$, $(\boldsymbol{a_n})$, $(\boldsymbol{b1})$ and $(\boldsymbol{b2})$ are satisfied, then the following inequalities hold*

$$\mathbb{E}^n \Big[\sup_{\phi \in \mathcal{K}'_{TV}(\alpha)} |U(t, \tilde{\boldsymbol{Y}}; \phi(\cdot, \tilde{Y}_t)) - U^n(t, \tilde{\boldsymbol{Y}}; \phi(\cdot, \tilde{Y}_t))| \Big] \le 2\alpha \, \mathbb{E}^n \big[\mathcal{Z}_t^n \big] + 2\alpha \, \mathbb{P}^n \big(\{ \tilde{X}_t \ne \tilde{X}_t^n \} \big),$$

$$\mathbb{E} \Big[\sup_{\phi \in \mathcal{K}'_{TV}(\alpha)} |U(t, \tilde{\boldsymbol{Y}}; \phi(\cdot, \tilde{Y}_t)) - U^n(t, \tilde{\boldsymbol{Y}}; \phi(\cdot, \tilde{Y}_t))| \Big] \le 2\alpha \, \mathbb{E} \big[\mathcal{Z}_t \big] + 2\alpha \, \mathbb{P} \big(\{ \tilde{X}_t \ne \tilde{X}_t^n \} \big),$$

$$\mathbb{E}^n \Big[\sup_{\phi \in \mathcal{K}'_{BL}(\alpha, \Lambda)} |U(t, \tilde{\boldsymbol{Y}}; \phi(\cdot, \tilde{Y}_t)) - U^n(t, \tilde{\boldsymbol{Y}}; \phi(\cdot, \tilde{Y}_t))| \Big] \le 2\alpha \, \mathbb{E}^n \big[\mathcal{Z}_t^n \big] + \Lambda \, \mathbb{E}^n \big[|\tilde{X}_t - \tilde{X}_t^n| \big],$$

$$\mathbb{E} \Big[\sup_{\phi \in \mathcal{K}'_{BL}(\alpha, \Lambda)} |U(t, \tilde{\boldsymbol{Y}}; \phi(\cdot, \tilde{Y}_t)) - U^n(t, \tilde{\boldsymbol{Y}}; \phi(\cdot, \tilde{Y}_t))| \Big] \le 2\alpha \, \mathbb{E} \big[\mathcal{Z}_t \big] + \Lambda \, \mathbb{E} \big[|\tilde{X}_t - \tilde{X}_t^n| \big],$$

where $\mathcal{Z}_t^n$ and $\mathcal{Z}_t$ are defined in (27) and (28), respectively.

Proof. It can be shown (e.g. by Lemma 4.1 of [8]) that, for any bounded function ϕ,

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

and

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

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

13

Taking into account that, if $\phi \in \mathcal{K}'_{TV}(\alpha)$ we have

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

or that, if $\phi \in \mathcal{K}'_{BL}(\alpha; \Lambda)$, we have

$$|\phi(\tilde{X}_t, \tilde{Y}_t) - \phi(\tilde{X}_t^n, \tilde{Y}_t)| \leq \Lambda \left|\tilde{X}_t - \tilde{X}_t^n\right|,$$

the previous observations can be used to get upper bounds for $E\big[dist(\pi_t, \tilde{\pi}_t^n)\big]$ and for $E\big[dist(\pi_t^n, \overline{\pi}_t^n)\big]$, when using the total variation or the bounded Lipschitz metric. Indeed, if we are interested in total variation one has either

$$\sup_{\phi \in \mathcal{K}'_{TV}(\alpha)} |U(t, \tilde{\boldsymbol{Y}}; \phi(\cdot, \tilde{Y}_t)) - U^n(t, \tilde{\boldsymbol{Y}}; \phi(\cdot, \tilde{Y}_t))| \leq 2\alpha \, \mathcal{Z}_t^n + 2\alpha \, \mathbb{P}^n\big(\{\tilde{X}_t \neq \tilde{X}_t^n\}/\mathcal{F}_t^{\tilde{Y}}\big),$$

or

$$\sup_{\phi \in \mathcal{K}'_{TV}(\alpha)} |U(t, \tilde{\boldsymbol{Y}}; \phi(\cdot, \tilde{Y}_t)) - U^n(t, \tilde{\boldsymbol{Y}}; \phi(\cdot, \tilde{Y}_t))| \leq 2\alpha \, \mathcal{Z}_t + 2\alpha \, \mathbb{P}\big(\{\tilde{X}_t \neq \tilde{X}_t^n\}/\mathcal{F}_t^{\tilde{Y}}\big),$$

and, if we are interested in bounded Lipschitz metric one has either

$$\sup_{\phi \in \mathcal{K}'_{BL}(\alpha, \Lambda)} |U(t, \tilde{\boldsymbol{Y}}; \phi(\cdot, \tilde{Y}_t)) - U^n(t, \tilde{\boldsymbol{Y}}; \phi(\cdot, \tilde{Y}_t))| \leq 2\alpha \, \mathcal{Z}_t^n + \Lambda \, \mathbb{E}^n\big[\big|\tilde{X}_t - \tilde{X}_t^n\big|/\mathcal{F}_t^{\tilde{Y}}\big],$$

or

$$\sup_{\phi \in \mathcal{K}'_{BL}(\alpha, \Lambda)} |U(t, \tilde{\boldsymbol{Y}}; \phi(\cdot, \tilde{Y}_t)) - U^n(t, \tilde{\boldsymbol{Y}}; \phi(\cdot, \tilde{Y}_t))| \leq 2\alpha \, \mathcal{Z}_t + \Lambda \, \mathbb{E}\big[\big|\tilde{X}_t - \tilde{X}_t^n\big|/\mathcal{F}_t^{\tilde{Y}}\big].$$

Taking the expectation we get the stated inequalities. $\qquad \square$

Remark 2.10. *Noting that $\mathcal{K}'_{BL}(\alpha, \Lambda)$ is contained in $\mathcal{K}'_{TV}(\alpha)$ from the upper bounds in the last two inequalities in the above Theorem 2.9 we get*

$$\mathbb{E}^n\Big[\sup_{\phi \in \mathcal{K}'_{BL}(\alpha, \Lambda)} |U(t, \tilde{\boldsymbol{Y}}; \phi(\cdot, \tilde{Y}_t)) - U^n(t, \tilde{\boldsymbol{Y}}; \phi(\cdot, \tilde{Y}_t))|\Big]$$

$$\leq 2\alpha \, \mathbb{E}^n\big[\mathcal{Z}_t^n\big] + \min\Big(2\alpha \, \mathbb{P}^n\big(\{\tilde{X}_t \neq \tilde{X}_t^n\}\big); \Lambda \, \mathbb{E}^n\big[\big|\tilde{X}_t - \tilde{X}_t^n\big|\big]\Big),$$

$$\mathbb{E}\Big[\sup_{\phi \in \mathcal{K}'_{BL}(\alpha, \Lambda)} |U(t, \tilde{\boldsymbol{Y}}; \phi(\cdot, \tilde{Y}_t)) - U^n(t, \tilde{\boldsymbol{Y}}; \phi(\cdot, \tilde{Y}_t))|\Big]$$

$$\leq 2\alpha \, \mathbb{E}\big[\mathcal{Z}_t\big] + \min\Big(2\alpha \, \mathbb{P}\big(\{\tilde{X}_t \neq \tilde{X}_t^n\}\big); \Lambda \, \mathbb{E}\big[\big|\tilde{X}_t - \tilde{X}_t^n\big|\big]\Big).$$

Furthermore, observing that for any $\phi \in \mathcal{K}'_{BL}(\alpha, \Lambda)$,

$$|\phi(x, y) - \phi(x', y)| \leq \max(2\alpha; \Lambda)\big(|x - x'| \wedge 1\big),$$

by similar arguments as in the proof of the above Theorem 2.9 we get

$$\mathbb{E}^n\Big[\sup_{\phi \in \mathcal{K}'_{BL}(\alpha, \Lambda)} |U(t, \tilde{\boldsymbol{Y}}; \phi(\cdot, \tilde{Y}_t)) - U^n(t, \tilde{\boldsymbol{Y}}; \phi(\cdot, \tilde{Y}_t))|\Big]$$

$$\leq 2\alpha \, \mathbb{E}^n\big[\mathcal{Z}_t^n\big] + \max(2\alpha; \Lambda) \, \mathbb{E}^n\big[\big|\tilde{X}_t - \tilde{X}_t^n\big| \wedge 1\big],$$

$$\mathbb{E}\Big[\sup_{\phi \in \mathcal{K}'_{BL}(\alpha, \Lambda)} |U(t, \tilde{\boldsymbol{Y}}; \phi(\cdot, \tilde{Y}_t)) - U^n(t, \tilde{\boldsymbol{Y}}; \phi(\cdot, \tilde{Y}_t))|\Big]$$

$$\leq 2\alpha \, \mathbb{E}\big[\mathcal{Z}_t\big] + \max(2\alpha; \Lambda) \, \mathbb{E}\big[\big|\tilde{X}_t - \tilde{X}_t^n\big| \wedge 1\big].$$

14

Moreover in the next section we will also use the following result.

Proposition 2.11. *Assume that conditions* $(\boldsymbol{a})$, $(\boldsymbol{a_n})$, $(\boldsymbol{b1})$ *and* $(\boldsymbol{b2})$ *are satisfied, then the following inequalities hold*

$$\sup_{\phi \in \mathcal{K}'_{TV}(\alpha)} \left| \mathbb{E}\big[\phi(\tilde{X}_t, \tilde{Y}_t)\big] - \mathbb{E}^n\big[\phi(\tilde{X}^n_t, \tilde{Y}_t)\big] \right| \leq \alpha\, \mathbb{E}^n\big[\mathcal{Z}^n_t\big] + 2\alpha\, \mathbb{P}^n\big(\{\tilde{X}_t \neq \tilde{X}^n_t\}\big),$$

$$\sup_{\phi \in \mathcal{K}'_{TV}(\alpha)} \left| \mathbb{E}\big[\phi(\tilde{X}_t, \tilde{Y}_t)\big] - \mathbb{E}^n\big[\phi(\tilde{X}^n_t, \tilde{Y}_t)\big] \right| \leq \alpha\, \mathbb{E}\big[\mathcal{Z}_t\big] + 2\alpha\, \mathbb{P}\big(\{\tilde{X}_t \neq \tilde{X}^n_t\}\big),$$

$$\sup_{\phi \in \mathcal{K}'_{BL}(\alpha,\Lambda)} \left| \mathbb{E}\big[\phi(\tilde{X}_t, \tilde{Y}_t)\big] - \mathbb{E}^n\big[\phi(\tilde{X}^n_t, \tilde{Y}_t)\big] \right| \leq \alpha\, \mathbb{E}^n\big[\mathcal{Z}^n_t\big] + \Lambda\, \mathbb{E}^n\big[\big|\tilde{X}_t - \tilde{X}^n_t\big|\big],$$

$$\sup_{\phi \in \mathcal{K}'_{BL}(\alpha,\Lambda)} \left| \mathbb{E}\big[\phi(\tilde{X}_t, \tilde{Y}_t)\big] - \mathbb{E}^n\big[\phi(\tilde{X}^n_t, \tilde{Y}_t)\big] \right| \leq \alpha\, \mathbb{E}\big[\mathcal{Z}_t\big] + \Lambda\, \mathbb{E}\big[\big|\tilde{X}_t - \tilde{X}^n_t\big|\big],$$

where $\mathcal{Z}^n_t$ *and* $\mathcal{Z}_t$ *are defined in (27) and (28), respectively.*

Proof. For any $\phi \in \mathcal{K}'_{TV}(\alpha)$

$$\left| \mathbb{E}\big[\phi(\tilde{X}_t, \tilde{Y}_t)\big] - \mathbb{E}^n\big[\phi(\tilde{X}^n_t, \tilde{Y}_t)\big] \right|$$

$$\leq \left| \mathbb{E}\big[\phi(\tilde{X}_t, \tilde{Y}_t)\big] - \mathbb{E}\big[\phi(\tilde{X}^n_t, \tilde{Y}_t)\big] \right|$$

$$+ \left| \mathbb{E}\big[\phi(\tilde{X}^n_t, \tilde{Y}_t)\big] - \mathbb{E}_{\mathbb{Q}}\Big[(d\mathbb{P}^n/d\mathbb{Q})|_{\mathcal{F}^{\tilde{X}^n,\tilde{Y}}_t}\phi(\tilde{X}^n_t, \tilde{Y}_t)\Big] \right|$$

$$\leq \left| \mathbb{E}\big[\phi(\tilde{X}_t, \tilde{Y}_t)\big] - \mathbb{E}\big[\phi(\tilde{X}^n_t, \tilde{Y}_t)\big] \right| + \alpha \mathbb{E}_{\mathbb{Q}}\Big[\big|(d\mathbb{P}/d\mathbb{Q})|_{\mathcal{F}^{\tilde{X}^n,\tilde{Y}}_t} - (d\mathbb{P}^n/d\mathbb{Q})|_{\mathcal{F}^{\tilde{X}^n,\tilde{Y}}_t}\big|\Big].$$

