Asymptotics of Ensemble Filters for Linear Stochastic Systems with Poisson-Sampled Observations⋆

Olga Yufereva[a], Aneel Tanwani[b]

[a]*Krasovskii Institute of Mathematics and Mechanics, Yekaterinburg, Russian Federation.*
Email: `olga.o.yufereva@gmail.com`

[b]*LAAS – CNRS, University of Toulouse, CNRS, Toulouse France.*
Email: `aneel.tanwani@cnrs.fr`

Abstract

For continuous-time linear stochastic dynamical systems driven by Wiener processes, we consider the problem of designing ensemble filters when the observation process is randomly time-sampled. For the design of ensemble filters, we consider a class of continuous-discrete diffusion processes with additive Gaussian noise and several design parameters, which are used to describe the evolution of the individual particles in the ensemble. These particles are coupled through the empirical covariance, and in some cases empirical mean as well, and require less computations for implementation than the optimal ones based on solving Riccati differential equations. For different choices of parameters, we can recover some common design techniques from the literature. Our focus in this work is on analyzing the asymptotic (in time) performance of these filters for sufficiently large number of particles. Using appropriate analysis tools, we derive differential equations to describe the expectation of empirical mean and sample covariance of the ensemble filters with respect to the sampling process and noise. The solutions of these differential equations (describing empirical moments) are shown to converge asymptotically to the mean and covariance of the optimal filter under certain conditions on the mean sampling rate of the observation process, and as the number of particles tends to infinity.

Keywords: Sub-optimal filtering; McKean–Vlasov type equation; Ensemble filters; Stochastic analysis; Random observations.

1. Introduction

The filtering problem relates to finding the hidden state of a stochastic process using some incomplete and noisy observations. For dynamical systems described by stochastic differential equations, the state evolves as a continuous random variable and the problem of inferring the state (which is optimal with respect to certain metric) using the partial noisy observations is formulated as computing the posterior distribution of the state process conditioned upon the measured observations. This problem has found relevance in several disciplines across mathematics and engineering, which has led to a variety of approaches to computing such posteriors, or their estimates. Some of these developments can be traced in the compilation [6].

Among the existing techniques for filtering, the use of Monte Carlo integration methods for approximating the optimal distribution has gained significant interest in the literature [10, 21, 22]. The basic idea of this technique is to draw samples from certain distributions of lower complexity and use the corresponding empirical measure to approximate of the posterior. In the same spirit, [11] introduced the technique of ensemble Kalman filters to develop filtering methods for large-scale applications related to geophysical sciences, so that the posterior is approximated from a *collection* of state estimators. Since

⋆The study of O. Yufereva has been funded by the Russian Science Foundation (project No. 24-11-00123). The work of A. Tanwani was partly supported by the project CyPhAI, financed by ANR-JST CREST program with grant number ANR-20-JSTM-0001.

then, the use of ensemble Kalman filters has had a notable impact in areas where estimation with noisy data is required in large-scale models [20] because the approximation is obtained by solving several lower-order equations rather than solving high-dimensional differential equations, which include the nonlinear ones (Riccati). From a technical viewpoint, the particle filtering algorithms require us to sample from a steady-state distribution and these samples are then used to construct an empirical approximation. In ensemble filtering algorithms, however, the question of choosing a distribution (from which to sample) is addressed by simulating the evolution of individual particles in the ensemble through differential equations that are coupled through the associated empirical mean and the empirical error covariance. Several review articles [5, 4] and the books [12] provide an overview of developments in that area. In most of these works, we do not find much details about the theoretical analysis of the proposed filtering techniques, and this area of mathematical analysis of the ensemble Kalman filters has gathered attention only very recently [8, 2, 3].

Over the past decade, ensemble filtering methods have been viewed from the lens of mean-field models described by stochastic differential equations, and the ensemble particles are simply the approximations of these mean-field models. Recent review articles which elaborate on this viewpoint are [2, 26]. In fact, the limiting behavior of these particles is described by a McKean–Vlasov type diffusion process, which is also referred to as the mean-field process. In the literature, this mean-field process is chosen in different ways, e.g., by adding noise in the prediction term and the correction term of the Kalman–Bucy process [9], or as a non-diffusion equation that is optimal in the measure transportation sense [27].

Another important research direction in the area of filtering is to study the problem with constraints on the information available for computing the optimal distribution. In particular, for implementation of filters subject to observations transmitted through some communication protocols, it is natural to stipulate that the observations arrive at some random time instants [23]. It is of interest to compute the conditional distribution of the state process conditioned upon this discrete observation process [15], and this results in continuous-discrete filters. In some recent work, [16] considers a continuous-time stochastic nonlinear system with discrete-time observations and presents an algorithm similar to the Kalman filter but using variational inference to approximate the conditional distribution using Gaussian approximation. For certain technical reasons and to better study the effect of mean sampling rate, we stipulate in our previous works that the sampling process is a Poisson counter. In particular, for a system class very close to the one studied in this paper, the authors have proposed a continuous-discrete filter in [29], where they analyze the boundedness of error covariance as a function of the mean sampling rate.

Our primary objective in this article is to develop and analyze ensemble filters for continuous-time stochastic processes subject to randomly time-sampled observations. In the literature, we find some variants of *continuous-discrete feedback particle filters* in different settings. The paper [1] provides one (and possibly the first) such example, where the authors use mollifiers in the particle equations to smoothen the dynamics, but no statements about the limiting process are provided. The paper [31] develops particle filters for nonlinear systems using the time-discretization procedure as a part of the derivation and studies convergence as the length of the sampling interval converges to zero. Recent conference papers by the authors [32, 30] provide a preliminary study of continuous-discrete counterparts of ensemble filters with randomly sampled observations. The paper [32] provides an adaptation of transport-inspired filter where the empirical moments exactly coincide with the optimal moments, whereas [30] is based on vanilla ensemble filters, which provides the first instance of asymptotic convergence but the results are limited to the scalar case. The thesis [14] merges these results and provides a framework for analyzing filtering problems with Poisson-sampled observations.

In this paper, we develop generic models for studying continuous-discrete ensemble filters with randomly sampled observations with analytical results about the performance of the proposed filters. For our purposes, the state process is modeled by linear continuous-time Ornstein–Uhlenbeck process and the sampling process for the observations is a Poisson counter. We develop ensemble filters which update their estimate whenever the Poisson counter increments due to the arrival of a new measurement from the observation process. In contrast to [32] and [30], our objective here is to study different types of ensemble filters which can be broadly categorized into four categories. The proposed ensemble filters have the common structure that each one of them contains a prediction part and a correction part. However, the four different mean-field models that we propose to describe these ensemble filters either contain no noise, or if they do, then the noise could be in the prediction part, the correction part, or in both of them. The presence of noise is motivated by the fact that it leads to positive definite (and hence invertible) error covariance matrix that is important for good performance of the ensemble particles, but at the same time, it is possible to consider designs which do not necessarily rely on injection of the noise to make the error covariance positive definite. We will explore these trade-offs by proposing different ensemble filters in our work.

From the analysis point of view, our goal is to provide a common approach to describe the performance of the proposed filtering algorithms. As a first result, we provide quantitative estimates on the difference between the empirical covariance of the ensemble filter and the optimal error covariance of the Kalman filter. These estimates are presented in a form so that it is easy to understand the effect of noise terms injected in the filtering algorithms. In essence, the results show that for appropriate sampling rate, if the number of the particles in the ensemble is large enough then the empirical covariance converges to the optimal covariance. A similar result is developed for the empirical mean, which is shown to converge to optimal mean asymptotically in time.

2. Overview and Problem setup

Let us begin with the description of the system class and the formulation of the basic filtering problem.

2.1. System Class

We consider dynamical systems modeled by linear stochastic differential equations of the form

$$\mathrm{d}X_t = AX_t\,\mathrm{d}t + B\,\mathrm{d}\omega_t \tag{1}$$

where $(X_t)_{t\geq 0}$ is an $\mathbb{R}^n$-valued diffusion process describing the state. Let $(\Omega, \mathcal{F}, \mathsf{P})$ denote the underlying probability space. It is assumed that, for each $t \geq 0$, $(\omega_t)_{t\geq 0}$ is a zero mean $\mathbb{R}^m$-valued standard Wiener process with the property that $\mathsf{E}[\mathrm{d}\omega_t\,\mathrm{d}\omega_t^\top] = I_m\mathrm{d}t$, for each $t \geq 0$. The matrices $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$ are taken as constant. The initial condition X_0 is assumed to be Gaussian and the process $(\omega_t)_{t\geq 0}$ does not depend on the state. In this setting, the solution of (1) is a stochastic process $(X_t)_{t\geq 0}$ that is adapted to the filtration generated by the initial condition and the process $(\omega_t)_{t\geq 0}$, see [18, Theorem 5.2.1].

2.2. Measurement process

Our goal is to study the state estimation problem when the output measurements are available only at random times. The motivation to work with randomly time-sampled measurements comes from several applications, such as communication over networks which allow information packets to be sent at some discrete randomly distributed time instants. Thus, we consider a nondecreasing monotone sequence $(\tau_k)_{k\in\mathbb{N}}$ taking values in $\mathbb{R}_{\geq 0}$ which denote the time instants at which observations are available for measurement. We introduce the process N_t defined as, $N_0 := 0$ and

$$N_t := \sup\big\{k \in \mathbb{N} \,\big|\, \tau_k \leq t\big\} \quad \text{for } t > 0, \tag{2}$$

and it is assumed that $(N_t)_{t\geq 0}$ is a Poisson process of intensity $\lambda > 0$ and it is independent of the noise and the state processes. Recall [24, Theorem 2.3.2] that the Poisson process of intensity $\lambda > 0$ is a continuous-time random process $\big(N_t\big)_{t\geq 0}$ taking values in $\mathbb{N}^* := \mathbb{N} \cup \{0\}$, with $N_0 = 0$, for every $n \in \mathbb{N}^*$ and $0 =: t_0 < t_1 < \cdots < t_n < +\infty$, the increments $\{N_{t_k} - N_{t_{k-1}}\}_{k=1}^n$ are independent, and $N_{t_k} - N_{t_{k-1}}$ is distributed as a Poisson-$\lambda(t_k - t_{k-1})$ random variable for each k. The Poisson process is among the most well studied processes, and standard results (see, e.g., [24, §2.3]) show that it is memoryless and Markovian.