Since, by Jensen inequality, $|E(Z/\mathcal{G})| = |E(E(Z/\mathcal{H})/\mathcal{G})| \leq E(|E(Z/\mathcal{H})|/\mathcal{G})$ for any integrable random variable Z and σ-algebras $\mathcal{G} \subseteq \mathcal{H} \subseteq \mathcal{F}$, we have

$$\mathbb{E}_{\mathbb{Q}}\Big[\big|(d\mathbb{P}/d\mathbb{Q})|_{\mathcal{F}^{\tilde{X}^n,\tilde{Y}}_t} - (d\mathbb{P}^n/d\mathbb{Q})|_{\mathcal{F}^{\tilde{X}^n,\tilde{Y}}_t}\big|\Big] \leq \mathbb{E}_{\mathbb{Q}}\Big[\big|(d\mathbb{P}/d\mathbb{Q})|_{\tilde{\mathcal{F}}_t} - (d\mathbb{P}^n/d\mathbb{Q})|_{\tilde{\mathcal{F}}_t}\big|\Big],$$

and therefore

$$\left| \mathbb{E}\big[\phi(\tilde{X}_t, \tilde{Y}_t)\big] - \mathbb{E}^n\big[\phi(\tilde{X}^n_t, \tilde{Y}_t)\big] \right| \leq \left| \mathbb{E}\big[\phi(\tilde{X}_t, \tilde{Y}_t)\big] - \mathbb{E}\big[\phi(\tilde{X}^n_t, \tilde{Y}_t)\big] \right| + \alpha \mathbb{E}_{\mathbb{Q}}\Big[\big|(d\mathbb{P}/d\mathbb{Q})|_{\tilde{\mathcal{F}}_t} - (d\mathbb{P}^n/d\mathbb{Q})|_{\tilde{\mathcal{F}}_t}\big|\Big].$$

As a consequence

$$\sup_{\phi \in \mathcal{K}'_{TV}(\alpha)} \left| \mathbb{E}\big[\phi(\tilde{X}_t, \tilde{Y}_t)\big] - \mathbb{E}^n\big[\phi(\tilde{X}^n_t, \tilde{Y}_t)\big] \right| \leq \alpha \mathbb{E}_{\mathbb{Q}}\Big[\big|(d\mathbb{P}/d\mathbb{Q})|_{\tilde{\mathcal{F}}_t} - (d\mathbb{P}^n/d\mathbb{Q})|_{\tilde{\mathcal{F}}_t}\big|\Big] + 2\,\alpha\, \mathbb{P}\big(\{\tilde{X}_t \neq \tilde{X}^n_t\}\big).$$

Interchanging the role of $\mathbb{P}$ and $\mathbb{P}^n$ we get

$$\sup_{\phi \in \mathcal{K}'_{TV}(\alpha)} \left| \mathbb{E}\big[\phi(\tilde{X}_t, \tilde{Y}_t)\big] - \mathbb{E}^n\big[\phi(\tilde{X}^n_t, \tilde{Y}_t)\big] \right| \leq \alpha \mathbb{E}_{\mathbb{Q}}\Big[\big|(d\mathbb{P}/d\mathbb{Q})|_{\tilde{\mathcal{F}}_t} - (d\mathbb{P}^n/d\mathbb{Q})|_{\tilde{\mathcal{F}}_t}\big|\Big] + 2\,\alpha\, \mathbb{P}^n\big(\{\tilde{X}_t \neq \tilde{X}^n_t\}\big).$$

By using similar arguments it is easy to get

$$\sup_{\phi \in \mathcal{K}'_{BL}(\alpha,\Lambda)} \left| \mathbb{E}\big[\phi(\tilde{X}_t, \tilde{Y}_t)\big] - \mathbb{E}^n\big[\phi(\tilde{X}^n_t, \tilde{Y}_t)\big] \right| \leq \alpha \mathbb{E}_{\mathbb{Q}}\Big[\big|(d\mathbb{P}/d\mathbb{Q})|_{\tilde{\mathcal{F}}_t} - (d\mathbb{P}^n/d\mathbb{Q})|_{\tilde{\mathcal{F}}_t}\big|\Big] + \Lambda\, \mathbb{E}\big[\big|\tilde{X}_t - \tilde{X}^n_t\big|\big]$$

and

$$\sup_{\phi \in \mathcal{K}'_{BL}(\alpha,\Lambda)} \left| \mathbb{E}\big[\phi(\tilde{X}_t, \tilde{Y}_t)\big] - \mathbb{E}^n\big[\phi(\tilde{X}^n_t, \tilde{Y}_t)\big] \right| \leq \alpha \mathbb{E}_{\mathbb{Q}}\Big[\big|(d\mathbb{P}/d\mathbb{Q})|_{\tilde{\mathcal{F}}_t} - (d\mathbb{P}^n/d\mathbb{Q})|_{\tilde{\mathcal{F}}_t}\big|\Big] + \Lambda\, \mathbb{E}^n\big[\big|\tilde{X}_t - \tilde{X}^n_t\big|\big].$$

$\square$

Remark 2.12. *As in the previous Remark 2.10 by similar arguments as in the proof of Theorem 2.9 we also get*

$$\sup_{\phi \in \mathcal{K}'_{BL}(\alpha, \Lambda)} \left| \mathbb{E}\big[\phi(\tilde{X}_t, \tilde{Y}_t)\big] - \mathbb{E}^n\big[\phi(\tilde{X}^n_t, \tilde{Y}_t)\big] \right|$$

$$\leq \alpha \mathbb{E}_{\mathbb{Q}}\Big[\big|(d\mathbb{P}/d\mathbb{Q})|_{\tilde{\mathcal{F}}_t} - (d\mathbb{P}^n/d\mathbb{Q})|_{\tilde{\mathcal{F}}_t}\big| \Big] + \max(2\alpha; \Lambda) \min\Big(\mathbb{E}^n\big[|\tilde{X}_t - \tilde{X}^n_t| \wedge 1\big];\ \mathbb{E}\big[|\tilde{X}_t - \tilde{X}^n_t| \wedge 1\big] \Big)$$

3 Approximation in filtering for Markov models with delayed observations

In this section we show how to use the approximations $\tilde{\pi}^n_t$ and $\overline{\pi}^n_t$ introduced in Section 2 for the filter in the case of Markov models with delayed observations, in order to obtain upper bounds for the expectations in (19) and (20) when dealing with the bounded Lipschitz distance or the total variation distance. In this purpose, we give some general upper bounds, depending on the actually observed trajectories (see (35) and (38) below) which lead to the main results of this section stated in Proposition 3.1 and Proposition 3.2. In the next section we use these bounds and we exploit the results about the sensitivity of the filter for Markov jump processes with counting observations given in [6], [7], and [8]. As we have already noticed at the end of Section 1, some approximation results hold also for different systems and can be obtained with different techniques.

Consider the systems $(\boldsymbol{X^n}, \boldsymbol{Y^n})$ and $(\boldsymbol{X}, \boldsymbol{Y})$, such that

$$(\boldsymbol{X^n}, \hat{\boldsymbol{Y}}^{\boldsymbol{n}}) \text{ is a Markov system with generator } \boldsymbol{L^n},$$

and

$$Y^n_t = \hat{Y}^n_{a(t)},$$

and analogously

$$(\boldsymbol{X}, \hat{\boldsymbol{Y}}) \text{ is a Markov system with generator } \boldsymbol{L},$$

and

$$Y_t = \hat{Y}_{a(t)},$$

where $a(\cdot)$ has the same properties as in Theorem 1.1.

As explained in Theorem 1.1 the filters π^n_t and π_t can be expressed in terms of the filters $\hat{\pi}^n_t$ and $\hat{\pi}_t$ and the semigroups $\exp\{\boldsymbol{L^n}s\}$ and $\exp\{\boldsymbol{L}s\}$ of the corresponding Markovian systems. More precisely (see (11))

$$\pi^n_t(\varphi) = \hat{\pi}^n_{a(t)}\big(\exp\{\boldsymbol{L^n}(t - a(t))\}\phi(\cdot, \hat{Y}^n_{a(t)})\big)$$

and

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

where $\phi(x, y) = \varphi(x)$, as in Theorem 1.1 and Section 2.3.

Therefore, on the one side, the convergence of the filters may be investigated in terms of the convergence of $\hat{\pi}^n_t$ to $\hat{\pi}_t$ and of $\exp\{\boldsymbol{L^n}s\}$ to $\exp\{\boldsymbol{L}s\}$. On the other side, we are interested in the approximations (15) and (16) of the filter, and therefore we need a representation formula for the functionals U^n and U. 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

$$\hat{\pi}^n_t = \hat{U}^n\big(t, \hat{\boldsymbol{Y}}^{\boldsymbol{n}}\big) \text{ and } \hat{\pi}_t = \hat{U}\big(t, \hat{\boldsymbol{Y}}\big),$$

where the functionals $\hat{U}^n$ and $\hat{U}$ depend on the initial distributions and the generators of the corresponding Markov systems.

As a consequence

$$\pi^n_t(\varphi) = \int_{\mathbb{R}^k} \hat{U}^n\big(a(t), \boldsymbol{Y^n} \circ \boldsymbol{\mathcal{A}^{-1}}; dx\big) \exp\{\boldsymbol{L^n}(t - a(t))\}\phi(x, Y^n_t)$$

and
$$\pi_t(\varphi) = \int_{\mathbb{R}^k} \hat{U}\big(a(t), \boldsymbol{Y} \circ \boldsymbol{\mathcal{A}^{-1}}; dx\big) \exp\{\boldsymbol{L}(t - a(t))\}\phi(x, Y_t),$$

where again $\phi(x, y) = \varphi(x)$ and hence, the right hand sides of the previous equalities define implicitly the functionals U^n and U, respectively.

Recalling that

$$(Y \circ \mathcal{A}^{-1})_s = Y_{a^{-1}(s)} = \hat{Y}_{a(a^{-1}(s))} = \hat{Y}_s, \tag{34}$$

since $a(a^{-1}(s)) = s$, $Y_t = \hat{Y}_{a(t)}$ and $Y_t^n = \hat{Y}_{a(t)}^n$, we can rewrite the above formulas as

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

and

$$\pi_t(\varphi) = \int_{\mathbb{R}^k} \hat{U}\big(a(t), \hat{\boldsymbol{Y}}; dx\big) \exp\{\boldsymbol{L}(t - a(t))\}\phi(x, \hat{Y}_{a(t)}),$$
$$= \hat{U}\big(r, \boldsymbol{y}; \exp\{\boldsymbol{L}(t - r)\}\phi(\cdot, y_r)\big)\big|_{r=a(t), \boldsymbol{y}=\hat{\boldsymbol{Y}}}.$$

Therefore, when we observe $\boldsymbol{Y}$ (situation $\boldsymbol{1}$ of Section 2), either the total variation distance or the bounded-Lipschitz distance between π_t and $\tilde{\pi}_t^n$ can be evaluated in terms of the supremum of

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

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

over φ in the class of functions $\mathcal{K}_{TV}$ or $\mathcal{K}_{BL}$ respectively. When instead we observe $\boldsymbol{Y^n}$ (situation $\boldsymbol{2}$ of Section 2), the distance between π_t^n and $\bar{\pi}_t^n$ can be evaluated in terms of the supremum of

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

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

over φ in the class of functions $\mathcal{K}_{TV}$ or $\mathcal{K}_{BL}$ respectively.

In situation $\boldsymbol{1}$ we have

$$|\pi_t(\varphi) - \tilde{\pi}_t^n(\varphi)| \tag{35}$$
$$\leq \left|\hat{U}\big(r, \boldsymbol{y}; \exp\{\boldsymbol{L}(t - r)\}\phi(\cdot, y_r)\big) - \hat{U}^n\big(r, \boldsymbol{y}; \exp\{\boldsymbol{L}(t - r)\}\phi(\cdot, y_r)\big)\right|_{r=a(t), \boldsymbol{y}=\hat{\boldsymbol{Y}}} \tag{36}$$
$$+ \left|\hat{U}^n\big(r, \boldsymbol{y}; \exp\{\boldsymbol{L}(t - r)\}\phi(\cdot, y_r)\big) - \hat{U}^n\big(r, \boldsymbol{y}; \exp\{\boldsymbol{L^n}(t - r)\}\phi(\cdot, y_r)\big)\right|_{r=a(t), \boldsymbol{y}=\hat{\boldsymbol{Y}}} \tag{37}$$

and in situation $\boldsymbol{2}$,

$$|\pi_t^n(\varphi) - \bar{\pi}_t^n(\varphi)| \tag{38}$$
$$\leq \left|\hat{U}^n\big(r, \boldsymbol{y}; \exp\{\boldsymbol{L^n}(t - r)\}\phi(\cdot, y_r)\big) - \hat{U}^n\big(r, \boldsymbol{y}; \exp\{\boldsymbol{L}(t - r)\}\phi(\cdot, y_r)\big)\right|_{r=a(t), \boldsymbol{y}=\hat{\boldsymbol{Y}}^{\boldsymbol{n}}} \tag{39}$$
$$+ \left|\hat{U}^n\big(r, \boldsymbol{y}; \exp\{\boldsymbol{L}(t - r)\}\phi(\cdot, y_r)\big) - \hat{U}\big(r, \boldsymbol{y}; \exp\{\boldsymbol{L}(t - r)\}\phi(\cdot, y_r)\big)\right|_{r=a(t), \boldsymbol{y}=\hat{\boldsymbol{Y}}^{\boldsymbol{n}}}. \tag{40}$$

Note that the two previous upper bounds differ only from the fact that they are evaluated on $\boldsymbol{y} = \hat{\boldsymbol{Y}}$ and $\boldsymbol{y} = \hat{\boldsymbol{Y}}^{\boldsymbol{n}}$. Moreover natural upper bounds for (37) are

$$\sup_x |\exp\{\boldsymbol{L}(t-r)\}\phi(x, y_r) - \exp\{\boldsymbol{L^n}(t-r)\}\phi(x, y_r)|_{r=a(t), \boldsymbol{y}=\hat{\boldsymbol{Y}}} \tag{41}$$

$$\leq \sup_{x,y} |\exp\{\boldsymbol{L}(t-r)\}\phi(x, y) - \exp\{\boldsymbol{L^n}(t-r)\}\phi(x, y)|_{r=a(t)}. \tag{42}$$

The same being valid for (39) when substituting $\boldsymbol{y} = \hat{\boldsymbol{Y}}$ with $\boldsymbol{y} = \hat{\boldsymbol{Y}}^{\boldsymbol{n}}$.

Then in both situations the distance between the filter and the approximation can be expressed by means of the distance between $\hat{U}^n$ and $\hat{U}$ and the distance between the semigroups $\exp\{\boldsymbol{L^n}s\}$ and $\exp\{\boldsymbol{L}s\}$ as can be deduced by the following results.

Proposition 3.1. *For the system with delayed observations described at the beginning of this section we have*

$$\sup_{\varphi \in \mathcal{K}_{TV}(\alpha)} |\pi_t(\varphi) - \tilde{\pi}_t^n(\varphi)|$$

$$\leq \sup_{\psi \in \mathcal{K}'_{TV}(\alpha)} \left| \hat{U}\big(r, \boldsymbol{y}; \psi(\cdot, y_r)\big) - \hat{U}^n\big(r, \boldsymbol{y}; \psi(\cdot, y_r)\big) \right|_{r=a(t), \boldsymbol{y}=\hat{\boldsymbol{Y}}} \tag{43}$$

$$+ \sup_{\phi \in \mathcal{K}'_{TV}(\alpha)} \sup_{x,y} |\exp\{\boldsymbol{L}(t-r)\}\phi(x, y) - \exp\{\boldsymbol{L^n}(t-r)\}\phi(x, y)|_{r=a(t)}, \tag{44}$$

where $\mathcal{K}_{TV}(\alpha)$ and $\mathcal{K}'_{TV}(\alpha)$ are defined as in Section 2.3.

Furthermore, if for all $u \geq 0$ there exists a constant $\Lambda'(u, \Lambda)$ such that

$$\exp\{\boldsymbol{L}u\}\big(\mathcal{K}'_{BL}(\alpha, \Lambda)\big) \subseteq \mathcal{K}'_{BL}(\alpha, \Lambda'(u, \Lambda)), \tag{45}$$

then we have

$$\sup_{\varphi \in \mathcal{K}_{BL}(\alpha, \Lambda)} |\pi_t(\varphi) - \tilde{\pi}_t^n(\varphi)|$$

$$\leq \sup_{\psi \in \mathcal{K}'_{BL}(\alpha, \Lambda'(t-r, \Lambda))} \left| \hat{U}\big(r, \boldsymbol{y}; \psi(\cdot, y_r)\big) - \hat{U}^n\big(r, \boldsymbol{y}; \psi(\cdot, y_r)\big) \right|_{r=a(t), \boldsymbol{y}=\hat{\boldsymbol{Y}}} \tag{46}$$

$$+ \sup_{\phi \in \mathcal{K}'_{BL}(\alpha, \Lambda)} \sup_{x,y} |\exp\{\boldsymbol{L}(t-r)\}\phi(x, y) - \exp\{\boldsymbol{L^n}(t-r)\}\phi(x, y)|_{r=a(t)}, \tag{47}$$

where $\mathcal{K}_{BL}(\alpha, \Lambda)$ and $\mathcal{K}'_{BL}(\alpha, \Lambda)$ are defined as in Section 2.3.

Proof. Taking into account (35) to (37) we get

$$\sup_{\varphi \in \mathcal{K}_{TV}(\alpha)} |\pi_t(\varphi) - \tilde{\pi}_t^n(\varphi)|$$

$$\leq \sup_{\phi \in \mathcal{K}'_{TV}(\alpha)} \left| \hat{U}\big(r, \boldsymbol{y}; \exp\{\boldsymbol{L}(t-r)\}\phi(\cdot, y_r)\big) - \hat{U}^n\big(r, \boldsymbol{y}; \exp\{\boldsymbol{L}(t-r)\}\phi(\cdot, y_r)\big) \right|_{r=a(t), \boldsymbol{y}=\hat{\boldsymbol{Y}}}$$

$$+ \sup_{\phi \in \mathcal{K}'_{TV}(\alpha)} \sup_{x,y} |\exp\{\boldsymbol{L}(t-r)\}\phi(x, y) - \exp\{\boldsymbol{L^n}(t-r)\}\phi(x, y)|_{r=a(t)}$$

$$\leq \sup_{\psi \in \mathcal{K}'_{TV}(\alpha)} \left| \hat{U}\big(r, \boldsymbol{y}; \psi(\cdot, y_r)\big) - \hat{U}^n\big(r, \boldsymbol{y}; \psi(\cdot, y_r)\big) \right|_{r=a(t), \boldsymbol{y}=\hat{\boldsymbol{Y}}}$$

$$+ \sup_{\phi \in \mathcal{K}'_{TV}(\alpha)} \sup_{x,y} |\exp\{\boldsymbol{L}(t-r)\}\phi(x, y) - \exp\{\boldsymbol{L^n}(t-r)\}\phi(x, y)|_{r=a(t)}$$

Moreover we also have

$$\sup_{\varphi \in \mathcal{K}_{BL}(\alpha,\Lambda)} |\pi_t(\varphi) - \tilde{\pi}_t^n(\varphi)|$$

$$\leq \sup_{\phi \in \mathcal{K}'_{BL}(\alpha,\Lambda)} \left| \hat{U}\big(r, \boldsymbol{y}; \exp\{\boldsymbol{L}(t-r)\}\phi(\cdot, y_r)\big) - \hat{U}^n\big(r, \boldsymbol{y}; \exp\{\boldsymbol{L}(t-r)\}\phi(\cdot, y_r)\big) \right|_{r=a(t), \boldsymbol{y}=\hat{\boldsymbol{Y}}}$$

$$+ \sup_{\phi \in \mathcal{K}'_{BL}(\alpha,\Lambda)} \sup_{x,y} |\exp\{\boldsymbol{L}(t-r)\}\phi(x,y) - \exp\{\boldsymbol{L^n}(t-r)\}\phi(x,y)|_{r=a(t)}$$

$$\leq \sup_{\psi \in \exp\{\boldsymbol{L}(t-r)\}\mathcal{K}'_{BL}(\alpha,\Lambda)} \left| \hat{U}\big(r, \boldsymbol{y}; \psi(\cdot, y_r)\big) - \hat{U}^n\big(r, \boldsymbol{y}; \psi(\cdot, y_r)\big) \right|_{r=a(t), \boldsymbol{y}=\hat{\boldsymbol{Y}}}$$

$$+ \sup_{\phi \in \mathcal{K}'_{BL}(\alpha,\Lambda)} \sup_{x,y} |\exp\{\boldsymbol{L}(t-r)\}\phi(x,y) - \exp\{\boldsymbol{L^n}(t-r)\}\phi(x,y)|_{r=a(t)},$$

and invoking (45) we get (46) by noticing that the supremum over $\exp\{\boldsymbol{L}(t-r)\}\big(\mathcal{K}'_{BL}(\alpha,\Lambda)\big)$ in the bound given above can be replaced by the supremum over $\mathcal{K}'_{BL}(\alpha, \Lambda'(t-r,\Lambda))$.

<div style="text-align:right">□</div>

Taking into account (38) to (40) and changing $\hat{\boldsymbol{Y}}$ into $\hat{\boldsymbol{Y}^n}$ we have the following result.

Proposition 3.2. *For the system with delayed observations described at the beginning of this section we have*

$$\sup_{\varphi \in \mathcal{K}_{TV}(\alpha)} |\pi_t^n(\varphi) - \overline{\pi}_t^n(\varphi)|$$

$$\leq \sup_{\psi \in \mathcal{K}'_{TV}(\alpha)} \left| \hat{U}\big(r, \boldsymbol{y}; \psi(\cdot, y_r)\big) - \hat{U}^n\big(r, \boldsymbol{y}; \psi(\cdot, y_r)\big) \right|_{r=a(t), \boldsymbol{y}=\hat{\boldsymbol{Y}^n}} \tag{48}$$

$$+ \sup_{\phi \in \mathcal{K}'_{TV}(\alpha)} \sup_{x,y} |\exp\{\boldsymbol{L}(t-r)\}\phi(x,y) - \exp\{\boldsymbol{L^n}(t-r)\}\phi(x,y)|_{r=a(t)}. \tag{49}$$

Furthermore if condition (45) holds, then we have

$$\sup_{\varphi \in \mathcal{K}_{BL}(\alpha,\Lambda)} |\pi_t^n(\varphi) - \overline{\pi}_t^n(\varphi)|$$

$$\leq \sup_{\psi \in \mathcal{K}'_{BL}(\alpha,\Lambda'(t-r,\Lambda))} \left| \hat{U}\big(r, \boldsymbol{y}; \psi(\cdot, y_r)\big) - \hat{U}^n\big(r, \boldsymbol{y}; \psi(\cdot, y_r)\big) \right|_{r=a(t), \boldsymbol{y}=\hat{\boldsymbol{Y}^n}} \tag{50}$$

$$+ \sup_{\phi \in \mathcal{K}'_{BL}(\alpha,\Lambda)} \sup_{x,y} |\exp\{\boldsymbol{L}(t-r)\}\phi(x,y) - \exp\{\boldsymbol{L^n}(t-r)\}\phi(x,y)|_{r=a(t)}. \tag{51}$$

The above upper bounds will be evaluated when studying filtering systems with counting observations in the following section. In particular upper bounds for the expectations of (43) and (46) will be obtained by a slight modification of the arguments used in the previous section (see Theorem 2.9). The same kind of consideration holds for the expectations of (44) and (47) (see Proposition 2.11).

4 Counting processes observations

4.1 Main results

In this section, we accomplish the above exposed general program in the case of a Markov jump process with counting delayed observations. More precisely we assume that the pairs $(\boldsymbol{X}, \hat{\boldsymbol{Y}})$ and $(\boldsymbol{X^n}, \hat{\boldsymbol{Y}^n})$ are Markov systems in $\mathbb{R} \times \mathbb{N}$, with respective initial conditions $\mu_0^X \otimes \delta_{\{y\}}$ 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

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

so that the predictable intensities of $\hat{\boldsymbol{Y}}$ and $\hat{\boldsymbol{Y}^n}$ are $\lambda_1(X_{t-},\hat{Y}_{t-})$ and $\lambda_1^n(X_{t-}^n,\hat{Y}_{t-}^n)$ respectively, and such that the following hypotheses are satisfied

($\boldsymbol{H_0}$) $0 \leq \underline{\lambda}_i \leq \lambda_i(x,y), \lambda_i^n(x,y) \leq \overline{\lambda}_i, \quad$ for $i = 0,1$;

($\boldsymbol{H_1}$) $0 < \lambda_1(x,y), \lambda_1^n(x,y)$;

($\boldsymbol{H_2}$) $\Delta_i^n := \sup\limits_{x,y}\boldsymbol{\kappa}\big(\mu_i(x,y;\cdot),\mu_i^n(x,y;\cdot)\big) < \infty, \quad$ for $i = 0,1$.

Note that the operator $\boldsymbol{L}$ given above can be expressed in the same form as the one defined by (4) and viceversa.

From now on we will say that the operator $\boldsymbol{L}$ is *bL-regular* when it satisfies the following conditions

($\boldsymbol{R_0}$) for every z the function $x \mapsto \lambda_i(x,z)$, $i = 0,1$, is bounded Lipschitz continuous and the Lipschitz constant is bounded from above by L_{λ_i},

($\boldsymbol{R_1}$) $\sup\limits_{y}\boldsymbol{\kappa}\big(\mu_i(x,y;\cdot),\mu_i(x',y;\cdot)\big) \leq \Gamma_{\mu_i}|x - x'|, \ $ for $i = 0,1$

($\boldsymbol{R_2}$) $\sup\limits_{x,y}\int |z - x|\mu_0(x,y;dz) \leq b_0, \qquad \int |z - x|\mu_1(x,y;dz) \leq a_1(|x| + y) + b_1$.

It is important to stress that, if $\boldsymbol{L}$ is bL-regular, then $\exp\{\boldsymbol{L}u\}\big(\mathcal{K}_{BL}'(\alpha,\Lambda)\big) \subseteq \mathcal{K}_{BL}'(\alpha,\Lambda e^{Mu})$ (see Corollary 4.9) for a suitable constant M (given in (67)), so that condition (45) is satisfied with

$$\Lambda'(u,\Lambda) = \Lambda e^{Mu}. \tag{52}$$

Finally we will denote by $\|\cdot\|_\infty$ the norm defined by $\|\mu\|_\infty = \sup_{x,y}\|\mu(x,y,\cdot)\|_{TV}$ for any regular nucleus $\mu(x,y;dz)$.

The central tool for the proofs of our bounds (see Theorem 4.5 below) is a particular construction of the pairs $(\boldsymbol{X^n},\hat{\boldsymbol{Y}^n})$ 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 2, and such that

($\hat{\boldsymbol{a}}$) under $\mathbb{P}$ the pair $(\boldsymbol{X},\hat{\boldsymbol{Y}})$ has generator $\boldsymbol{L}$ and initial condition $\mu = \mu_0^X \otimes \delta_{\{y\}}$,

($\hat{\boldsymbol{a}}_{\boldsymbol{n}}$) under $\mathbb{P}^n$ the pair $(\boldsymbol{X^n},\hat{\boldsymbol{Y}})$ has generator $\boldsymbol{L}^n$ and initial condition $\mu^n = \mu_0^{X^n} \otimes \delta_{\{y\}}$,

($\hat{\boldsymbol{b}}$) the probability measures $\mathbb{Q}$, $\mathbb{P}$ and $\mathbb{P}^n$ are equivalent on $= \mathcal{F}_t^{X,X^n,\hat{Y}}$.