The discretized and noisy observation process under consideration is defined as

$$y_{\tau_{N_t}} = CX_{\tau_{N_t}} + \nu_{N_t}, \qquad t \geq 0, \tag{3}$$

where $C \in \mathbb{R}^{p\times n}$ is a constant matrix, and ν_k is a sequence of i.i.d. Gaussian noise processes and $\nu_0 \sim \mathcal{N}(0, V)$. Equation (3) is motivated by the fact that a continuous observation process $\mathrm{d}z_t = CX_t\mathrm{d}t + dv_t$ with a Wiener process $(v_t)_{t\geq 0}$ is formally equivalent to $y_t = CX_t + \nu_t$, for any $t \geq 0$, with the identifications $y_t \sim \frac{\mathrm{d}z_t}{\mathrm{d}t}$ and $\nu_t \sim \frac{\mathrm{d}v_t}{\mathrm{d}t}$, see [15, Chapter 4] for further details. Our goal is to construct the estimate $\widehat{X}_t$, which minimizes the mean-square estimation error, using the observations $\mathcal{Y}_t := \{y_{\tau_k} \,|\, k \leq N_t\}$.

2.3. Optimal filter

The basic problem in filter design is to find an estimate of the state process which minimizes the mean-square estimation error, and is described by the expectation of the state process $(X_t)_{t\geq 0}$ conditioned upon the measurements observed over the interval $[0,t]$, that is, $\mathcal{Y}_t$. In particular, with the structure imposed on the system dynamics in this section, the conditional expectation is Gaussian and the two moments are simulated through ordinary differential equations with updates at times when a new measurement arrives. For an arbitrary strictly increasing real-valued sequence $(\tau_k)_{k\in\mathbb{N}^*}$, this procedure is also proposed in [15, Thm. 7.1]. If we specify a sequence $(\tau_k)_{k\in\mathbb{N}^*}$ so that it corresponds to the arrival times of a Poisson process, we simulate the mean of the conditional distribution as:

$$\dot{\widehat{X}}_t = A\widehat{X}_t \mathrm{d}t, \quad t \in [\tau_{N_t}, \tau_{1+N_t}[\tag{4a}$$

$$\widehat{X}_t^+ = \widehat{X}_t + L_t^{\mathrm{opt}}(y_t - C\widehat{X}_t), \quad t = \tau_{N_t} \tag{4b}$$

where the injection gain $L_t^{\mathrm{opt}} = P_t C^\top (CP_tC^\top + V)^{-1}$, and the error covariance process $(P_t)_{t\geq 0}$ is described as

$$\dot{P}_t = (AP_t + P_tA^\top + GG^\top), \quad t \in [\tau_{N_t}, \tau_{1+N_t}[\tag{5a}$$

$$P_t^+ = P_t - P_tC^\top(CP_tC^\top + V)^{-1}CP_t, \quad t = \tau_{N_t}. \tag{5b}$$

To make the presentation compact later on, we adopt the formalism of writing the continuous-discrete equations (4a) and (4b) together in a single differential equation, when the jumps are driven by a Poisson counter N_t :

$$\mathrm{d}\widehat{X}_t = A\widehat{X}_t\mathrm{d}t + L_t^{\mathrm{opt}}(y_t - C\widehat{X}_t)\mathrm{d}N_t. \tag{6}$$

Similarly, using this formalism, equations (5a) and (5b) can be written in a combined form as,

$$\mathrm{d}P_t = (AP_t + P_tA^\top + BB^\top)\mathrm{d}t - P_tC^\top(CP_tC^\top + V)^{-1}CP_t\,\mathrm{d}N_t. \tag{7}$$

Remark 2.1. In this paper, we will often consider differential equations driven by $\mathrm{d}N_t$, where N_t is a Poisson counter of given intensity $\lambda > 0$. Taking (6) as an example of such differential equation, we see that its solution is precisely given by (4). In particular, for the external processes that get multiplied by $\mathrm{d}N_t$ ($(L_t^{\mathrm{opt}})_{t\geq 0}$ and $(y_t)_{t\geq 0}$ in (6)), it suffices to know their value at $t = \tau_{N_t}$ to define the solution of these equations.

In the foregoing discussion, one makes the observation that the optimal conditional distribution, that is, $\mathrm{Law}(X_t|\mathcal{Y}_t)$ is Gaussian for each realization of $(N_t)_{t\geq 0}$ despite the fact that the mean and covariance are discontinuous along each sample path.

From analysis viewpoint, it is important to look at the expectation of the process $(P_t)_{t\geq 0}$ with respect to the sampling times $(\tau_{N_t})_{t\geq 0}$. In our previous work [29], we showed that the expectation of piecewise deterministic process P_t, denoted by $\mathcal{P}_t$, is described by the following differential equation:

$$\dot{\mathcal{P}}_t = A\mathcal{P}_t + \mathcal{P}_tA^\top + BB^\top - \lambda\mathcal{P}_tC^\top(C\mathcal{P}_tC^\top + V)^{-1}C\mathcal{P}_t. \tag{8}$$

Moreover, in [29] and [7], we provide conditions in terms of the bounds on the mean sampling rate $\lambda > 0$ and the structural assumptions on controllability and observability of the pairs (A, B) and (A, C) that guarantee boundedness of $\mathcal{P}_t$, and show convergence of $\mathcal{P}_t$ towards the steady state. The boundedness is also important for asymptotic analysis of the first moment of the error process $(X_t - \widehat{X}_t)_{t\geq 0}$.

2.4. Ensemble filtering approach

In the previous section, we saw that the implementation of conventional optimal filter, even for linear systems, is computationally heavy since it requires simulating a Riccati differential equation to compute the injection gains. For a state process evolving in $\mathbb{R}^n$, this optimal filter involves solving $(n^2+n)/2$ differential equations for the covariance process $(P_t)_{t\geq 0}$ and n differential equations for the first moment of the estimate $(\widehat{X}_t)_{t\geq 0}$. This can be quite cumbersome for large values of n especially when we take into consideration the additional operations involved in computing these processes. For this reason, there has been extensive research for other methods to implement optimal filter, or its approximation. One such technique is based on the use of the *ensemble*, or feedback particle, filters. The basic idea of the particle filters is to simulate a collection of particles through stochastic differential equations which are coupled to each other through joint statistics of the population.

The first step in computing an approximation, via *ensemble filtering* technique, is to find a process S_t with certain properties that allows for good approximation and at the same time, it implicitly describes the optimal filter. In particular, for a filtering system with the hidden state process $(X_t)_{t\geq 0}$ and the observation process $(y_{\tau_{N_t}})_{t\geq 0}$, a stochastic process $(S_t)_{t\geq 0}$ satisfying

$$\mathrm{Law}(S_t \mid \mathcal{Y}_t) = \mathrm{Law}(X_t \mid \mathcal{Y}_t) \qquad \forall t \geq 0 \tag{9}$$

is called an *exact filter*. Different processes might play the role of exact filters. It turns out that there are several ways to construct exact filters that can be approximated by computation-friendly methods and the resulting approximations provide sufficiently good performance when compared to the exact solution. For our purposes, such approximations are provided by an ensemble of particles whose limiting behavior (as the number of particles tend to infinity) converges to the exact filter. The computational advantage of this approach is that the simulation of particles, in general, is more efficient and potentially applicable in nonlinear systems as well (although there are very few instances of formal analysis in nonlinear setting). As an example of approximating a process using the particles, one can take for instance a result from [25, Theorem 1.4, p.172] where we consider a process S_t that satisfies a simple version of McKean–Vlasov type equation

$$\mathrm{d}S_t = \mathrm{d}\omega_t + \Big(\int b(S_t, \bar{S})\mathfrak{m}_t(\mathrm{d}\bar{S})\Big)\mathrm{d}t. \tag{10}$$

In (10), $\mathfrak{m}_t(\mathrm{d}\bar{S})$ is the law of S_t, ω_t is a standard Brownian motion, the function b is bounded and Lipschitz continuous and initial distribution is given. The process S_t can be considered as a *mean-field process* as it can be approximated by an ensemble of M interacting particles S_t^i, $i = 1, \ldots, M$, with the dynamics

$$\mathrm{d}S_t^i = \mathrm{d}\omega_t^i + \frac{1}{M}\sum_{j=1}^{M} b(S_t^i, S_t^j)\mathrm{d}t, \quad i = 1, \ldots, M,$$

where ω_t^i is an independent copy of the process ω_t appearing in (10), and the initial condition S_0^i is the same as S_0. In particular, when the number of particles M tends to ∞, each S_t^i approaches a process which is an independent copy of the process S_t.