Note that the use of $(\boldsymbol{X^n},\hat{\boldsymbol{Y}^n})$ and $(\boldsymbol{X},\hat{\boldsymbol{Y}})$ is a slight abuse of notation: as at the end of the previous section we should write instead $(\tilde{\boldsymbol{X}}^n,\tilde{\hat{\boldsymbol{Y}}}^n)$ and $(\tilde{\boldsymbol{X}},\tilde{\hat{\boldsymbol{Y}}})$. Nevertheless we prefer the slight abuse of notation, in order to avoid the notation $\tilde{\hat{Y}}$.

Therefore, when dealing with the triplet $(\boldsymbol{X},\boldsymbol{X^n},\hat{\boldsymbol{Y}})$ instead of $(\tilde{\boldsymbol{X}},\tilde{\boldsymbol{X}}^n,\tilde{\boldsymbol{Y}})$, conditions **($\boldsymbol{a}$)**, **($\boldsymbol{a_n}$)**, **($\boldsymbol{b1}$)** and **($\boldsymbol{b2}$)** are satisfied. Then, in this framework, Theorem 2.9 applies with $(\boldsymbol{X},\boldsymbol{X^n},\hat{\boldsymbol{Y}})$ instead of $(\tilde{\boldsymbol{X}},\tilde{\boldsymbol{X}}^n,\tilde{\boldsymbol{Y}})$ and with $\hat{U}$ and $\hat{U}^n$ instead of U and U^n, respectively, the same considerations being valid for Proposition 2.11. The explicit expression of $\hat{U}$ is given by

$$\hat{U}(t,\boldsymbol{y};\varphi) = \begin{cases} \hat{\Pi}_t(\varphi|\boldsymbol{s}), & \text{for } y(t) = \sum_{i=1}^{\infty}\mathbb{I}_{[0,s_i]}(t), \\ \varphi(x_0), & \text{otherwise.} \end{cases}$$

(see (9) and (10)). Obviously the definition of the probability measure $\hat{U}(t, \boldsymbol{y}; dx)$ on the complement of the set $\{\boldsymbol{y}$ such that $y(t) = \sum_{i=1}^{\infty} \mathbb{I}_{[0,s_i]}(t)\}$ is not unique. An analogous representation holds for $\hat{U}^n$.

Basically the above conditions coincide with **Condition A** of [8], and the construction we use is the same as the one in Section 3 of [8], the only difference being that in [8] we assume $\hat{Y}_0 = 0$. By this construction the model $(\boldsymbol{X}, \boldsymbol{Y})$ is defined on $(\Omega, \mathcal{F}, \mathbb{P})$ taking $Y_t = \hat{Y}_{a(t)}$ and the model $(\boldsymbol{X}^n, \boldsymbol{Y}^n)$ is defined on $(\Omega, \mathcal{F}, \mathbb{P}^n)$ taking $\hat{\boldsymbol{Y}}^{\boldsymbol{n}} = \hat{\boldsymbol{Y}}$ and $Y_t^n = \hat{Y}_{a(t)}^n$, so that $(\boldsymbol{X}^n, \boldsymbol{Y}^n)$ coincides with $(\boldsymbol{X}^n, \boldsymbol{Y})$ on $(\Omega, \mathcal{F}, \mathbb{P}^n)$. Moreover if $\mu_0^X = \delta_{\{x\}}$ and $\mu_0^{X^n} = \delta_{\{x\}}$, then the semigroups can be represented respectively as

$$\exp\{\boldsymbol{L}t\}\phi(x,y) = \mathbb{E}\big[\phi(X_t, \hat{Y}_t)\big], \tag{53}$$

and

$$\exp\{\boldsymbol{L}^{\boldsymbol{n}}t\}\phi(x,y) = \mathbb{E}^n\big[\phi(X_t^n, \hat{Y}_t)\big]. \tag{54}$$

The construction of the triplet $(\boldsymbol{X}, \boldsymbol{X}^n, \hat{\boldsymbol{Y}})$ is possible under the assumptions $(\boldsymbol{H_0})$ and $(\boldsymbol{H_1})$.

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 $ds\, d\zeta$, and $\mathcal{N}_1(ds, d\zeta)$ on $[0, T] \times [0, 1]$, with intensity measure $ds\, d\zeta$. Then for suitable functions K_0, K_1, K_0^n, K_1^n the triplet $(\boldsymbol{X}, \boldsymbol{X}^n, \hat{\boldsymbol{Y}})$ is

$$X_t = X_0 + \int_0^t \int_0^{\bar{\lambda}_0} K_0(X_{s-}, X_{s-}^n, \hat{Y}_{s-}; \zeta) \mathcal{N}_0(ds, d\zeta) \tag{55}$$

$$+ \int_0^t \int_0^1 K_1(X_{s-}, X_{s-}^n, \hat{Y}_{s-}; \zeta) \mathcal{N}_1(ds, d\zeta),$$

$$X_t^n = X_0^n + \int_0^t \int_0^{\bar{\lambda}_0} K_0^n(X_{s-}, X_{s-}^n, \hat{Y}_{s-}; \zeta) \mathcal{N}_0(ds, d\zeta) \tag{56}$$

$$+ \int_0^t \int_0^1 K_1^n(X_{s-}, X_{s-}^n, \hat{Y}_{s-}; \zeta) \mathcal{N}_1(ds, d\zeta),$$

$$\hat{Y}_t = y + \int_0^t \int_0^1 \mathcal{N}_1(ds, d\zeta) = y + \mathcal{N}_1((0, t] \times [0, 1]), \tag{57}$$

with initial conditions (X_0, X_0^n) independent of $\mathcal{N}_0$ and $\mathcal{N}_1$, and

$$\frac{d\mathbb{P}}{d\mathbb{Q}}\Big|_{\mathcal{F}_t^{X, X^n, \hat{Y}}} = \exp\left\{ \int_0^t \log\left(\lambda_1(X_{s-}, \hat{Y}_{s-})\right) d\hat{Y}_s - \int_0^t \left(\lambda_1(X_{s-}, \hat{Y}_{s-}) - 1\right) ds \right\},$$

$$\frac{d\mathbb{P}^n}{d\mathbb{Q}}\Big|_{\mathcal{F}_t^{X, X^n, \hat{Y}}} = \exp\left\{ \int_0^t \log\left(\lambda_1^n(X_{s-}^n, \hat{Y}_{s-})\right) d\hat{Y}_s - \int_0^t \left(\lambda_1^n(X_{s-}^n, \hat{Y}_{s-}) - 1\right) ds \right\}.$$

Remark 4.1. *By using Lemma 3.7 in [8], explicit expressions for the functions K_0, K_1, K_0^n, K_1^n can be given. In particular $K_0(x, \tilde{x}, y)$ and $K_0^n(x, \tilde{x}, y)$ are constructed using $\lambda(x) = \lambda_0(x, y)$, $\tilde{\lambda}(\tilde{x}) = \lambda_0^n(\tilde{x}, y)$, $\bar{\lambda} = \bar{\lambda}_0$, $\mu(x; dz) = \mu_0(x, y; dz)$, and $\tilde{\mu}(\tilde{x}; dz) = \mu_0^n(\tilde{x}, y; dz)$. While $K_1(x, \tilde{x}, y)$ and $K_1^n(x, \tilde{x}, y)$ are constructed using $\lambda(x) = \tilde{\lambda}(\tilde{x}) = \bar{\lambda} = 1$, $\mu(x; dz) = \mu_1(x, y; dz)$, and $\tilde{\mu}(\tilde{x}; dz) = \mu_1^n(\tilde{x}, y; dz)$.*

Note that this construction relies on the fact that the state space is one dimensional. Similar expressions could be obtained in higher dimensions.

Firstly, we deduce from the above Radon-Nikodym derivatives some upper bounds which will play a key role in the sequel. By using Lemma 4.7 in [8] we have

$$\mathbb{E}^n\left[\left|\frac{d\mathbb{P}}{d\mathbb{P}^n}\Big|_{\mathcal{F}_t^{X, X^n, \hat{Y}}} - 1\right|\right] = \mathbb{E}\left[\left|\frac{d\mathbb{P}^n}{d\mathbb{P}}\Big|_{\mathcal{F}_t^{X, X^n, \hat{Y}}} - 1\right|\right]$$

$$\leq 2\mathbb{E} \wedge \mathbb{E}^n\left[\int_0^t \left|\lambda_1(X_s, \hat{Y}_s) - \lambda_1^n(X_s^n, \hat{Y}_s)\right| ds\right] \tag{58}$$

$$\leq 2\int_0^t \mathbb{E} \vee \mathbb{E}^n\left[\left|\lambda_1(X_s, \hat{Y}_s) - \lambda_1^n(X_s^n, \hat{Y}_s)\right|\right] ds, \tag{59}$$

where $\mathbb{E} \wedge \mathbb{E}^n[Z] = \min\{\mathbb{E}[Z], \mathbb{E}^n[Z]\}$, and $\mathbb{E} \vee \mathbb{E}^n[Z] = \max\{\mathbb{E}[Z], \mathbb{E}^n[Z]\}$, for any random variable Z.

Secondly, note that since in our construction $\hat{\boldsymbol{Y}}^n = \hat{\boldsymbol{Y}}$, (43) and (48) coincide, and so the first upper bounds in Proposition 3.1 and Proposition 3.2 also coincide, and the same holds for (46) and (50) and consequently for the last upper bounds. Then, taking $r = a(t)$, the following result can be used to obtain upper bounds for the expectations of (43) and (46).

Proposition 4.2. *Under the hypotheses $(\boldsymbol{H_0})$ and $(\boldsymbol{H_1})$*

$$\mathbb{E} \vee \mathbb{E}^n \left[\sup_{\psi \in \mathcal{K}'_{TV}(\alpha)} \left| \hat{U}\left(r, \boldsymbol{y}; \psi(\cdot, y_r)\right) - \hat{U}^n\left(r, \boldsymbol{y}; \psi(\cdot, y_r)\right) \right|_{\boldsymbol{y} = \hat{\boldsymbol{Y}}} \right] \tag{60}$$

$$\leq \alpha \Big[4\|\lambda_1 - \lambda_1^n\| r +$$

$$\exp\left\{(\bar{\lambda}_1 - \underline{\lambda}_1)\, r\right\} \Big(\|\mu_0^X - \mu_0^{X^n}\|_{TV} \Big(\frac{2\bar{\lambda}_1}{\bar{\lambda}_1 - \underline{\lambda}_1} + 1 \Big) + (\bar{\lambda}_0 + \bar{\lambda}_1) J^n\, A(r) \Big) \Big]$$

where

$$A(r) = 2r + 4\bar{\lambda}_1 \frac{\exp\left\{-(\bar{\lambda}_1 - \underline{\lambda}_1)\, r\right\} + (\bar{\lambda}_1 - \underline{\lambda}_1)r - 1}{(\bar{\lambda}_1 - \underline{\lambda}_1)^2} \tag{61}$$

and

$$J^n = \max(\|p_0 - p_0^n\|_\infty, \|\mu_1 - \mu_1^n\|_\infty)/2, \tag{62}$$

with

$$p_0(x, y; dx') = \big(1 - (\lambda_0(x, y)/\bar{\lambda}_0)\big)\delta_x(dx') + \big(\lambda_0(x, y)/\bar{\lambda}_0\big)\, \mu_0(x, y; dx') \tag{63}$$

$$p_0^n(x, y; dx') = \big(1 - (\lambda_0^n(x, y)/\bar{\lambda}_0)\big)\delta_x(dx') + \big(\lambda_0^n(x, y)/\bar{\lambda}_0\big)\, \mu_0^n(x, y; dx'). \tag{64}$$

If furthermore, the operator $\boldsymbol{L}$ is bL-regular assumption $(\boldsymbol{H_2})$ holds, and the initial distributions μ_0^X and $\mu_0^{X^n}$ have finite first moments, then

$$\mathbb{E} \vee \mathbb{E}^n \left[\sup_{\psi \in \mathcal{K}'_{BL}(\alpha, \Lambda')} \left| \hat{U}\left(r, \boldsymbol{y}; \psi(\cdot, y_r)\right) - \hat{U}^n\left(r, \boldsymbol{y}; \psi(\cdot, y_r)\right) \right|_{\boldsymbol{y} = \hat{\boldsymbol{Y}}} \right] \tag{65}$$

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

$$+ \kappa\big(\mu_0^X, \mu_0^{X^n}\big) \Big(4\alpha L_{\lambda_1} \boldsymbol{\mathcal{E}}_1^M(r) + \Lambda' \boldsymbol{\mathcal{E}}_0^M(r) \Big)$$

where

$$H^n = \bar{\lambda}_0 \Delta_0^n + b_0\, \|\lambda_0 - \lambda_0^n\| + \bar{\lambda}_1 \Delta_1^n, \tag{66}$$

$$M = \bar{\lambda}_0(\Gamma_{\mu_0} - 1)^+ - \underline{\lambda}_0(1 - \Gamma_{\mu_0})^+ + b_0 L_{\lambda_0} + \bar{\lambda}_1(\Gamma_{\mu_1} - 1)^+ - \underline{\lambda}_1(1 - \Gamma_{\mu_1})^+, \tag{67}$$

and

$$\boldsymbol{\mathcal{E}}_k^M(t) := \frac{1}{M^k} \sum_{h=k}^{\infty} \frac{(tM)^h}{h!} = \frac{1}{M^k} \Big(\exp\{tM\} - \sum_{h=0}^{k-1} \frac{(tM)^h}{h!} \Big). \tag{68}$$

The proof of the above result is given in the next subsection.