The above discussion shows the underlying principle of ensemble filters that relies on approximating the solution of a McKean–Vlasov differential equation using particles that are driven by stochastic differential equations coupled together by empirical mean of all the particles in the population. To design ensemble filters, we also look for different types McKean–Vlasov type equations which serve as exact filters, and then the ensemble filters are proposed to approximate these exact filters. For the setting described above, the basic problem studied in this article is to design ensemble filters for the continuous-time system (1) with the discrete observation process (3). The main steps involved in doing so are the following:

- Find the process(es) S_t such that $\mathsf{E}(S_t \mid \mathcal{Y}_t) \sim \mathsf{E}(X_t \mid \mathcal{Y}_t)$.
- Describe the ensemble of particles S_t^i, $i = 1, \ldots, M$ coupled to each other via empirical mean and empirical variance, such that, each S_t^i represents an independent copy of S_t when $M \to \infty$.
- Show that the empirical covariance and the empirical mean of the particles is consistent with the optimal solution to the filtering problem.

3. Mean-Field Model

As we discussed in the previous section, the design of ensemble filters is based on approximating an exact filter which provides the optimal performance. This exact filter serves as the mean-field process that we approximate later using a collection of particles. For the filtering problem with continuous state process and discrete observations, we propose a class of exact filters in this section. Generically, our exact filter $(S_t)_{t\geq 0}$ is a continuous-discrete process with several parameters $\mathbf{D}_t, \mathbf{E}_t, \mathbf{F}_t, \mathbf{G}_t, \mathbf{H}_t, \mathbf{J}_t$, which are càdlàg $\mathcal{Y}_t$-measurable processes. The equation describing this exact filter is defined as follows:

$$\mathrm{d}S_t := AS_t\mathrm{d}t + \mathbf{D}_t(S_t - \widehat{S}_t)\mathrm{d}t + \mathbf{E}_t\mathrm{d}\bar{\omega}_t + \Big[\mathbf{F}_tS_t + \mathbf{G}_t\widehat{S}_t + \mathbf{H}_ty_t + \mathbf{J}_t\bar{\nu}_t\Big]\,\mathrm{d}N_t \tag{11a}$$

$$\widehat{S}_t := \mathsf{E}[S_t \mid \mathcal{Y}_t], \tag{11b}$$

$$Q_t := \mathsf{E}[(S_t - \widehat{S}_t)(S_t - \widehat{S}_t)^\top \mid \mathcal{Y}_t], \tag{11c}$$

where $\bar{\omega}_t$ is an independent copy of ω_t, and $\bar{\nu}_t := \bar{\nu}_{N_t}$ with $\bar{\nu}_{N_t} \sim \mathcal{N}(0, V)$ being an independent copy of ν_{N_t}. Here, $\widehat{S}_t$ denotes the mean of S_t conditioned upon the observations $\mathcal{Y}_t$, and Q_t denotes the corresponding conditional covariance. It is seen that the basic structure of the proposed mean-filed model comprises of two parts: the prediction term and the correction term. The prediction term corresponds to the continuous flow and involves integration due to Lebesgue measure $\mathrm{d}t$ and the noise process $\bar{\omega}$. The correction term corresponds to the jumps in S_t because the measurements are only available at discrete times governed by the Poisson counter N_t. Later in this section, we will see how this generalized model corresponds to different design methods for ensemble filtering in the literature for different values of the parameters.

For the process S_t to be an exact filter, we obviously require that $\widehat{S} = \widehat{X}$, where $\widehat{X}_t$ is the optimal mean obtained from Kalman filter and given in (6). In other words, we want $\widehat{S}_t$ in (11b) and Q_t in (11c) to satisfy the following differential equations:

$$\mathrm{d}\widehat{S}_t = A\widehat{S}_t\mathrm{d}t + Q_tC^\top(CQ_tC^\top + V)^{-1}(y_t - C\widehat{S}_t)\mathrm{d}N_t, \tag{12a}$$

$$\mathrm{d}Q_t = (AQ_t + Q_tA^\top + BB^\top)\mathrm{d}t - Q_tC^\top(CQ_tC^\top + V)^{-1}CQ_t\mathrm{d}N_t. \tag{12b}$$

In the following proposition, we provide the conditions on the parameters in (11a) so that $(S_t)_{t\geq 0}$ is an exact filter.

Proposition 3.1. *Let $\mathbf{D}_t, \mathbf{E}_t, \mathbf{F}_t, \mathbf{G}_t, \mathbf{H}_t, \mathbf{J}_t$ be càdlàg $\mathcal{Y}_t$-measurable processes. The processes $(\widehat{S}_t)_{t\geq 0}$ defined in* (11b) *and $(Q_t)_{t\geq 0}$ defined in* (11c) *satisfy* (12a) *and* (12b)*, respectively, if for all $t > 0$, we have*

$$BB^\top = \mathbf{D}_tQ_t + Q_t\mathbf{D}_t^\top + \mathbf{E}_t\mathbf{E}_t^\top \tag{13a}$$

and for jump times τ_k, $k \in \mathbb{N}^*$, it holds that

$$\mathbf{H}_{\tau_k} = Q_{\tau_k}C^\top(CQ_{\tau_k}C^\top + V)^{-1}, \tag{13b}$$

$$\mathbf{F}_{\tau_k} + \mathbf{G}_{\tau_k} = -Q_{\tau_k}C^\top(CQ_{\tau_k}C^\top + V)^{-1}C, \tag{13c}$$

$$-Q_{\tau_k}C^\top(CQ_{\tau_k}C^\top + V)^{-1}CQ_{\tau_k} = (I + \mathbf{F}_{\tau_k})Q_{\tau_k}(I + \mathbf{F}_{\tau_k})^\top - Q_{\tau_k} + \mathbf{J}_{\tau_k}V\mathbf{J}_{\tau_k}^\top. \tag{13d}$$

Proof. Based on the definition of $(\widehat{S}_t)_{t\geq 0}$ in (11a), we get the following differential equation for the evolution of conditional mean:

$$\mathrm{d}\widehat{S}_t = A\widehat{S}_t\mathrm{d}t + \Big[(\mathbf{F}_t + \mathbf{G}_t)\widehat{S}_t + \mathbf{H}_ty_t\Big]\,\mathrm{d}N_t. \tag{14}$$

Clearly, by choosing the matrices $\mathbf{F}_t$, $\mathbf{G}_t$, and $\mathbf{H}_t$ that satisfy (13b) and (13c), we see that the solution of (14) coincides with the optimal solution given in (12a).

Next, for the error covariance, we consider the process $e_t := S_t - \widehat{S}_t$, which satisfies the equation

$$\mathrm{d}e_t = (A + \mathbf{D}_t)e_t\mathrm{d}t + \mathbf{E}_t\mathrm{d}\bar{\omega}_t + [\mathbf{F}_te_t + \mathbf{J}_t\bar{\nu}_t]\,\mathrm{d}N_t.$$

Denote the i-th entry of the vector e_t by $[e_t]_i$ and the (i,j)-th entry of the matrix $e_te_t^\top$ by $[e_te_t^\top]_{ij}$. Using Ito's chain rule (see Proposition A.1 in Appendix), the last equation then yields

$$\begin{aligned}\mathrm{d}[e_te_t^\top]_{ij} = \mathrm{d}[e_t]_i[e_t]_j &= [e_t]_i[(A + \mathbf{D}_t)e_t\mathrm{d}t + \mathbf{E}_t\mathrm{d}\bar{\omega}_t]_j + [e_t]_j[(A + \mathbf{D}_t)e_t\mathrm{d}t + \mathbf{E}_t\mathrm{d}\omega_t]_i \\ &+ \frac{1}{2}\sum_k[\mathbf{E}_t]_{ik}[\mathbf{E}_t]_{jk}\mathrm{d}t + \Big[\big[e_t + \mathbf{F}_te_t + \mathbf{J}_t\bar{\nu}_t\big]_i\big[e_t + \mathbf{F}_te_t + \mathbf{J}_t\bar{\nu}_t\big]_j - [e_t]_i[e_t]_j\Big]\mathrm{d}N_t.\end{aligned}$$

The evolution equation for the conditional covariance $(Q_t)_{t\geq 0}$ given in (11c) is described by the equation:

$$\begin{aligned}\mathrm{d}Q_t = \mathrm{d}\mathsf{E}[e_t e_t^\top \mid \mathcal{Y}_t] &= (AQ_t + Q_t A^\top + \mathbf{D}_t Q_t + Q_t \mathbf{D}_t^\top + \mathbf{E}_t \mathbf{E}_t^\top)\mathrm{d}t \\ &\quad + \mathsf{E}\Big[\big(e_t + \mathbf{F}_t e_t + \mathbf{J}_t \bar{\nu}_t\big)\big(e_t + \mathbf{F}_t e_t + \mathbf{J}_t \bar{\nu}_t\big)^\top - e_t e_t^\top \mid \mathcal{Y}_t\Big]\mathrm{d}N_t \\ &= (AQ_t + Q_t A^\top + \mathbf{D}_t Q_t + Q_t \mathbf{D}_t^\top + \mathbf{E}_t \mathbf{E}_t^\top)\mathrm{d}t \\ &\quad + \Big[(I + \mathbf{F}_t)Q_t(I + \mathbf{F}_t)^\top - Q_t + \mathbf{J}_t V \mathbf{J}_t^\top\Big]\mathrm{d}N_t.\end{aligned}$$

Using the conditions in (13a) and (13d), it follows that Q_t coincides with the solution (12b). □

3.1. Special Cases

Proposition 3.1 provides generic conditions under which the conditional mean and conditional variance obtained from the process of type (11a) resembles the optimal mean and error variance of the Kalman filter. With this generic result, we can indeed recover a few special cases which closely resemble the mean-field processes for ensemble filtering in continuous-time systems. In what follows, we use the notation:

$$L_t := Q_t C^\top (C Q_t C^\top + V)^{-1},$$

so that L_t describes the injection gain for the innovation term used in the Kalman filter.