Remark 4.3. *As a consequence, taking $\alpha = 1$, we can obtain*

$$\mathbb{E} \vee \mathbb{E}^n \left[\|\hat{U}(r, \boldsymbol{y}) - \hat{U}^n(r, \boldsymbol{y})\|_{TV} \Big|_{\boldsymbol{y} = \hat{\boldsymbol{Y}}} \right] \tag{69}$$

$$\leq \Big[4\|\lambda_1 - \lambda_1^n\|r +$$

$$\exp \{(\bar{\lambda}_1 - \underline{\lambda}_1)\, r\} \Big(\|\mu_0^X - \mu_0^{X^n}\|_{TV} \Big(\frac{2\bar{\lambda}_1}{\bar{\lambda}_1 - \underline{\lambda}_1} + 1 \Big) + (\bar{\lambda}_0 + \bar{\lambda}_1) J^n\, A(r) \Big) \Big],$$

and similarly, taking $\Lambda' = 1$,

$$\mathbb{E} \vee \mathbb{E}^n \left[d_{BL} \big(\hat{U}(r, \boldsymbol{y}), \hat{U}^n(r, \boldsymbol{y}) \big) \Big|_{\boldsymbol{y} = \hat{\boldsymbol{Y}}} \right] \tag{70}$$

$$\leq 4 \Big[\|\lambda_1 - \lambda_1^n\|r + L_{\lambda_1} H^n\, \boldsymbol{\mathcal{E}}_2^M(r) \Big] + H^n\, \boldsymbol{\mathcal{E}}_1^M(r)$$

$$+ \boldsymbol{\kappa}\big(\mu_0^X, \mu_0^{X^n}\big) \Big(4 L_{\lambda_1} \boldsymbol{\mathcal{E}}_1^M(r) + \boldsymbol{\mathcal{E}}_0^M(r) \Big).$$

The above upper bounds are interesting by themselves, since they correspond to upper bounds for the approximations of the system without delay. The last bound was essentially obtained in [8], Remark 2.7. However in that bound there was an imprecision, and (70) corrects it.

Note that the upper bound for (69) grows exponentially fast in r, and therefore is valid only in a finite interval of time, namely when the l.h.s. is less or equal than 2. A similar consideration holds for the upper bound for (60). Moreover note that we have used implicitly that $\bar{\lambda}_1 - \underline{\lambda}_1 > 0$. The upper bound (69) grows exponentially fast also when $\boldsymbol{L} = \boldsymbol{L^n}$, i.e. when the approximating system and the limit one differ only in the distribution of the initial condition, but it simplifies to

$$\mathbb{E} \vee \mathbb{E}^n \left[\|\hat{U}(r, \boldsymbol{y}) - \hat{U}^n(r, \boldsymbol{y})\|_{TV} \Big|_{\boldsymbol{y} = \hat{\boldsymbol{Y}}} \right] \leq \exp \{(\bar{\lambda}_1 - \underline{\lambda}_1)\, r\} \Big(\frac{2\bar{\lambda}_1}{\bar{\lambda}_1 - \underline{\lambda}_1} + 1 \Big) \|\mu_0^X - \mu_0^{X^n}\|_{TV}, \tag{71}$$

since in this case $\|\lambda_1 - \lambda_1^n\| = 0$ and $J^n = 0$.
As far as the bounds (70) and (65) are concerned, note that if $\boldsymbol{L} = \boldsymbol{L^n}$, then the bound (70) simplifies to

$$\mathbb{E} \vee \mathbb{E}^n \left[d_{BL} \big(\hat{U}(r, \boldsymbol{y}), \hat{U}^n(r, \boldsymbol{y}) \big) \Big|_{\boldsymbol{y} = \hat{\boldsymbol{Y}}} \right] \leq \boldsymbol{\kappa}\big(\mu_0^X, \mu_0^{X^n}\big) \Big(4 L_{\lambda_1} \boldsymbol{\mathcal{E}}_1^M(r) + \boldsymbol{\mathcal{E}}_0^M(r) \Big), \tag{72}$$

since in this case $\|\lambda_1 - \lambda_1^n\| = 0$ and $H^n = 0$.

Moreover note that the constants H^n and M correspond to the constants defined in (2.9) and (2.10) of Theorem 2.1 in [8] respectively, and that the constant M can be either positive or negative: for instance if $\Gamma_{\mu_1} < 1$, then for any $\lambda_0(x, y)$ and $\mu_0(x, y; dx')$, $M = \bar{\lambda}_0 (\Gamma_{\mu_0} - 1)^+ - \underline{\lambda}_0 (1 - \Gamma_{\mu_0})^+ + b_0 L_{\lambda_0} - \underline{\lambda}_1 (1 - \Gamma_{\mu_1})$, so that for $\underline{\lambda}_1$ sufficiently large, the constant M is negative.
Furthermore it is interesting to note that in the case of the observation of a fixed parameter, i.e. $X_t = X_0$, and $X_t^n = X_0^n$, observed at times which form a standard Poisson process independent of $\boldsymbol{X}$ then $M = 0$, since in this case $\lambda_0(x) = \lambda_0^n(x) = 0$, $L_{\lambda_0} = 0$, $\lambda_1(x) = \lambda_1^n(x) = 1$, and $\Gamma_i = 1$.
The condition that the value of $\underline{\lambda}_1$ is sufficiently large can be interpreted as the condition that the times of observation are sufficiently frequent. In this case the error evaluated in the bounded-Lipschitz metric is of the same order of the error at time zero, evaluated in the Kantorovitch metric. If instead $\Gamma_{\mu_1} \geq 1$, but $\Gamma_{\mu_0} < 1$, then $M = -\underline{\lambda}_0 (1 - \Gamma_{\mu_0})^+ + b_0 L_{\lambda_0} + \bar{\lambda}_1 (\Gamma_{\mu_1} - 1)$, and so for $\underline{\lambda}_0$ sufficiently large, again M is negative.

Finally note that for any $k \geq 0$ and for positive M, $\boldsymbol{\mathcal{E}}_k^M(t)$ increases to infinity exponentially fast, with $\boldsymbol{\mathcal{E}}_k^M(t) \leq \dfrac{t^k \exp\{tM\}}{k!}$. Hence the bounds (70) and (65) are meaningful only in a finite interval of time, as the bound for the total variation distance. When M is negative, we have that

$$\lim_{t \to \infty} \boldsymbol{\mathcal{E}}_k^M(t) = \infty, \qquad \text{for } k \geq 2,$$

23

but the rate of convergence is polynomial (indeed $\lim_{t\to\infty} \mathcal{E}_k^M(t)/t^{k-1} = (k-1)!/|M|$), while

$$\lim_{t\to\infty} \mathcal{E}_0^M(t) = \lim_{t\to\infty} e^{Mt} = 0,$$

and

$$\lim_{t\to\infty} \mathcal{E}_1^M(t) = \lim_{t\to\infty} \frac{e^{Mt}-1}{M} = \frac{1}{|M|}.$$

Hence, for negative M, though the general bounds (70) and (65) are meaningful only in a finite interval of time, when the systems differ only for the initial conditions, i.e. when $\boldsymbol{L} = \boldsymbol{L^n}$, we can use (72) and get that

$$\lim_{r\to\infty} \mathbb{E} \vee \mathbb{E}^n \big[d_{BL}\big(\hat{U}(r,\boldsymbol{y}),\hat{U}^n(r,\boldsymbol{y})\big)\big]\Big|_{\boldsymbol{y}=\hat{\boldsymbol{Y}}} \leq \boldsymbol{\kappa}\big(\mu_0^X,\mu_0^{X^n}\big)\frac{4L_{\lambda_1}}{|M|}. \tag{73}$$

The latter result can be interpreted as a kind of stability property for the filter with respect to its initial condition. Asymptotic stability properties with respect to the Hilbert projective metric have been studied by Budhiraja and Kushner [3] in the case when $\boldsymbol{X}$ is an ergodic Markov process with generator $\boldsymbol{A}$, and the system $(\boldsymbol{X},\boldsymbol{Y})$ is described by $\boldsymbol{L}\phi(x,y) = \big(\boldsymbol{A}\phi(\cdot,y)\big)(x) + \lambda_1(x,y)[\phi(x,y+1) - \phi(x,y)]$. Their result does not require $\boldsymbol{X}$ to be a jump process, however when this is the case simultaneous jump times of $\boldsymbol{X}$ and $\boldsymbol{Y}$ are not allowed. In order to compare their result with (73) note that on the one hand we need $\boldsymbol{X}$ to be a jump process, while on the other hand simultaneous jump times are not prohibited and we do not need $\boldsymbol{X}$ to be Markov.

The same kind of considerations hold for the next upper bounds and can be used to give upper bounds for (44) and (47).

Proposition 4.4. *Under the assumptions* $(\boldsymbol{H_0}), (\boldsymbol{H_1})$ *and the notations of Proposition 4.2,*

$$\sup_{\phi\in\mathcal{K}'_{TV}(\alpha)} \sup_{x,y} |\exp\{\boldsymbol{L}t\}\phi(x,y) - \exp\{\boldsymbol{L^n}t\}\phi(x,y)|$$
$$\leq \alpha\Big(2\|\lambda_1 - \lambda_1^n\|\, t + \exp\{(\bar{\lambda}_1 - \underline{\lambda}_1)t\}(\bar{\lambda}_0 + \bar{\lambda}_1)\, J^n\, B(t)\Big) \tag{74}$$

where

$$B(t) = 2t + 2\bar{\lambda}_1 \frac{\exp\{-(\bar{\lambda}_1 - \underline{\lambda}_1)\,t\} + (\bar{\lambda}_1 - \underline{\lambda}_1)t - 1}{(\bar{\lambda}_1 - \underline{\lambda}_1)^2} = t + A(t)/2. \tag{75}$$

If furthermore the operator $\boldsymbol{L}$ *is bL-regular, assumption* $(\boldsymbol{H_2})$ *holds, and the initial distributions* μ_0^X *and* $\mu_0^{X^n}$ *have finite first moments, then*

$$\sup_{\phi\in\mathcal{K}'_{BL}(\alpha,\Lambda)} \sup_{x,y} |\exp\{\boldsymbol{L}t\}\phi(x,y) - \exp\{\boldsymbol{L^n}t\}\phi(x,y)|$$
$$\leq 2\alpha\Big[\|\lambda_1 - \lambda_1^n\|t + L_{\lambda_1}H^n\, \mathcal{E}_2^M(t)\Big] + \Lambda H^n\, \mathcal{E}_1^M(t). \tag{76}$$

The proof of the above result is given in the next subsection.

Using the results and the notations of Proposition 3.1 and Proposition 3.2, with $\alpha = 1$, $\Lambda = 1$ and $\Lambda'(t-r,1) = e^{M(t-r)}$ (see (52)), then using Proposition 4.2 with $\Lambda' = \Lambda'(t-r,1)$ and Proposition 4.4, we get immediately the main result of this section.

Theorem 4.5. *Under the assumptions* $(\boldsymbol{H_0})$ *and* $(\boldsymbol{H_1})$, *and the notations introduced in the previous Propositions 4.2 and 4.4*

$$E\big[\|\pi_t - \tilde{\pi}_t^n\|_{TV}\big] \vee E\big[\|\pi_t^n - \overline{\pi}_t^n\|_{TV}\big]$$

$$\leq \Big[2\|\lambda_1 - \lambda_1^n\|(t+r) + \|\mu_0^X - \mu_0^{X^n}\|_{TV} \exp\{(\overline{\lambda}_1 - \underline{\lambda}_1)\,r\}\Big(\frac{2\overline{\lambda}_1}{\overline{\lambda}_1 - \underline{\lambda}_1} + 1\Big) + \tag{77}$$

$$(\overline{\lambda}_0 + \overline{\lambda}_1)J^n\,\big(\exp\{(\overline{\lambda}_1 - \underline{\lambda}_1)\,r\}A(r) + \exp\{(\overline{\lambda}_1 - \underline{\lambda}_1)\,(t-r)\}B(t-r)\big)\Big]\Big|_{r=a(t)}.$$

Furthermore assume $(\boldsymbol{H_2})$, *the bL-regularity of* $\boldsymbol{L}$, *and that the initial distributions* μ_0^X *and* $\mu_0^{X^n}$ *have finite first moments. Then*

$$E\big[d_{BL}(\pi_t, \tilde{\pi}_t^n)\big] \vee E\big[d_{BL}(\pi_t^n, \overline{\pi}_t^n)\big]$$

$$\leq 4\Big[\|\lambda_1 - \lambda_1^n\|r + L_{\lambda_1}H^n\,\boldsymbol{\mathcal{E}}_2^M(r)\Big] + \exp\{M(t-r)\}H^n\,\boldsymbol{\mathcal{E}}_1^M(r)$$

$$+ \boldsymbol{\kappa}\big(\mu_0^X, \mu_0^{X^n}\big)\Big(4L_{\lambda_1}\boldsymbol{\mathcal{E}}_1^M(r) + \boldsymbol{\mathcal{E}}_0^M(t)\Big) \tag{78}$$

$$+ 2\Big(\|\lambda_1 - \lambda_1^n\|(t-r) + L_{\lambda_1}H^n\,\boldsymbol{\mathcal{E}}_2^M(t-r)\Big) + H^n\,\boldsymbol{\mathcal{E}}_1^M(t-r)\Big|_{r=a(t)}.$$

To conclude this section we work out an example highlighting the bounds in (78) by using a particular case of the model considered in Remark 2.5 in [8].

Example Assume that the measures $\mu_i(x, y, \cdot)$ are Gaussian with mean $m_i(x, y)$, and variance $\sigma_i^2(x, y)$, and that the functions $\lambda_i(z, y)$, $m_i(z, y)$, and $\sigma_i(z, y)$ are Lipschitz in z, uniformly in y, with Lipschitz constant L_{λ_i}, L_{m_i} and L_{σ_i}, respectively. We assume also that the functions $\lambda_i(z, y)$, $\sigma_0(z, y)$ and $\mathsf{m}_0(z, y) := m_0(z, y) - z$ are uniformly bounded by $\overline{\lambda}_i$, $\overline{\sigma}_0$ and $\overline{\mathsf{m}}_0$, respectively, and that $\sigma_1(0, y) \leq \mathsf{K}_1^\sigma(y+1)$ and $|m_1(0, y)| \leq \mathsf{K}_1^m(y+1)$.