(MF-1) Vanilla mean-field process (`Vanilla`) : As the first case of the exact filter $(S_t)_{t\geq 0}$ described in (11), we take $\mathbf{J}_t \neq 0$, $\mathbf{D}_t = 0$, and $\mathbf{G}_t = 0$ so $\mathbf{E}_t = B$, $\mathbf{F}_t = -L_t C$, $H_t = L_t$ and one can take $\mathbf{J}_t = -L_t$

$$\mathrm{d}S_t := AS_t \mathrm{d}t + B\mathrm{d}\bar{\omega}_t + L_t\Big(y_t - CS_t - \bar{\nu}_t\Big)\mathrm{d}N_t. \tag{15}$$

This choice of parameters therefore corresponds to creating a copy of the system dynamics with same noise statistics and then introducing an innovation term which also contains the correction of the noise from the observation process. The resulting exact filter (15) is the foundation for the so-called *vanilla* ensemble filter as proposed initially in [11] for discrete-time systems.

(MF-2) Mean-field process with Noisy Prediction and Deterministic Correction (`NPDC`) : Next, we assume that $\mathbf{J}_t = 0$, $\mathbf{D}_t = 0$, so $\mathbf{E}_t = B$ and

$$\mathrm{d}S_t = AS_t \mathrm{d}t + B\mathrm{d}\bar{\omega}_t + \Big(L_t y_t - L_t C\widehat{S}_t + \mathbf{F}_t(S_t - \widehat{S}_t)\Big)\mathrm{d}N_t, \tag{16a}$$

$$\mathbf{F}_{\tau_k} \text{ satisfies } Q_{\tau_k}\mathbf{F}_{\tau_k}^\top + \mathbf{F}_{\tau_k} Q_{\tau_k} \mathbf{F}_{\tau_k}^\top + \mathbf{F}_{\tau_k} Q_{\tau_k} = -L_{\tau_k} C Q_{\tau_k}. \tag{16b}$$

In contrast to the `Vanilla` process, the exact filter proposed in (16), denoted `NPDC`, does not contain a noise term in the correction due to observations. The noise process $(\bar{\omega}_t)_{t\geq 0}$ only appears in the prediction part, which is a copy of the noise term in the system dynamics. In continuous-time, ensemble filters based on such processes were studied in [3, 2].

(MF-3) Mean-field process with Deterministic Prediction and Noisy Correction (`DPNC`): In contrast to the previous case, we consider the process with $\mathbf{D}_t = \frac{1}{2}BB^\top Q_t^{-1}$, $\mathbf{E}_t = 0$, $\mathbf{J}_t \neq 0$, $\mathbf{F}_t = -L_t C$, $\mathbf{J}_t = -L_t$

$$\mathrm{d}S_t = AS_t \mathrm{d}t + \frac{1}{2}BB^\top Q_t^{-1}(S_t - \widehat{S}_t)\mathrm{d}t + L_t\Big(y_t - CS_t - \bar{\nu}_t\Big)\mathrm{d}N_t. \tag{17}$$

The process (17), denoted `DPNC`, describes an exact filter where the noise term only appears in the correction, just as in `Vanilla` process. However, unlike `Vanilla` case, the prediction term does not contain the noise term and hence it has been modified by adding the term $\frac{1}{2}BB^\top Q_t^{-1}(S_t - \widehat{S}_t)$. This sort of modification, where a noise term of the system dynamics is compensated by a deterministic term depending on conditional mean and covariance, appeared in [27] and is also observed in the exact filters that we propose next in (18).

(MF-4) Deterministic Transport-inspired mean-field process (`DeT`): Lastly, we consider the case where $\mathbf{J}_t = 0$ and $\mathbf{E}_t = 0$ so one can take $\mathbf{D}_t = \frac{1}{2} BB^\top Q_t^{-1}$ and obtain

$$\mathrm{d}S_t = AS_t\mathrm{d}t + \frac{1}{2}BB^\top Q_t^{-1}(S_t - \widehat{S}_t)\mathrm{d}t + \Big(L_t y_t - L_t C\widehat{S}_t + \mathbf{F}_t(S_t - \widehat{S}_t)\Big)\mathrm{d}N_t, \tag{18a}$$

$$\mathbf{F}_{\tau_k} \text{ satisfies } Q_{\tau_k}\mathbf{F}_{\tau_k}^\top + \mathbf{F}_{\tau_k}Q_{\tau_k}\mathbf{F}_{\tau_k}^\top + \mathbf{F}_{\tau_k}Q_{\tau_k} = -L_{\tau_k}CQ_{\tau_k}. \tag{18b}$$

The three types of exact filters presented above contain a noise term either in the prediction part, or in the correction part. In the literature on continuous-time ensemble filters, we find another technique which basically views the evolution of conditional posterior using the optimal transport framework. This eventually leads to a deterministic mean-field process as described in [27]. Our `DPNC` filter is partly inspired by this technique as we introduce a deterministic term in the prediction part to compensate for the process noise. On the other hand, `DeT` process describes an exact filter where we do not inject any noise in the prediction or correction part. The only source of randomness in these filters is due to the presence of the noisy observations (y_t), and the randomness of the jump times due to arrival of the observations.

4. Ensemble Filters

In the previous section, we provided a general expression for mean-field processes that can be simulated as an alternative to Kalman filters to obtain the optimal estimate. Realization of such processes remains a computationally difficult task and, therefore, we look for simpler numerical procedures to approximate the solution of these mean-field processes. One possible way is to do so by simulating an *ensemble* of particles that are driven by differential equations derived from mean-field processes.

More precisely, we consider a collection of M particles, denoted by S^i, $i = 1, \ldots, M$, and each of these particles is a stochastic process described by a solution to a stochastic differential equation. For these particles, we consider the empirical mean $\widehat{S}_t^M$ and the empirical covariance Q_t^M at time t, described by the following relations:

$$\widehat{S}_t^M = \frac{1}{M}\sum_{i=1}^M S_t^i, \tag{19a}$$

$$Q_t^M = \frac{1}{M}\sum_{i=1}^M (S_t^i - \widehat{S}_t^M)(S_t^i - \widehat{S}_t^M)^\top. \tag{19b}$$

Using these definitions, the system of coupled stochastic differential equations used for simulating sample paths of the particles is described as follows:

$$\mathrm{d}S_t^i := AS_t^i\mathrm{d}t + \mathbf{D}_t^M(S_t^i - \widehat{S}_t^M)\mathrm{d}t + \mathbf{E}_t^M\mathrm{d}\omega_t^i + \Big[\mathbf{F}_t^M S_t^i + \mathbf{G}_t^M\widehat{S}_t^M + \mathbf{H}_t^M y_t + \mathbf{J}_t^M\nu_t^i\Big]\mathrm{d}N_t \tag{20}$$

where ω_t^i are independent copies of ω_t and $\nu_t^i := \nu_{N_t}^i$ with $\nu_{N_t}^i \sim \mathcal{N}(0, V)$ representing independent copies of ν_{N_t}, for $i = 1, \cdots, M$. The Poisson process N_t is common for all the equations. The processes $\mathbf{D}_t^M$, $\mathbf{E}_t^M$, $\mathbf{F}_t^M$, $\mathbf{G}_t^M$, $\mathbf{H}_t^M$, and $\mathbf{J}_t^M$ are chosen as a function of the empirical mean $\widehat{S}_t^M$ and empirical covariance Q_t^M. Keeping the result of Proposition 3.1, we choose them so that the empirical counterparts of the conditions (13) hold, that is,

$$BB^\top = \mathbf{D}_t^M Q_t^M + Q_t^M {\mathbf{D}_t^M}^\top + \mathbf{E}_t^M {\mathbf{E}_t^M}^\top, \quad t \geq 0, \tag{21a}$$

and for jump times τ_k, it holds that

$$\mathbf{H}_{\tau_k}^M = Q_{\tau_k}^M C^\top (CQ_{\tau_k}^M C^\top + V)^{-1}, \tag{21b}$$

$$\mathbf{F}_{\tau_k}^M + \mathbf{G}_{\tau_k}^M = -Q_{\tau_k}^M C^\top (CQ_{\tau_k}^M C^\top + V)^{-1} C, \tag{21c}$$

$$-Q_{\tau_k}^M C^\top (CQ_{\tau_k}^M C^\top + V)^{-1} CQ_{\tau_k}^M = (I + \mathbf{F}_{\tau_k}^M)Q_{\tau_k}^M (I + \mathbf{F}_{\tau_k})^\top - Q_{\tau_k}^M + \mathbf{J}_{\tau_k}^M V {\mathbf{J}_{\tau_k}^M}^\top. \tag{21d}$$

It will be shown in the subsequent sections that the ensemble filters provide an approximation of the mean field process, and consequently, the empirical mean and empirical covariance approach the optimal mean and optimal covariance of the posterior distribution obtained from the Kalman filter.

4.1. Special cases

For the specific mean-field processes considered in Section 3, we now consider the corresponding ensemble filters. The basic idea of the ensemble filters is that by imposing the structure similar to a mean-field process, we can obtain an approximation of the exact solution. For simplicity, we introduce the notation L_t^M to denote the empirical gain at time t:

$$L_t^M := Q_t^M C^\top (C Q_t^M C^\top + V)^{-1}. \tag{22}$$

Also, in the four special cases that we present below, ω_t^i are independent copies of ω_t and $\nu_t^i = \nu_{N_t}^i \sim \mathcal{N}(0, V)$ represent independent copies ν_{N_t}, while the Poisson process N_t is common for all the equations.