Then conditions $(\boldsymbol{R_0})$, $(\boldsymbol{R_1})$ and $(\boldsymbol{R_2})$ are satisfied with

$$\Gamma_{\mu_i} = \sqrt{2/\pi}\,L_{\sigma_i} + L_{m_i},$$

$$b_0 = \overline{\sigma}_0 + \overline{\mathsf{m}}_0,\ b_1 = \mathsf{K}_1^\sigma + \mathsf{K}_1^m,\ \text{and}\ a_1 = \big(L_{\sigma_1} + L_{m_1}\big) \vee b_1$$

as can be easily deduced by the assumptions, the upper bound

$$\boldsymbol{\kappa}\big(\mu_i(x, y; \cdot), \mu_i(x', y; \cdot)\big) \leq \sqrt{2/\pi}\,|\sigma_i(x, y) - \sigma_i(x', y)| + |m_i(x, y) - m_i(x', y)|$$

$$\leq \Big(\sqrt{2/\pi}\,L_{\sigma_i} + L_{m_i}\Big)\,|x - x'|,$$

and the bound

$$\int_{\mathbb{R}} |z - x|\,\mu_i(x, y; dz) \leq \int_{\mathbb{R}} |z - m_i(x, y)|\,\mu_i(x, y; dz) + \int_{\mathbb{R}} |m_i(x, y) - x|\,\mu_i(x, y; dz)$$

$$\leq \sigma_i(x, y) + |m_i(x, y) - x|.$$

As far as the approximating process is concerned we define

$$\lambda_i^n(x, y) = \lambda\big(\tfrac{\lfloor nx \rfloor}{n}, y\big), \quad \mu_i^n(x, y; A) := T^n \mu_i\big(\tfrac{\lfloor nx \rfloor}{n}, y; A\big), \quad \text{and} \quad \pi_0^n(A) = T^n \pi_0(A)$$

where, for any probability measure ν,

$$T^n \nu(A) = \sum_k \nu\Big(\big[\tfrac{k}{n}, \tfrac{k+1}{n}\big)\Big)\,\delta_{\{\frac{k}{n}\}}(A).$$

Then condition $(\boldsymbol{H_0})$ is obviously satisfied, condition $(\boldsymbol{H_1})$ is satisfied if we assume furthermore $\lambda_1(x,y)$ to be strictly positive, and finally condition $(\boldsymbol{H_2})$ is satisfied with

$$\Delta_i^n = \left(\sqrt{2/\pi}\, L_{\sigma_i} + L_{m_i} + 1\right)\frac{1}{n}$$

since $\boldsymbol{\kappa}\big(\nu, \mathcal{T}^n\nu\big) \le \frac{1}{n}$, so that

$$\boldsymbol{\kappa}\big(\mu_i(x,y;\cdot), \mu_i^n(x,y;\cdot)\big) \le \boldsymbol{\kappa}\big(\mu_i(x,y;\cdot), \mu_i(\tfrac{\lfloor nx \rfloor}{n}, y;\cdot)\big) + \boldsymbol{\kappa}\big(\mu_i(\tfrac{\lfloor nx \rfloor}{n}, y;\cdot), \mathcal{T}^n\mu_i(\tfrac{\lfloor nx \rfloor}{n}, y;\cdot)\big)$$

$$\le \left(\sqrt{2/\pi}\, L_{\sigma_i} + L_{m_i} + 1\right)\frac{1}{n}.$$

Then $E\big[d_{BL}(\pi_t, \tilde{\pi}_t^n)\big]$ and $E\big[d_{BL}(\pi_t^n, \overline{\pi}_t^n)\big]$ are of order $\frac{1}{n}$, since the quantities involved in the upper bound (78) are of order $\frac{1}{n}$. Indeed

$$\|\lambda_i - \lambda_i^n\| \le L_{\lambda_i}\frac{1}{n}, \quad \boldsymbol{\kappa}(\pi_0, \pi_0^n) \le \frac{1}{n}$$

and finally, recalling (66) and the previously obtained upper bounds,

$$H^n \le \left[\bar{\lambda}_0\big(\sqrt{2/\pi}\, L_{\sigma_0} + L_{m_0} + 1\big) + (\overline{\sigma}_0 + \overline{m}_0)\, L_{\lambda_0} + \bar{\lambda}_1\big(\sqrt{2/\pi}\, L_{\sigma_1} + L_{m_1} + 1\big)\right]\frac{1}{n}.$$

4.2 Proofs of Propositions 4.2 and 4.4

Using the construction of the triplet $(\boldsymbol{X}, \boldsymbol{X^n}, \hat{\boldsymbol{Y}})$, it is easy to prove the following auxiliary results. In the first two results we need the processes

$$N_0(t) = \mathcal{N}_0\big((0,t] \times [0, \bar{\lambda}_0]\big), \quad N_1(t) = \mathcal{N}_1\big((0,t] \times [0,1]\big),$$

which are independent Poisson processes under $\mathbb{Q}$, with parameter $\bar{\lambda}_0$ and 1, respectively.

Proposition 4.6. *If the assumptions $(\boldsymbol{H_0}), (\boldsymbol{H_1})$ hold and if the initial conditions (X_0, X_0^n) in (55) and (56) are coupled in such a way that $\mathbb{Q}(X_0 \ne X_0^n) = \|\mu_0^X - \mu_0^{X^n}\|_{TV}/2$ (see e.g. [8]), then*

$$\mathbb{P} \vee \mathbb{P}^n\big(X_t \ne X_t^n\big) \le \exp\{(\bar{\lambda}_1 - \underline{\lambda}_1)\, t\}\Big(\|\mu_0^X - \mu_0^{X^n}\|_{TV}/2 + (\bar{\lambda}_0 + \bar{\lambda}_1)\, t J^n\Big)$$

where the constant J^n has been introduced in (62) of Proposition 4.2.

Proof. To get a bound for $\mathbb{P} \vee \mathbb{P}^n\big(X_t \ne X_t^n\big)$ observe that

$$\mathbb{P}\big(X_t \ne X_t^n\big) = \mathbb{E}^{\mathbb{Q}}\big[(d\mathbb{P}/d\mathbb{Q})|_{\mathcal{F}_t^{X,X^n,\hat{Y}}}\, \mathbb{I}_{\{X_t \ne X_t^n\}}\big]$$

and that

$$(d\mathbb{P}/d\mathbb{Q})|_{\mathcal{F}_t^{X,X^n,\hat{Y}}} = \exp\left\{\int_0^t \log\big(\lambda_1(X_{s-}, \hat{Y}_{s-})\big)\, d\hat{Y}_s - \int_0^t \big(\lambda_1(X_{s-}, \hat{Y}_{s-}) - 1\big)\, ds\right\}$$

$$= \exp\left\{\int_0^t \log\big(\lambda_1(X_{s-}, \hat{Y}_{s-})\big)\, dN_1(s) - \int_0^t \big(\lambda_1(X_{s-}, \hat{Y}_{s-}) - 1\big)\, ds\right\}$$

$$\le \big(\bar{\lambda}_1\big)^{N_1(t)} \exp\{-(\underline{\lambda}_1 - 1)t\}.$$

As a consequence

$$\mathbb{P}\big(X_t \ne X_t^n\big) \le \mathbb{E}^{\mathbb{Q}}\big[\big(\bar{\lambda}_1\big)^{N_1(t)} \mathbb{I}_{\{X_t \ne X_t^n\}}\big] \exp\{-(\underline{\lambda}_1 - 1)t\}. \tag{79}$$

In a similar way we prove that

$$\mathbb{P}^n\big(X_t \ne X_t^n\big) \le \mathbb{E}^{\mathbb{Q}}\big[\big(\bar{\lambda}_1\big)^{N_1(t)} \mathbb{I}_{\{X_t \ne X_t^n\}}\big] \exp\{-(\underline{\lambda}_1 - 1)t\}. \tag{80}$$

The result then follows immediately from the next lemma, which is proved in the Appendix. $\square$

Lemma 4.7. *Under the hypotheses and notations of Proposition 4.6*

$$\mathbb{E}^{\mathbb{Q}}\big[\,(\bar{\lambda}_1)^{N_1(t)}\,\mathbb{I}_{\{X_t \neq X_t^n\}}\big] \leq \exp\{(\bar{\lambda}_1 - 1)\,t\}\Big(\|\mu_0^X - \mu_0^{X^n}\|_{TV}/2 + (\bar{\lambda}_0 + \bar{\lambda}_1)\,tJ^n\Big). \tag{81}$$

The following result plays a key role in our analysis in order to get the upper bounds when dealing with the bounded Lipschitz metric.

Proposition 4.8. *If the operator $\mathbf{L}$ is bL-regular, the hypotheses $(\mathbf{H_0}) - (\mathbf{H_2})$ hold, and if the initial conditions (X_0, X_0^n) in (55) and (56) have finite first moments and are coupled in such a way that $\mathbb{E}_Q\big[\,|X_0 - X_0^n|\,\big] = \boldsymbol{\kappa}\big(\mu_0^X, \mu_0^{X^n}\big)$ (see e.g. [8]), then*

$$\mathbb{E} \vee \mathbb{E}^n\big[\,|X_t - X_t^n|\,\big] \leq \boldsymbol{\kappa}\big(\mu_0^X, \mu_0^{X^n}\big)\boldsymbol{\mathcal{E}}_0^M(t) + H^n\boldsymbol{\mathcal{E}}_1^M(t)$$

where the constants H^n, M and the functions $\boldsymbol{\mathcal{E}}_k^M$ have been introduced in Proposition 4.2.

Proof. The proof of this result is the same as the proof of Lemma 5.1 in [8], with obvious changes in the notations, in particular we have to substitute $\tilde{X}$ with X^n. Indeed, as in that proof, one can show

$$\mathbb{E}\big[|X_t - X_t^n|\big] \leq \mathbb{E}\big[|X_0 - X_0^n|\big] + \int_0^t \mathbb{E}\big[H + M|X_s - X_s^n|\big]ds,$$

and a similar inequality w.r.t. $\mathbb{E}^n$. Then the statement follows by Gronwall inequality. We note also that there was a mistake in the statement of Lemma 5.1 of [8], and this result corrects it. $\square$

As a consequence of Proposition 4.8 we get immediately that the class of bounded functions ψ which are Lipschitz in x for all y, uniformly in y is invariant under the action of the semigroup $\exp\{\mathbf{L}t\}$.

Corollary 4.9. *Let $\mathbf{L}$ be a bL-regular operator, with λ_i, $i = 0, 1$ satisfying hypotheses $(\mathbf{H_0})$ and $(\mathbf{H_1})$, let ψ be a measurable bounded function such that, for all x, x', y, $|\psi(x,y) - \psi(x',y)| \leq L_\psi|x - x'|$.
Then*

$$|\exp\{\mathbf{L}t\}\psi(x,y) - \exp\{\mathbf{L}t\}\psi(x',y)| \leq e^{Mt}L_\psi|x - x'|, \qquad \text{for all } x, x', y,$$

and

$$\exp\{\mathbf{L}t\}\big(K'(\alpha, \Lambda)\big) \subseteq K'(\alpha, e^{Mt}\Lambda).$$

Proof. Assume that $\mathbf{L^n} = \mathbf{L}$, so that $H^n = 0$, and assume that $\mu_0^X = \delta_{\{x\}}$, $\mu_0^{X^n} = \delta_{\{x'\}}$, so that $\boldsymbol{\kappa}\big(\mu_0^X, \mu_0^{X^n}\big) = |x - x'|$. Then

$$|\exp\{\mathbf{L}t\}\psi(x,y) - \exp\{\mathbf{L}t\}\psi(x',y)|$$
$$= \Big|\mathbb{E}\big[\psi(X_t, \hat{Y}_t)\big] - \mathbb{E}\big[\psi(X_t^n, \hat{Y}_t)\big]\Big| \leq L_\psi\mathbb{E}[|X_t - X_t^n|],$$

and, observing that $(\mathbf{H_2})$ is trivially satisfied, and that $\boldsymbol{\mathcal{E}}_0^M(t) = e^{Mt}$, Proposition 4.8 implies the result.
The second statement is an immediate consequence of the first one. $\square$

Remark 4.10. *We note that the boundedness condition on the function ψ is not necessary we could ask for a bound of the kind $|\psi(x,y)| \leq A(|x| + y) + B$: indeed, under the conditions of the previous Corollary 4.9, $\exp(\mathbf{L}t)\psi(x,y)$ is finite for any ψ Lipschitz in x uniformly in y: in Proposition 4.8 the condition on the first moments of (X_0, X_0^n) guarantees the integrability of (X_t, X_t^n) for all $t \geq 0$, and when $(X_0, X_0^n) = (x, x')$ this condition is obviously satisfied and implies that $\exp(\mathbf{L}t)\psi(x,y) = \mathbb{E}\big[\psi(X_t, \hat{Y}_t)\big]$ is finite, for all such functions which are Lipschitz in x uniformly in y.*

Proof. **(of Proposition 4.2)**
In order to get the first estimate, we can use the first part of Theorem 2.9 with $(\boldsymbol{X}, \boldsymbol{X^n}, \hat{\boldsymbol{Y}})$ and r instead of $(\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{X}}^{\boldsymbol{n}}, \tilde{\boldsymbol{Y}})$ and t, and get that (60) is bounded above by

$$2\alpha \max \left(\mathbb{E}[\hat{\mathcal{Z}}_r] + \mathbb{P}(X_r \neq X_r^n) \, ; \, \mathbb{E}^n[\hat{\mathcal{Z}}_r^n] + \mathbb{P}^n(X_r \neq X_r^n) \right),$$

where $\hat{\mathcal{Z}}_r$ and $\hat{\mathcal{Z}}_r^n$ are defined as $\mathcal{Z}_r$ and $\mathcal{Z}_r^n$, with obvious changes. Using Remark 2.4 we get $\mathbb{E}[\hat{\mathcal{Z}}_r]$ and $\mathbb{E}^n[\hat{\mathcal{Z}}_r^n]$ are equal to

$$\mathbb{E}^n \left[\left| \left. \frac{d\mathbb{P}}{d\mathbb{P}^n} \right|_{\mathcal{F}_t^{X,X^n,\hat{Y}}} - 1 \right| \right] = \mathbb{E} \left[\left| \left. \frac{d\mathbb{P}^n}{d\mathbb{P}} \right|_{\mathcal{F}_t^{X,X^n,\hat{Y}}} - 1 \right| \right].$$

Therefore, by (59), (60) is bounded above by

$$2\alpha \left(2 \int_0^r \mathbb{E} \vee \mathbb{E}^n \left[\left| \lambda_1(X_s, \hat{Y}_s) - \lambda_1^n(X_s^n, \hat{Y}_s) \right| \right] ds + \mathbb{P} \vee \mathbb{P}^n (X_r \neq X_r^n) \right)$$

$$\leq 4\alpha \int_0^r \left(\|\lambda_1 - \lambda_1^n\| + \bar{\lambda}_1 \mathbb{P} \vee \mathbb{P}^n (X_s \neq X_s^n) \right) ds + 2\alpha \, \mathbb{P} \vee \mathbb{P}^n (X_r \neq X_r^n).$$

The thesis is achieved by applying Proposition 4.6.