(EnF-1) Vanilla Ensemble Filter (`Vanilla-EnF`): The particle equations are obtained from the process (15), that is,

$$\mathrm{d}S_t^i = A S_t^i \mathrm{d}t + B \mathrm{d}\omega_t^i + L_t^M \Big(y_t - C S_t^i - \nu_t^i \Big) \mathrm{d}N_t. \tag{23}$$

We see that the difference compared to (15) is that the injection gain L_t^M is now driven by the empirical covariance of the population and one does not need to compute the error covariance exactly anymore.

(EnF-2) Ensemble Filter with Noisy Prediction and Deterministic Correction (`NPDC-EnF`): The particles equations, in this case, are obtained from the `NPDC` process in (16):

$$\mathrm{d}S_t^i := A S_t^i \mathrm{d}t + B \mathrm{d}\omega_t^i + \Big(L_t^M (y_t - C \widehat{S}_t^M) + \mathbf{F}_t^M (S_t - \widehat{S}_t^M) \Big) \mathrm{d}N_t, \tag{24a}$$

$$\mathbf{F}_{\tau_k}^M : \quad Q_{\tau_k}^M (\mathbf{F}_{\tau_k}^M)^\top + \mathbf{F}_{\tau_k}^M Q_{\tau_k}^M (\mathbf{F}_{\tau_k}^M)^\top + \mathbf{F}_{\tau_k}^M Q_{\tau_k}^M = -L_{\tau_k}^M C Q_{\tau_k}^M. \tag{24b}$$

The coupling in the particles in `NPDC-EnF` is not only due to the empirical covariance but also due to the empirical mean.

(EnF-3) Ensemble Filter with Deterministic Prediction and Noisy Correction (`DPNC-EnF`): This is obtained by using `DPNC` process as the base for describing the evolution of the particles, where we once again see the coupling due to empirical moments:

$$\mathrm{d}S_t^i = A S_t^i \mathrm{d}t + \frac{1}{2} B B^\top (Q_t^M)^{-1} (S_t^i - \widehat{S}_t^M) \mathrm{d}t + L_t^M \Big(y_t - C S_t^i - \nu_t^i \Big) \mathrm{d}N_t \tag{25}$$

(EnF-4) Deterministic Transport-inspired Ensemble Filter (`DeT-EnF`) Finally, we consider the ensemble filters obtained from `DeT` mean-field process. The equation used for simulating the particles is,

$$\mathrm{d}S_t^i := A S_t^i \mathrm{d}t + \frac{1}{2} B B^\top (Q_t^M)^{-1} (S_t^i - \widehat{S}_t^M) \mathrm{d}t + \Big(L_t^M (y_t - C \widehat{S}_t^M) + \mathbf{F}_t^M (S_t^i - \widehat{S}_t^M) \Big) \mathrm{d}N_t, \tag{26a}$$

$$\mathbf{F}_{\tau_k}^M : \quad Q_{\tau_k}^M (\mathbf{F}_{\tau_k}^M)^\top + \mathbf{F}_{\tau_k}^M Q_{\tau_k}^M (\mathbf{F}_{\tau_k}^M)^\top + \mathbf{F}_{\tau_k}^M Q_{\tau_k}^M = -L_{\tau_k}^M C Q_{\tau_k}^M. \tag{26b}$$

Once again, we see the coupling among the particles due to the terms depending on empirical mean and covariance. One important thing to note in (26), as well as (25), is that the prediction part contains a term depending on $(Q_t^M)^{-1}$. We recall that Q_t^M is defined as a sum of rank one matrices in (19b). So, we need to simulate a large enough number of particles, at least greater than n (the dimension of the state), to make sure that Q_t^M is invertible.

4.2. Simulation of an academic example

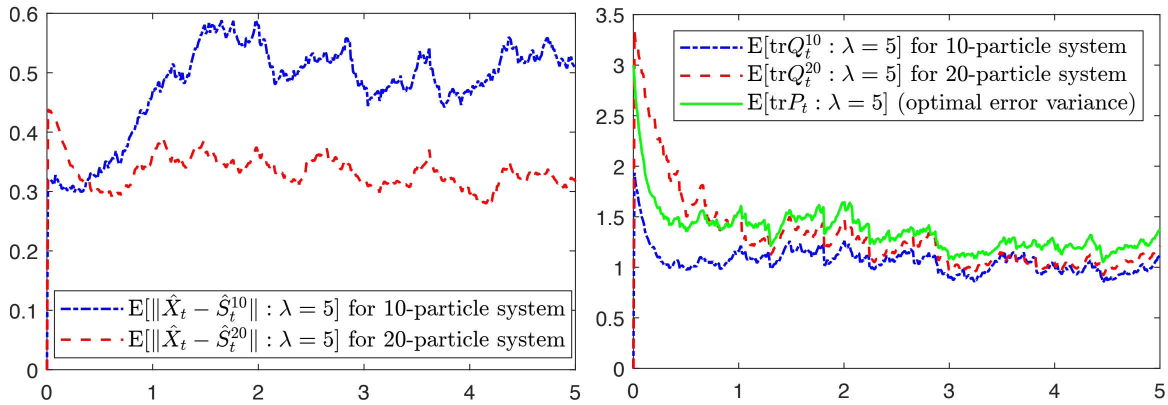
For the purpose of illustration, let us show the simulation results of `Vanilla` ensemble filters given by (23). We consider an academic example where the system is described by the equations:

$$\mathrm{d}x_t = A x_t \mathrm{d}t + B \mathrm{d}\omega_t \tag{27a}$$

$$y_{\tau_k} = C x_{\tau_k} + \nu_k, \tag{27b}$$

with $A = \begin{bmatrix} 0 & 3 & 1 \\ 2 & -2 & 1 \\ -2 & 1 & -3 \end{bmatrix}$, $C = \begin{bmatrix} 1 & -1 & 2 \\ 1 & 0 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 0.5 & 0.5 & 0.5 \end{bmatrix}^\top$, and for each $k \in \mathbb{N}^*$, ν_k is normally distributed with mean $(0,0)^\top$ and the constant variance $V = \begin{bmatrix} 0.5 & 0.1 \\ 0.1 & 0.5 \end{bmatrix}$.

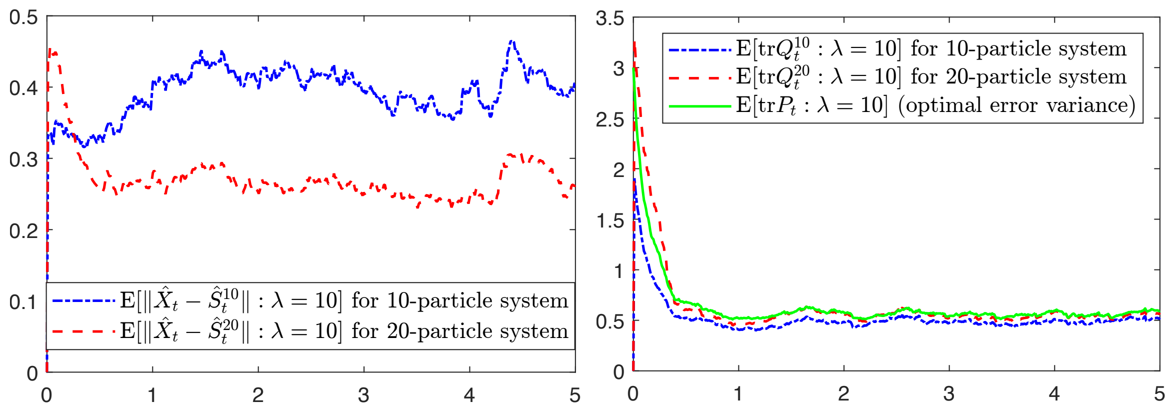
To measure the effectiveness of the ensemble filter, we compare it with the optimal estimator. This is done by looking at, firstly, the difference between the optimal mean $\widehat{X}_t$ from (4) and the empirical mean $\widehat{S}_t^M$ defined in (19a), and secondly, the optimal variance P_t from (5) in comparison with the empirical variance Q_t^M defined in (19b), for different number of particles M and different values of mean sampling rate λ. Figure 1 shows the plots of our simulation over 100 sample paths of the sampling process N_t with intensity $\lambda = 5$, while choosing the number of particles equal to 10 and 20. Figure 1a shows the expectation of $\|\widehat{X}_t - \widehat{S}_t^M\|$ and Fig. 1b plots the expectation of optimal and empirical variances. Figure 2 shows similar quantities, but with the intensity of the sampling process N_t increasing to $\lambda = 10$. The simulation results reported in Figure 1 and Figure 2 are consistent with the analytical results reported in the subsequent sections.



(a) The difference of optimal mean and empirical mean is plotted for 10-particles (blue curve) and 20 particles (red curve).

(b) The plot shows the optimal covariance (green); empirical covariance for 10-particles (blue) and 20 particles (red).

Figure 1: Comparison of mean and covariance for the optimal and `Vanilla-EnF` with mean sampling rate $\lambda = 5$.



(a) The difference of optimal mean and empirical mean is plotted for 10-particles (blue curve) and 20 particles (red curve).

(b) The plot shows the optimal covariance (green); empirical covariance for 10-particles (blue) and 20 particles (red).

Figure 2: Comparison of mean and covariance for the optimal and `Vanilla-EnF` with mean sampling rate $\lambda = 10$.

In the sequel, we will develop a unified approach to analyze the performance of the ensemble filters proposed in this section. It is useful to point out that some preliminary investigation for `Vanilla-EnF` in the scalar case was carried out in our conference paper [30], and we initially proposed the structure of continuous-discrete `DeT-EnF` in [32]. In the case of `DeT-EnF`, we will see that it results in exact optimal variance, but for the other cases, we develop a more general and common approach for asymptotic analysis that allows us to better understand the utility of the proposed filters.

5. Evolution of Empirical Moments

As a first step in studying the performance of the ensemble filters proposed in the previous section, we study the evolution of the empirical moments defined in (19a)–(19b) for the generalized ensemble system (20). This is done by deriving the corresponding differential equations. The later sections will then address the question of asymptotic behavior of these differential equations.

5.1. Empirical Mean

For the first moment, $\widehat{S}^M$ defined in (19a), we simply obtain

$$\mathrm{d}\widehat{S}_t^M = A\widehat{S}_t^M \mathrm{d}t + \frac{1}{\sqrt{M}}\mathbf{E}_t^M \mathrm{d}\widetilde{\omega}_t^M + \Big((\mathbf{F}_t^M + \mathbf{G}_t^M)\widehat{S}_t^M + \mathbf{H}_t^M y_t + \frac{1}{\sqrt{M}}\mathbf{J}_t^M \widetilde{\nu}_t^M\Big)\mathrm{d}N_t \tag{28}$$

where $\widetilde{\omega}_t^M = \frac{1}{\sqrt{M}}\sum_{i=1}^{M}\omega_t^i$ so that $\mathsf{E}[\mathrm{d}\widetilde{\omega}_t^M(\mathrm{d}\widetilde{\omega}_t^M)^\top] = I_m \mathrm{d}t$ and for $t = \tau_k$, $\widetilde{\nu}_t^M = \frac{1}{\sqrt{M}}\sum_{i=1}^{M}\nu_t^i \sim \mathcal{N}(0, V)$. Using the conditions listed in (21b), (21c) and the notation L_t^M, introduced in (22), we get

$$\mathrm{d}\widehat{S}_t^M = A\widehat{S}_t^M \mathrm{d}t + \frac{1}{\sqrt{M}}\mathbf{E}_t^M \mathrm{d}\widetilde{\omega}_t^M + \Big(L_t^M(y_t - C\widehat{S}_t^M) + \frac{1}{\sqrt{M}}\mathbf{J}_t^M \widetilde{\nu}_t^M\Big)\mathrm{d}N_t. \tag{29}$$

We note that the different values of $\mathbf{E}_t^M$ and $\mathbf{J}_t^M$, that were used in Section 3 and Section 4, allow us to get more tailored expressions for each of the four cases discussed there. We will revisit this equation in Section 7 to carry out asymptotic analysis of the first moment for each of the four ensemble filters introduced in Section 4. For the time being, however, we use (29) to derive the differential equation for evolution of empirical covariance.

5.2. Empirical Covariance

For working out the differential equation for the evolution of empirical covariance, we consider the auxiliary process $q_t^\ell := S_t^\ell - \widehat{S}_t^M$, for $\ell = 1, \dots, M$. It is observed that

$$\mathrm{d}q_t^\ell = (A + \mathbf{D}_t^M)q_t^\ell \mathrm{d}t + \mathbf{E}_t^M \mathrm{d}(\omega_t^\ell - \frac{1}{\sqrt{M}}\widetilde{\omega}_t^M) + \Big[\mathbf{F}_t^M q_t^\ell + \mathbf{J}_t^M(\bar{\nu}_t^\ell - \frac{1}{\sqrt{M}}\widetilde{\nu}_t^M)\Big]\,\mathrm{d}N_t.$$

We can apply Ito's chain rule from Proposition A.1 in the Appendix to get

$$\begin{aligned}
\mathrm{d}q_t^\ell(q_t^\ell)^\top &= \Big((A + \mathbf{D}_t)q_t^\ell(q_t^\ell)^\top + q_t^\ell(q_t^\ell)^\top(A + \mathbf{D}_t)^\top + (1 - \frac{1}{M})\mathbf{E}_t^M(\mathbf{E}_t^M)^\top\Big)\,\mathrm{d}t \\
&\quad + \mathbf{E}_t^M \mathrm{d}(\omega_t^\ell - \frac{1}{\sqrt{M}}\widetilde{\omega}_t^M)(q_t^\ell)^\top + q_t^\ell\big(\mathbf{E}_t^M \mathrm{d}(\omega_t^\ell - \frac{1}{\sqrt{M}}\widetilde{\omega}_t^M)\big)^\top \\
&\quad + \left[\Big(q_t^\ell + \mathbf{F}_t^M q_t^\ell + \mathbf{J}_t^M(\nu_t^\ell - \frac{1}{\sqrt{M}}\widetilde{\nu}_t^M)\Big)\Big(q_t^\ell + \mathbf{F}_t^M q_t^\ell + \mathbf{J}_t^M(\nu_t^\ell - \frac{1}{\sqrt{M}}\widetilde{\nu}_t^M)\Big)^\top - q_t^\ell(q_t^\ell)^\top\right]\mathrm{d}N_t.
\end{aligned}$$

Note that, we can write $Q_t^M = \frac{1}{M}\sum_{\ell=1}^{M} q_t^\ell(q_t^\ell)^\top$ and use (21a). Furthermore, since $\sum_{\ell=1}^{M} q_t^\ell = 0$, the terms like $\sum_\ell \widetilde{\omega}_t^M(q_t^\ell)^\top$ eliminate, so it follows that

$$\begin{aligned}
\mathrm{d}Q_t^M &= \Big(AQ_t^M + Q_t^M A^\top + BB^\top - \frac{1}{M}\mathbf{E}_t^M(\mathbf{E}_t^M)^\top\Big)\,\mathrm{d}t + \frac{1}{M}\sum_{\ell=1}^{M}\mathbf{E}_t^M \mathrm{d}\omega_t^\ell(q_t^\ell)^\top + \frac{1}{M}\sum_{\ell=1}^{M} q_t^\ell(\mathbf{E}_t^M \mathrm{d}\omega_t^\ell)^\top \\
&\quad + \Big[(I + \mathbf{F}_t^M)Q_t^M(I + \mathbf{F}_t^M)^\top - Q_t^M \\
&\qquad + \frac{1}{M}\sum_{\ell=1}^{M}\Big((q_t^\ell + \mathbf{F}_t^M q_t^\ell)(\nu_t^\ell - \frac{1}{\sqrt{M}}\widetilde{\nu}_t^M)^\top(\mathbf{J}_t^M)^\top + \mathbf{J}_t^M(\nu_t^\ell - \frac{1}{\sqrt{M}}\widetilde{\nu}_t^M)(q_t^\ell + \mathbf{F}_t^M q_t^\ell)^\top \\
&\qquad + \mathbf{J}_t^M(\nu_t^\ell - \frac{1}{\sqrt{M}}\widetilde{\nu}_t^M)(\nu_t^\ell - \frac{1}{\sqrt{M}}\widetilde{\nu}_t^M)^\top(\mathbf{J}_t^M)^\top\Big)\Big]\mathrm{d}N_t.
\end{aligned}$$

In compact form, we can write

$$
\begin{aligned}
\mathrm{d}Q_t^M &= \Big(AQ_t^M + Q_t^M A^\top + BB^\top - \frac{1}{M}\mathbf{E}_t^M(\mathbf{E}_t^M)^\top\Big)\,\mathrm{d}t + \frac{1}{M}\mathrm{d}\mathcal{M}_t \\
&\quad + \Big[(I+\mathbf{F}_t^M)Q_t^M(I+\mathbf{F}_t^M)^\top - Q_t^M + \mathbf{J}_t^M\Big(\frac{1}{M}\sum_{\ell=1}^M(\nu_t^\ell - \frac{1}{\sqrt{M}}\widetilde{\nu}_t^M)(\nu_t^\ell - \frac{1}{\sqrt{M}}\widetilde{\nu}_t^M)^\top)\Big)(\mathbf{J}_t^M)^\top\Big]\mathrm{d}N_t
\end{aligned}
$$

where the martingale $\mathcal{M}_t$ is defined as

$$
\begin{aligned}
\mathrm{d}\mathcal{M}_t &= \sum_{\ell=1}^M \mathbf{E}_t^M \mathrm{d}\omega_t^\ell (q_t^\ell)^\top + \sum_{\ell=1}^M q_t^\ell(\mathbf{E}_t^M\mathrm{d}\omega_t^\ell)^\top \\
&\quad + \sum_{\ell=1}^M\Big[(q_t^\ell + \mathbf{F}_t^M q_t^\ell)(\nu_t^\ell - \frac{1}{\sqrt{M}}\widetilde{\nu}_t^M)^\top(\mathbf{J}_t^M)^\top + \mathbf{J}_t^M(\nu_t^\ell - \frac{1}{\sqrt{M}}\widetilde{\nu}_t^M)(q_t^\ell + \mathbf{F}_t^M q_t^\ell)^\top\Big]\mathrm{d}N_t.
\end{aligned}
$$

Using (21d), and introducing the notation

$$
V_t^M := \frac{1}{M}\sum_{\ell=1}^M(\nu_t^\ell - \frac{1}{\sqrt{M}}\widetilde{\nu}_t^M)(\nu_t^\ell - \frac{1}{\sqrt{M}}\widetilde{\nu}_t^M)^\top, \tag{30}
$$

we get

$$
\begin{aligned}
\mathrm{d}Q_t^M &= \Big(AQ_t^M + Q_t^M A^\top + BB^\top - \frac{1}{M}\mathbf{E}_t^M(\mathbf{E}_t^M)^\top\Big)\,\mathrm{d}t + \frac{1}{M}\mathrm{d}\mathcal{M}_t \\
&\qquad + \left[-L_t^M C Q_t^M - \mathbf{J}_t^M V(\mathbf{J}_t^M)^\top + \mathbf{J}_t^M V_t^M(\mathbf{J}_t^M)^\top\right]\mathrm{d}N_t.
\end{aligned}
\tag{31}
$$

We can thus summarize the foregoing calculations in the form of following result:

Proposition 5.1. *Consider the ensemble filters described by* (20)*, with empirical mean* $\widehat{S}_t^M$ *and covariance* Q_t^M *described by* (19a) *and* (19b)*, respectively. If conditions in* (21) *hold, then* $\widehat{S}_t^M$ *satisfies* (29) *and* Q_t^M *satisfies* (31).