To obtain the second estimate, we can use the second part of Theorem 2.9, again with $(\boldsymbol{X}, \boldsymbol{X^n}, \hat{\boldsymbol{Y}})$ and r instead of $(\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{X}}^{\boldsymbol{n}}, \tilde{\boldsymbol{Y}})$ and t, and get that (65) is bounded above by

$$\max \left(2\alpha \, \mathbb{E}[\hat{\mathcal{Z}}_r] + \Lambda \mathbb{E}(|X_r - X_r^n|) \, ; \, 2\alpha \, \mathbb{E}^n[\hat{\mathcal{Z}}_r^n] + \Lambda \, \mathbb{E}^n(|X_r - X_r^n|) \right).$$

Therefore, by (59), (65) is bounded above by

$$4\alpha \int_0^r \mathbb{E} \vee \mathbb{E}^n \left[\left| \lambda_1(X_s, \hat{Y}_s) - \lambda_1^n(X_s^n, \hat{Y}_s) \right| \right] ds + \Lambda \, \mathbb{E} \vee \mathbb{E}^n[|X_r - X_r^n|]$$

$$\leq 4\alpha \int_0^r \left(\|\lambda_1 - \lambda_1^n\| + L_{\lambda_1} \mathbb{E} \vee \mathbb{E}^n \left[|X_s - X_s^n| \right] \right) ds + \Lambda \, \mathbb{E} \vee \mathbb{E}^n \left[|X_r - X_r^n| \right].$$

The thesis is achieved by applying Proposition 4.8, and using the fact that $\boldsymbol{\mathcal{E}}_{k+1}^M(r) = \int_0^r \boldsymbol{\mathcal{E}}_k^M(s) \, ds$. $\qquad \square$

The proof of Proposition 4.4 is similar to the previous one, but instead of using Theorem 2.9 we use Proposition 2.11, with initial distributions $\mu_0^X = \mu_0^{X^n} = \delta_{\{x\}}$. The last fact explains the similarity of the upper bounds in Propositions 4.2 and Proposition 4.4.

Remark 4.11. *When the first moments are not finite the result of Proposition 4.8 cannot be used. However we can prove that*

$$\mathbb{E} \vee \mathbb{E}^n \left[|X_t - X_t^n| \wedge 1 \right] \leq d_{BL}\left(\mu_0^X, \mu_0^{X^n} \right) \boldsymbol{\mathcal{E}}_0^{\bar{M}}(t) + \bar{H}^n \boldsymbol{\mathcal{E}}_1^{\bar{M}}(t) \tag{82}$$

where $\bar{M} = (M_0)^+ + \bar{\lambda}_1 (\Gamma_{\mu_1} - 1)^+$ and $\bar{H}^n = H_0^n \wedge \bar{\lambda}_0 + \bar{\lambda}_1(\Delta_1^n \wedge 1)$, with $H_0^n = \bar{\lambda}_0 \Delta_0^n + \|\lambda_0 - \lambda_0^n\| b_0$, and $M_0 = \bar{\lambda}_0 (\Gamma_{\mu_0} - 1)^+ - \underline{\lambda}_0 (1 - \Gamma_{\mu_0})^+ + b_0 L_{\lambda_0}$. Indeed using the same techniques as in the proofs of Lemma 5.1 in [8] and Lemma 4.3 in [7] when setting $g_1(x, x') = |x - x'| \wedge 1$ we get

$$\mathbb{L}g_1(x, x') \leq \left(H_0^n + M_0|x - x'| + \bar{\lambda}_0|x - x'| \right) \wedge \bar{\lambda}_0 - \bar{\lambda}_0(|x - x'| \wedge 1)$$
$$+ \lambda_1(x, y) \left[(\Delta_1^n + \Gamma_{\mu_1}|x - x'|) \wedge 1 - |x - x'| \wedge 1 \right].$$

Then taking into account that $M_0 + \bar{\lambda}_0 \geq 0$, and that if $a, b, \lambda \geq 0$, then $(a + b) \wedge \lambda \leq a \wedge \lambda + b \wedge \lambda$, and $(ab) \wedge \lambda \leq (a \vee \lambda)(b \wedge 1)$, we have

$$\mathbb{L}g_1(x, x') \leq H_0^n \wedge \bar{\lambda}_0 + \left((M_0 + \bar{\lambda}_0) \vee \bar{\lambda}_0 \right) (|x - x'| \wedge 1) - \bar{\lambda}_0(|x - x'| \wedge 1)$$
$$+ \lambda_1(x, y) \left[\Delta_1^n \wedge 1 + (\Gamma_{\mu_1} \vee 1)(|x - x'| \wedge 1) - (|x - x'| \wedge 1) \right]$$
$$= H_0^n \wedge \bar{\lambda}_0 + \lambda_1(x, y)(\Delta_1^n \wedge 1) + \left((M_0)^+ + \lambda_1(x, y)(\Gamma_{\mu_1} - 1)^+ \right) (|x - x'| \wedge 1)$$
$$\leq \bar{H}^n + \bar{M}(|x - x'| \wedge 1).$$

Therefore, when the hypothesis on the first moments does not hold, taking (82) *and the estimates in Remarks 2.5, 2.10 and 2.12 into account, the upper bound* (78) *of Theorem 4.5 remains valid with minor changes in the constants, and substituting M and H^n with $\bar{M}$ and $\bar{H}^n$, respectively.*

4.3 Appendix: Proof of Lemma 4.7

In order to prove (81) it is sufficient to prove that

$$\mathbb{E}^{\mathbb{Q}}\big[\,(\bar{\lambda}_1)^{N_1(t)}\,\mathbb{I}_{\{X_t \neq X_t^n\}}\big] \leq \mathbb{E}^{\mathbb{Q}}\big[\,(\bar{\lambda}_1)^{N_1(t)}\big]\|\mu_0^X - \mu_0^{X^n}\|_{TV}/2 + \mathbb{E}^{\mathbb{Q}}\big[\big(N_0(t)+N_1(t)\big)\,(\bar{\lambda}_1)^{N_1(t)}\big]\,J^n. \quad (83)$$

Indeed the r.h.s. of (83) is equal to the r.h.s. of (81), since $\boldsymbol{N_0}$ and $\boldsymbol{N_1}$ are independent under $\mathbb{Q}$ and for any Poisson process $M(t)$ of parameter ρ, and for any $a > 0$, it holds $\mathbb{E}[M(t)] = \rho t$, $\mathbb{E}\big[a^{M(t)}\big] = \exp\{\rho(a-1)t\}$, $\mathbb{E}\big[M(t)\,a^{M(t)}\big] = \rho a\, t \exp\{\rho(a-1)t\}$.

Let T_k denote the jump times of the Poisson process $N(t) = N_0(t) + N_1(t)$, where we recall that $N_0(t) = \mathcal{N}_0\big([0, \bar{\lambda}_0] \times (0,t]\big)$ and $N_1(t) = \mathcal{N}_1\big([0,1] \times (0,t]\big)$. Then the processes defined by (55), (56) and (57), i.e.

$$\big(X_t, X_t^n, \hat{Y}_t\big) = \big(X_t, X_t^n, y + N_1(t)\big)$$

are obtained as a Markov chain $(\eta_k, \eta_k^n, \hat{y}_k)$ evaluated at time $k = N(t)$, so that $(\eta_k, \eta_k^n, \hat{y}_k) = \big(X_{T_k}, X_{T_k}^n, \hat{Y}_{T_k}\big)$. Moreover the Markov chain and the Poisson process $\{N(t), t > 0\}$ are independent.

Let $S \in \{T_h,\ h \geq 0\}$ be the first time when $\{X_t \neq X_t^n\}$, then

$$
\begin{aligned}
\mathbb{E}^{\mathbb{Q}}\big[\,(\bar{\lambda}_1)^{N_1(t)}\,\mathbb{I}_{\{X_t \neq X_t^n\}}\big] &\leq \mathbb{E}^{\mathbb{Q}}\big[\,(\bar{\lambda}_1)^{N_1(t)}\,\mathbb{I}_{\{S \leq t\}}\big] \\
&= \mathbb{E}^{\mathbb{Q}}\big[\,(\bar{\lambda}_1)^{N_1(t)}\,\mathbb{I}_{\{X_0 \neq X_0^n\}}\big] + \sum_{k=1}^{\infty}\mathbb{E}^{\mathbb{Q}}\Big[\mathbb{I}_{\{N(t)=k\}}\sum_{h=1}^{k}(\bar{\lambda}_1)^{\hat{Y}_t - y}\,\mathbb{I}_{\{S=T_h\}}\Big] \\
&= \mathbb{E}^{\mathbb{Q}}\big[\,(\bar{\lambda}_1)^{N_1(t)}\big]\mathbb{Q}(X_0 \neq X_0^n) + \sum_{k=1}^{\infty}\sum_{h=1}^{k}\mathbb{E}^{\mathbb{Q}}\big[\mathbb{I}_{\{N(t)=k\}}\,(\bar{\lambda}_1)^{\hat{y}_k - y}\,\mathbb{I}_{\{\eta_h \neq \eta_h^n,\,\eta_l = \eta_l^n,\,l \leq h-1\}}\big]. \quad (84)
\end{aligned}
$$

The first addend is equal to the first addend in inequality (83), since $\mathbb{Q}(X_0 \neq X_0^n) = \|\mu_0^X - \mu_0^{X^n}\|_{TV}/2$. Moreover, for any k and $1 \leq h \leq k$, we have

$$\mathbb{E}^{\mathbb{Q}}\big[\mathbb{I}_{\{N(t)=k\}}\,(\bar{\lambda}_1)^{\hat{y}_k - y}\,\mathbb{I}_{\{\eta_h \neq \eta_h^n,\,\eta_l = \eta_l^n,\,l \leq h-1\}}\big] = \mathbb{Q}(N(t)=k)\,\mathbb{E}^{\mathbb{Q}}\big[\,(\bar{\lambda}_1)^{\hat{y}_k - y}\,\mathbb{I}_{\{\eta_h \neq \eta_h^n,\,\eta_l = \eta_l^n,\,l \leq h-1\}}\big],$$

since under $\mathbb{Q}$ the Poisson process $N(t)$ and the Markov chain $(\eta_k, \eta_k^n, \hat{y}_k)$ are independent.

Therefore, according to Lemma 4.12 below, the second addend in (84) is bounded above by

$$
\begin{aligned}
\sum_{k=1}^{\infty}\mathbb{Q}(N(t)=k)\,k\,\mathbb{E}^{\mathbb{Q}}\Big[\,(\bar{\lambda}_1)^{\hat{y}_k - y}\Big]\,J^n & \\
= \sum_{k=1}^{\infty}\mathbb{E}^{\mathbb{Q}}\Big[\mathbb{I}_{\{N(t)=k\}}\,k\,(\bar{\lambda}_1)^{\hat{y}_k - y}\Big]\,J^n &= \mathbb{E}^{\mathbb{Q}}\Big[N(t)\,(\bar{\lambda}_1)^{N_1(t)}\Big]\,J^n,
\end{aligned}
$$

and this ends the proof of inequality (83).

Lemma 4.12. *For any $k \geq 1$ and $1 \leq h \leq k$,*

$$\mathbb{E}^{\mathbb{Q}}\big[\,(\bar{\lambda}_1)^{\hat{y}_k - y}\,\mathbb{I}_{\{\eta_h \neq \eta_h^n,\,\eta_l = \eta_l^n,\,l \leq h-1\}}\big] \leq \mathbb{E}^{\mathbb{Q}}\Big[\,(\bar{\lambda}_1)^{\hat{y}_k - y}\Big]\,J^n. \quad (85)$$

To prove this inequality we need the following technical result.