The result in Proposition 5.1 describes the evolution of empirical moments conditioned upon the sampling process. It is instructive to compare the resulting equations (29) and (31) with their optimal counterparts given in (6) and (7), respectively. The differential equations for empirical moments can be seen as a perturbation of the optimal moments, and we expect these perturbations to diminish (in appropriate sense) as the number of particles M gets large.

5.3. Expectation of Moments with respect to Sampling Process

As a qualitative indicator of the performance of the ensemble filters proposed in the previous section, we now look at the expectation of empirical moments with respect to the sampling process. This will allow us to compare expectation of empirical moments with the expectation of optimal moments. It must also be noted that, in the analysis of ensemble filters for continuous-time systems [9], the differential equation for the empirical covariance describes a stochastic fluctuation around a deterministic Riccati equation. However, for our purposes, we will only look at the expectation of the empirical covariance Q_t^M so that the randomness due to noise processes $(\omega_t^\ell)_{t\geq 0}$, $(\nu_t^\ell)_{t\geq 0}$, $\ell = 1,\dots,M$, and the sampling process N_t is averaged out and the resulting equation is purely deterministic.

To derive this differential equation, and to make our results more tailored to the class of four ensemble filters introduced in Section 4, we fix $\mathbf{E}_t^M$ and $\mathbf{J}_t^M$ as,

$$
\mathbf{E}_t^M = \mathbf{1}_\omega B, \quad \mathbf{J}_t^M = -\mathbf{1}_v L_t^M \tag{32}
$$

where $\mathbf{1}_\omega$ and $\mathbf{1}_v$ can either be 1 or 0, depending on whether the noise term is present in the prediction and correction part, respectively, for the ensemble filter being considered. This choice of $\mathbf{E}_t^M$ basically corresponds to driving the prediction term (resp., correction term) with the same noise statistics that appear in the system dynamics. The choice of $\mathbf{J}_t^M$ means that we introduce observation noise in the correction term and the injection gain gets multiplied by the noise. As seen earlier, these choices correspond to the particular cases of ensemble filters.

We note that the evolution of $\widehat{S}_t^M$ and Q_t^M, as desribed in (29) and (31), respectively, involves the martingale $\mathcal{M}_t$ and the noise terms. In this paper, we do not carry out the stochastic analysis, but rather look at their expectations. This expectation is taken with respect to the noise terms, and the sampling process driven by Poisson counter that describes the correction times. We are thus interested in calculating $\mathcal{Q}_t^M := \mathsf{E}[Q_t^M \mid Q_0^M]$, and $\widehat{\mathcal{S}}_t^M := \mathsf{E}[\widehat{S}_t^M \mid \widehat{S}_0^M]$ and we conclude this section by providing two main results that describe the differential equations governing the evolution of expected empirical mean and covariance. The primary tool used in the derivation of these results is similar and is described in detail in Proposition A.3.

Theorem 5.2. *Consider the ensemble filters described by* (20) *under the conditions* (21)*, with empirical covariance* Q_t^M *described by* (31)*. Using the values in* (32)*, we obtain*

$$\boxed{\frac{\mathrm{d}\mathcal{Q}_t^M}{\mathrm{d}t} = A\mathcal{Q}_t^M + \mathcal{Q}_t^M A^\top + BB^\top - \frac{\mathbf{1}_\omega}{M}BB^\top - \lambda\Big(\mathcal{L}_t^M C\mathcal{Q}_t^M - \frac{\mathbf{1}_v}{M}\mathcal{L}_t^M V(\mathcal{L}_t^M)^\top\Big).} \tag{33}$$

where $\mathcal{L}_t^M := \mathcal{Q}_t^M C^\top (C\mathcal{Q}_t^M C + V)^{-1}$*, and* $\mathcal{Q}_0^M = Q_0^M$*.*

Proof. We recall the equation for Q_t^M and note that

$$\mathsf{E}[V_t^M] = \mathsf{E}\left(\frac{1}{M}\sum_{\ell=1}^M (\nu_t^\ell - \frac{1}{\sqrt{M}}\widetilde{\nu}_t^M)(\nu_t^\ell - \frac{1}{\sqrt{M}}\widetilde{\nu}_t^M)^\top\right) = (1-\frac{1}{M})V.$$

Using this property, we can now compute the expectation of the empirical covariance with respect to the sampling process N_t, the process noise ω_t^ℓ and the observation noise ν_t^ℓ. The underlying derivation is carried out in the proof of Proposition A.3 and its application to (31) yields (33). □

Similarly, for the expectation of first moment $\widehat{\mathcal{S}}_t^M$, we adopt a similar approach and arrive at the following result:

Theorem 5.3. *Consider the ensemble filters described by* (20) *under the conditions* (21)*, with empirical mean* $\widehat{S}_t^M$ *described by* (29)*. Using the values in* (32)*, we obtain*

$$\boxed{\frac{\mathrm{d}\widehat{\mathcal{S}}_t^M}{\mathrm{d}t} = A\widehat{\mathcal{S}}_t^M + \lambda\mathcal{L}_t^M C(\mathsf{E}[X_t] - \widehat{\mathcal{S}}_t^M).} \tag{34}$$

The proof of Theorem 5.3 is again carried out using Proposition A.3. This is because the dynamics of the empirical mean along a sample path is described by (29) which has the same structure as (A.1). In the following sections, we will analyze the asymptotic behavior of the differential equations (33) and (34) with respect to time and the number of particles. This allows us to compare the performance of the proposed ensemble filters with their optimal counterparts in Section 2.

6. Asymptotic Convergence of Expected Covariance

As described in Section 2, the evolution of the expectation of the optimal second moment with respect to the sampling process is described by the following differential equation:

$$\dot{\mathcal{P}}_t = A\mathcal{P}_t + \mathcal{P}_t A^\top + BB^\top - \lambda\mathcal{P}_t C^\top (C\mathcal{P}_t C^\top + V)^{-1} C\mathcal{P}_t. \tag{35}$$

Our primary goal in this section is to compare the expected error covariance resulting from the ensemble filters (33) with the optimal value (35).

The existence of solution for (35) and convergence of the solution of $\mathcal{P}_t$ to a fixed point as $t \to \infty$ has been a topic of our previous work [29] and [7]. In those works, we provide conditions on system data

(A, B, C), the noise statistics, and the mean sampling rate λ such that the resulting solution converges to a fixed point asymptotically. In this article, our focus is on showing that $\mathcal{Q}_t^M$ actually converges to $\mathcal{P}_t$ as the number of particles M gets large, and moreover the difference due to the initial conditions between the two systems decays exponentially with time.

To formalize this convergence result, we introduce some basic notation and some assumptions. In what follows, we will denote the $n \times n$ positive semidefinite matrices by $\mathbb{S}^{n\times n}$, and the positive definite matrices by $\mathbb{S}_+^{n\times n}$. The eigenvalue of a matrix will be denoted by μ, and the largest and smallest eigenvalues are denoted by $\mu_{\max}$ and $\mu_{\min}$, respectively.

Assumption 6.1. The pair (A, B) is controllable. The pair (A, C) is observable. The mean sampling rate λ is large enough to satisfy the inequality

$$\lambda > \mu_{\max}(A + A^\top), \tag{36}$$

where $\mu_{\max}(\cdot)$ denotes the largest eigenvalue of its matrix argument.

Assumption 6.1 provides three basic assumptions that are used in establishing the existence of steady state solutions of (35) and the convergence of trajectories towards that steady state. Such results have been established in [29] and [7].

Assumption 6.2. For a symmetric positive semidefinite matrix $Q_0 \in \mathbb{S}^{n\times n}$, consider the differential equation,

$$\frac{\mathrm{d}\mathcal{Q}_t}{\mathrm{d}t} = A\mathcal{Q}_t + \mathcal{Q}_t A^\top + BB^\top - \lambda\Big(\mathcal{L}_t C\mathcal{Q}_t - \frac{1}{M}\mathcal{L}_t V(\mathcal{L}_t)^\top\Big), \quad \mathcal{Q}_0 = Q_0, \tag{37}$$

where $\mathcal{L}_t := \mathcal{Q}_t C^\top (C\mathcal{Q}_t C^\top + V)^{-1}$. It is assumed that the solution $\mathcal{Q}_t$ satisfying (37) is defined for all $t \geq 0$, and there exists $\overline{\mathcal{Q}}_0 \in \mathbb{S}_+^{n\times n}$ such that $\mathcal{Q}_t \leq \overline{\mathcal{Q}}_0$ holds for every $M \geq 1$.

Assumption 6.2 basically requires a uniform upper bound on the solution of expected error covariance resulting from the ensemble filters. It must be noted that, for equation (37), uniformity with respect to M does not introduce any restriction because the right-hand side decreases with increase in the value of M. Due to Assumption 6.2, it also follows that (33) has a unique solution $\mathcal{Q}_t^M$ bounded by some positive definite matrix $\overline{\mathcal{Q}}_0$.