Lemma 4.13. *Under the hypotheses and notations of Proposition 4.6, the following upper bound holds for any $a > 0$*

$$\mathbb{E}^{\mathbb{Q}}\big[a^{\hat{y}_{k+1}-\hat{y}_k}\mathbb{I}_{\{\eta_{k+1}\neq\eta^n_{k+1}\}}|\eta_k = z, \eta^n_k = z, \hat{y}_k = w\big]$$

$$= a^1\,\frac{1}{1+\bar{\lambda}_0}\|\mu_1(z,w,\cdot)-\mu^n_1(z,w,\cdot)\|_{TV}/2 \tag{86}$$

$$+ a^0\,\frac{\bar{\lambda}_0}{1+\bar{\lambda}_0}\|p_0(z,w,\cdot)-p^n_0(z,w,\cdot)\|_{TV}/2$$

$$\leq \mathbb{E}^{\mathbb{Q}}\big[a^{\hat{y}_{k+1}-\hat{y}_k}\big]\max(\|p_0-p^n_0\|_\infty, \|\mu_1-\mu^n_1\|_\infty)/2. \tag{87}$$

Proof. Without loss of generality we can assume that $\mathcal{N}_0 = \{(T^0_k, \zeta^0_k), k \geq 1\}$ is the restriction to $(0,\infty)\times(0,\bar{\lambda}_0]$ of a Poisson measure $\mathcal{N} = \{(T_k,\zeta_k), k \geq 1\}$ on $(0,\infty)\times(-1,\bar{\lambda}_0]$. Moreover, if $\overline{\mathcal{N}}_1 = \{(\bar{T}^1_k,\bar{\zeta}^1_k), k \geq 1\}$ is the restriction of $\mathcal{N}$ to $(0,\infty)\times(-1,0]$, we can assume that $\mathcal{N}_1 = \{(T^1_k,\zeta^1_k), k \geq 1\}$ is obtained by taking $(T^1_k,\zeta^1_k) = (\bar{T}^1_k, -\bar{\zeta}^1_k)$.

We recall that under $\mathbb{Q}$ the processes

$$\big(X_t, X^n_t, \hat{Y}_t\big) = \big(X_t, X^n_t, y + N_1(t)\big)$$

are constructed by defining K_0, K^n_0, K_1 and K^n_1 as explained in Definition 3.8 of [8], that is K_0 and K^n_0 are defined by means of Lemma 3.7 of [8] using the measures $p_0(x, y; dx')$ and $p^n_0(x, y; dx')$ defined by (63) and (64) respectively, while K_1 and K^n_1 are defined using the measures $\mu_1(x, y; dz)$ and $\mu^n_1(x, y; dz)$.

Equality (86) can be derived as follows

$$\mathbb{E}^{\mathbb{Q}}\big[a^{\hat{y}_{k+1}-\hat{y}_k}\mathbb{I}_{\{\eta_{k+1}\neq\eta^n_{k+1}\}}|\ \eta_k = z, \eta^n_k = z, \hat{y}_k = w\big]$$

$$= \mathbb{E}^{\mathbb{Q}}\big[a^1\,\mathbb{I}_{\{\zeta_{k+1}\in(-1,0]\}}\,\mathbb{I}_{\{K_1(z,z,w;-\zeta_{k+1})\neq K^n_1(z,z,w;-\zeta_{k+1})\}}\big]$$

$$+ \mathbb{E}^{\mathbb{Q}}\big[a^0\,\mathbb{I}_{\{\zeta_{k+1}\in(0,\bar{\lambda}_0]\}}\,\mathbb{I}_{\{K_0(z,z,w;\zeta_{k+1})\neq K^n_0(z,z,w;\zeta_{k+1})\}}\big]$$

$$= a^1\mathbb{Q}(\zeta_{k+1}\in(-1,0], K_1(z,z,w;-\zeta_{k+1})\neq K^n_1(z,z,w;-\zeta_{k+1}))$$

$$+ a^0\mathbb{Q}(\zeta_{k+1}\in(0,\bar{\lambda}_0], K_0(z,z,w;\zeta_{k+1})\neq K^n_0(z,z,w;\zeta_{k+1}))$$

$$= a^1\,\frac{1}{1+\bar{\lambda}_0}\int_0^1 \mathbb{I}_{\{u:\,K_1(z,z,w;u)\neq K^n_1(z,z,w;u)\}}(v)\,dv$$

$$+ a^0\,\frac{\bar{\lambda}_0}{1+\bar{\lambda}_0}\int_0^{\bar{\lambda}_0}\mathbb{I}_{\{u:\,K_0(z,z,w;u)\neq K^n_0(z,z,w;u)\}}(v)\,\frac{dv}{\bar{\lambda}_0}$$

$$= a^1\,\frac{1}{1+\bar{\lambda}_0}\|\mu_1(z,w,\cdot)-\mu^n_1(z,w,\cdot)\|_{TV}/2 + a^0\,\frac{\bar{\lambda}_0}{1+\bar{\lambda}_0}\|p_0(z,w,\cdot)-p^n_0(z,w,\cdot)\|_{TV}/2,$$

where the last equality is due to the specific choice of K_0, K^n_0, K_1 and K^n_1. Observing that

$$a^1\,\frac{1}{1+\bar{\lambda}_0} + a^0\,\frac{\bar{\lambda}_0}{1+\bar{\lambda}_0} = \mathbb{E}^{\mathbb{Q}}\big[a^{\hat{y}_{k+1}-\hat{y}_k}\big],$$

inequality (87) is an immediate consequence of (86). $\qquad\square$

Proof. **(of Lemma 4.12)**
We observe that

$$\mathbb{E}^{\mathbb{Q}}\Big[\big(\bar{\lambda}_1\big)^{\hat{y}_k-y}\mathbb{I}_{\{\eta_h\neq\eta^n_h,\ \eta_l=\eta^n_l,\ l\leq h-1\}}\Big]$$

$$= \mathbb{E}^{\mathbb{Q}}\Big[\big(\bar{\lambda}_1\big)^{\hat{y}_h-y}\ \mathbb{I}_{\{\eta_h\neq\eta^n_h,\ \eta_l=\eta^n_l,\ l\leq h-1\}}\ \mathbb{E}^{\mathbb{Q}}\big[\big(\bar{\lambda}_1\big)^{\hat{y}_k-\hat{y}_h}\ |\mathcal{F}^{\eta,\eta^n,\hat{y}}_h\big]\Big]$$

$$= \mathbb{E}^{\mathbb{Q}}\Big[\big(\bar{\lambda}_1\big)^{\hat{y}_h-y}\ \mathbb{I}_{\{\eta_h\neq\eta^n_h,\ \eta_l=\eta^n_l,\ l\leq h-1\}}\Big]\mathbb{E}^{\mathbb{Q}}\big[\big(\bar{\lambda}_1\big)^{\hat{y}_k-\hat{y}_h}\ \big]$$

$$\leq \mathbb{E}^{\mathbb{Q}}\Big[\big(\bar{\lambda}_1\big)^{\hat{y}_h-y}\ \mathbb{I}_{\{\eta_h\neq\eta^n_h,\ \eta_{h-1}=\eta^n_{h-1}\}}\Big]\mathbb{E}^{\mathbb{Q}}\big[\big(\bar{\lambda}_1\big)^{\hat{y}_k-\hat{y}_h}\ \big].$$

Then (85) follows taking into account that

$$\mathbb{E}^{\mathbb{Q}}\Big[\big(\bar{\lambda}_1\big)^{\hat{y}_h - y} \, \mathbb{I}_{\{\eta_h \neq \eta_h^n, \, \eta_{h-1} = \eta_{h-1}^n\}} \Big]$$

$$= \mathbb{E}^{\mathbb{Q}}\Big[\mathbb{I}_{\{\eta_{h-1} = \eta_{h-1}^n\}} \, \big(\bar{\lambda}_1\big)^{\hat{y}_{h-1} - y} \, \mathbb{E}^{\mathbb{Q}}\big[\big(\bar{\lambda}_1\big)^{\hat{y}_h - \hat{y}_{h-1}} \, \mathbb{I}_{\{\eta_h \neq \eta_h^n\}} \, |\mathcal{F}_{h-1}^{\eta, \eta^n, \hat{y}} \big] \Big]$$

$$= \mathbb{E}^{\mathbb{Q}}\Big[\big(\bar{\lambda}_1\big)^{\hat{y}_{h-1} - y} \mathbb{I}_{\{\eta_{h-1} = \eta_{h-1}^n\}} \, \mathbb{E}^{\mathbb{Q}}\big[\big(\bar{\lambda}_1\big)^{\hat{y}_h - \hat{y}_{h-1}} \, \mathbb{I}_{\{\eta_h \neq \eta_h^n\}} \, |\eta_{h-1}, \eta_{h-1}^n, \hat{y}_{h-1} \big] \Big],$$

and that, using inequality (87) of Lemma 4.13, with $a = \bar{\lambda}_1$, the above quantity is bounded above by

$$\mathbb{E}^{\mathbb{Q}}\Big[\big(\bar{\lambda}_1\big)^{\hat{y}_{h-1} - y} \mathbb{I}_{\{\eta_{h-1} = \eta_{h-1}^n\}} \, \mathbb{E}^{\mathbb{Q}}\big[\big(\bar{\lambda}_1\big)^{\hat{y}_h - \hat{y}_{h-1}} \big] \max(\|p_0 - p_0^n\|_\infty, \|\mu_1 - \mu_1^n\|_\infty)/2 \Big]$$

$$\leq \mathbb{E}^{\mathbb{Q}}\Big[\big(\bar{\lambda}_1\big)^{\hat{y}_{h-1} - y} \Big] \mathbb{E}^{\mathbb{Q}}\Big[\big(\bar{\lambda}_1\big)^{\hat{y}_h - \hat{y}_{h-1}} \Big] J^n = \mathbb{E}^{\mathbb{Q}}\Big[\big(\bar{\lambda}_1\big)^{\hat{y}_h - y} \Big] \, J^n,$$

where in the last equality we have used the independence of the increments $\hat{y}_{h-1} - y$ and $\hat{y}_h - \hat{y}_{h-1}$ under $\mathbb{Q}$.

$\square$

References

[1] Bhatt, A. G., Kallianpur, G., and Karandikar, R. L. Robustness of the nonlinear filter. *Stochastic Process. Appl. 81*, 2 (1999), 247–254.

[2] Bhatt, A. G., and Karandikar, R. L. Robustness of the nonlinear filter: the correlated case. *Stochastic Process. Appl. 97*, 1 (2002), 41–58.

[3] Budhiraja, A., and Kushner, H. J. Robustness of nonlinear filters over the infinite time interval. *SIAM J. Control Optim. 36*, 5 (1998), 1618–1637 (electronic).

[4] Calzolari, A., Florchinger, P., and Nappo, G. Nonlinear filtering for Markov systems with delayed observations. preprint, Dip. Mat. - Univ. di Roma La Sapienza - http://www.mat.uniroma1.it/people/nappo/papers.pdf/DelayMarkov.pdf, 2004.

[5] Calzolari, A., Florchinger, P., and Nappo, G. Convergence in nonlinear filtering for stochastic delay systems. new version of the preprint, Dip. Mat. - Univ. di Roma La Sapienza (2003) - http://www.mat.uniroma1.it/people/nappo/papers.pdf/CFN-conv-delay-rate.pdf (submitted), 2005.

[6] Calzolari, A., and Nappo, G. A filtering problem with counting observations: error bounds due to the uncertainty on the infinitesimal parameters. *Stochastics Stochastics Rep. 61*, 1-2 (1997), 1–19.

[7] Calzolari, A., and Nappo, G. Robust approximation in a filtering problem with real state space and counting observations. *Appl. Math. Optim. 42*, 1 (2000), 51–71.

[8] Calzolari, A., and Nappo, G. Counting observations: a note on state estimation sensitivity with an L^1-bound. *Appl. Math. Optim. 44*, 2 (2001), 177–201.

[9] Ceci, C., Gerardi, A., and Tardelli, P. An estimate of the approximation error in the filtering of a discrete jump process. *Math. Models Methods Appl. Sci. 11*, 2 (2001), 181–198.

[10] Chang, M. H. Discrete approximation of nonlinear filtering for stochastic delay equations. *Stochastic Anal. Appl. 5*, 3 (1987), 267–298.

[11] Del Moral, P. Measure-valued processes and interacting particle systems. Application to nonlinear filtering problems. *Ann. Appl. Probab. 8*, 2 (1998), 438–495.

[12] Di Masi, G. B., Pratelli, M., and Runggaldier, W. J. An approximation for the nonlinear filtering problem, with error bound. *Stochastics 14*, 4 (1985), 247–271.

[13] Elliott, R. J., and Glowinski, R. Approximations to solutions of the Zakai filtering equation. *Stochastic Anal. Appl. 7*, 2 (1989), 145–168.

[14] Kirch, M., and Runggaldier, W. J. Efficient hedging when asset prices follow a geometric Poisson process with unknown intensities. *SIAM J. Control Optim. 43*, 4 (2004/05), 1174–1195 (electronic).

[15] Kliemann, W., Koch, G., and Marchetti, F. On the unnormalized solution of the filtering problem with counting observations. *IEEE Trans. Inform. Theory 316*, 6 (1990), 1415–1425.

[16] Kurtz, T. G., and Ocone, D. L. Unique characterization of conditional distributions in nonlinear filtering. *Ann. Probab. 16*, 1 (1988), 80–107.

[17] Kushner, H. J. *Probability methods for approximations in stochastic control and for elliptic equations*. Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1977. Mathematics in Science and Engineering, Vol. 129.

[18] Le Gland, F. Time discretization of nonlinear filtering equations. In *Proceedings of the 28th IEEE Conference on Decision and Control, Vol. 1–3 (Tampa, FL, 1989)* (New York, 1989), IEEE, pp. 2601–2606.

[19] Lototsky, S., Mikulevicius, R., and Rozovskii, B. L. Nonlinear filtering revisited: a spectral approach. *SIAM J. Control Optim. 35*, 2 (1997), 435–461.

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

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

[22] Twardowska, K., Marnik, T., and Pasławaska-Połuniak, M. Approximation of the Zakai equation in a nonlinear filtering problem with delay. *Int. J. Appl. Math. Comput. Sci. 13*, 2 (2003), 151–160.