The main result of this section provides a quantitative estimate on the difference between the solutions of (33) and (35). To state this result, we need following statements:

Lemma 6.3 ([17, Prop. 3.1])**.** *Consider the matrices $Q \in \mathbb{S}_+^{n\times n}$, $C \in \mathbb{R}^{p\times n}$, and $V \in \mathbb{S}_+^{p\times p}$, and let $L_Q := QC^\top(CQC^\top + V)^{-1}$. Then, it holds that,*

$$\|L_Q\| \leq \frac{\|C\|}{\mu_{\min}(V)}\|Q\| =: \gamma_Q. \tag{38}$$

Moreover, for $Q_1, Q_2 \in \mathbb{S}_+^{n\times n}$, we have

$$L_{Q_1} - L_{Q_2} = (I - L_{Q_2}C)(Q_1 - Q_2)C^\top(CQ_1C^\top + V)^{-1},$$

and

$$\|L_{Q_1} - L_{Q_2}\| \leq \|I - L_{Q_2}C\| \, \frac{\|C\|}{\mu_{\min}(V)} \, \|Q_1 - Q_2\| .$$

Lemma 6.4. *Consider the solution $\mathcal{Q}_t^M$ of (33), and suppose that Assumption 6.2 holds, so that $\mathcal{Q}_t^M \leq \overline{\mathcal{Q}}_0$, for all $t \geq 0$, and some $\overline{\mathcal{Q}}_0 \in \mathbb{S}_+^{n\times n}$. Then, there exists a constant scalar $\kappa_{\overline{\mathcal{Q}}_0}$ such that*

$$\big\|I - \mathcal{Q}_t^M C^\top (C\mathcal{Q}_t^M C^\top + V)^{-1}C\big\| \leq \kappa_{\overline{\mathcal{Q}}_0} := 1 + \|\overline{\mathcal{Q}}_0\| \cdot \|C^\top V^{-1} C\|, \tag{39}$$

for $t \in [0, \infty[$.

Theorem 6.5. *Consider the solution $\mathcal{P}_t$ of the equation (35) with initial condition $\mathcal{P}_0 \in \mathbb{S}^{n\times n}$ and the process $\mathcal{Q}_t^M$ solving (33) with initial condition $\mathcal{Q}_0 \in \mathbb{S}^{n\times n}$. Suppose that Assumption 6.1 and Assumption 6.2 hold, so that $\mathcal{Q}_t^M \leq \overline{\mathcal{Q}}_0$, for all $t \geq 0$. If $\mathcal{P}_0 \geq \mathcal{Q}_0$, then*

$$\big\|\mathcal{P}_t - \mathcal{Q}_t^M\big\| \leq \exp(-\alpha t)\,(1 + \beta_0 t)\,\|\mathcal{P}_0 - \mathcal{Q}_0\| + \frac{\beta_0}{\alpha^2 M}(\|BB^\top\|\mathbf{1}_\omega + \|\bar{V}_\lambda\|\mathbf{1}_v) \tag{40}$$

where $\alpha := \lambda - \mu_{\max}(A + A^\top)$*,* $\beta_0 = \lambda\kappa^2_{\overline{\mathcal{Q}}_0} \exp\left(\frac{\lambda\kappa^2_{\overline{\mathcal{Q}}_0}}{\alpha}\right)$*, and* $\bar{V}_\lambda = \lambda\mu_{\max}(V)\gamma^2_{\overline{\mathcal{Q}}_0} I$*, with* $\gamma_{\overline{\mathcal{Q}}_0}$ *and* $\kappa_{\overline{\mathcal{Q}}_0}$ *defined as in* (38) *and* (39)*, respectively.*

The proof of Theorem 6.5 builds on several intermediate developments which appear in the sequel. It will be shown in Section 6.4 that the proof essentially follows from Proposition 6.7 and Proposition 6.8 which appear in Section 6.2 and Section 6.3, respectively.

6.1. Monotonic Properties

As a first step in the proof of Theorem 6.5, we need to analyze the equation (33) carefully, and establish some monotonic properties of its solution with respect to initial conditions. For the optimal filter, the continuity on initial values was shown in [7]. In our case, the dynamics to be analyzed are different and to study the convergence in this case, let us rewrite the differential equation (35) as follows, where we use the notation $\mathcal{L}_t = \mathcal{P}_t C^\top (C\mathcal{P}_t C^\top + V)^{-1}$:

$$\begin{aligned}
\dot{\mathcal{P}}_t &= A\mathcal{P}_t + \mathcal{P}_t A^\top + BB^\top - \lambda(\mathcal{L}_t C\mathcal{P}_t + \mathcal{P}_t C^\top \mathcal{L}_t^\top) + \lambda\mathcal{L}_t V\mathcal{L}_t^\top + \lambda\mathcal{L}_t C\mathcal{P}_t C^\top \mathcal{L}^\top \\
&= A\mathcal{P}_t + \mathcal{P}_t A^\top + BB^\top + \lambda\mathcal{L}_t V\mathcal{L}_t^\top + \lambda(I - \mathcal{L}_t C)\mathcal{P}_t(I - \mathcal{L}_t C)^\top - \lambda\mathcal{P}_t \\
&= A\mathcal{P}_t + \mathcal{P}_t A^\top + BB^\top + \lambda\mathcal{L}_t V\mathcal{L}_t^\top + \lambda(I - \mathcal{L}_t C)\mathcal{P}_t(I - \mathcal{L}_t C)^\top - \frac{\lambda}{2}\mathcal{P}_t - \frac{\lambda}{2}\mathcal{P}_t \\
&= \left(A - \frac{\lambda}{2}I\right)\mathcal{P}_t + \mathcal{P}_t\left(A - \frac{\lambda}{2}I\right)^\top + BB^\top + \lambda\mathcal{L}_t V\mathcal{L}_t^\top + \lambda(I - \mathcal{L}_t C)\mathcal{P}_t(I - \mathcal{L}_t C)^\top.
\end{aligned}$$

Thus, by letting $A_\lambda := \left(A - \frac{\lambda}{2}I\right)$, (35) can now be equivalently expressed as

$$\dot{\mathcal{P}}_t = A_\lambda\mathcal{P}_t + \mathcal{P}_t A_\lambda^\top + BB^\top + \lambda\mathcal{L}_t V\mathcal{L}_t^\top + \lambda(I - \mathcal{L}_t C)\mathcal{P}_t(I - \mathcal{L}_t C)^\top. \tag{41}$$

Let us denote the flow of (35) at time $t \geq 0$, starting with initial condition $\mathcal{P}_0$, by $\phi_t(\mathcal{P}_0)$, with $\phi_0(\mathcal{P}_0) = \mathcal{P}_0$ and

$$\phi_t(\mathcal{P}_0) = \mathcal{E}_t^\lambda \mathcal{P}_0 \mathcal{E}_t^{\lambda\top} + \int_0^t \mathcal{E}_{t-s}^\lambda \Big[\lambda(I - \mathcal{L}_s C)\phi_s(\mathcal{P}_0)(I - \mathcal{L}_s C)^\top + \lambda\mathcal{L}_s V\mathcal{L}_s^\top + BB^\top\Big]\mathcal{E}_{t-s}^{\lambda\top}\mathrm{d}s \tag{42}$$

where

$$\mathcal{E}_{s,t}^\lambda := \exp\left[\left(A - \frac{\lambda}{2}I\right)(t - s)\right]$$

and we let $\mathcal{E}_t^\lambda := \mathcal{E}_{0,t}^\lambda$. Similarly, the equation (33) can be written as

$$\dot{\mathcal{Q}}_t^M = A_\lambda\mathcal{Q}_t^M + \mathcal{Q}_t^M A_\lambda^\top + \left(1 - \frac{\mathbf{1}_\omega}{M}\right)BB^\top + \lambda\left(1 + \frac{\mathbf{1}_v}{M}\right)\mathcal{L}_t^M V(\mathcal{L}_t^M)^\top + \lambda(I - \mathcal{L}_t^M C)\mathcal{Q}_t^M(I - \mathcal{L}_t^M C)^\top. \tag{43}$$

We denote the solution of (43) at time $t \geq 0$, starting with initial condition $\mathcal{Q}_0$, by $\varphi_t(\mathcal{Q}_0)$. In the analysis, we also consider the case where $\mathbf{1}_v = 0$ and in that case we denote the solution of (43) by $\varphi_t^\omega(\mathcal{Q}_0)$.

Using Lemma B.2 in Appendix, we can get bounds on the difference of flows $\phi_t(\mathcal{P}_0) - \varphi_t^\omega(\mathcal{Q}_0)$. The details are similar to ones provided in [7, Proof of Prop. 2]. In particular, as a lower bound, we get

$$\begin{aligned}
\phi_t(\mathcal{P}_0) - \varphi_t^\omega(\mathcal{Q}_0) \geq \mathcal{E}_t^\lambda(\mathcal{P}_0 - \mathcal{Q}_0)(\mathcal{E}_t^\lambda)^\top &+ \frac{\mathbf{1}_\omega}{M}\int_0^t \mathcal{E}_{t-s}^\lambda BB^\top(\mathcal{E}_{t-s}^\lambda)^\top\mathrm{d}s \\
&+ \int_0^t \mathcal{E}_{t-s}^\lambda[\lambda(I - \mathcal{L}_s C)(\phi_s(\mathcal{P}_0) - \varphi_s^\omega(\mathcal{Q}_0))(I - \mathcal{L}_s C)^\top](\mathcal{E}_{t-s}^\lambda)^\top\mathrm{d}s.
\end{aligned} \tag{44}$$

Similarly, an upper bound on the difference of two flows is given by

$$\phi_t(\mathcal{P}_0) - \varphi_t^\omega(\mathcal{Q}_0) \leq \mathcal{E}_t^\lambda(\mathcal{P}_0 - \mathcal{Q}_0)(\mathcal{E}_t^\lambda)^\top + \frac{\mathbf{1}_\omega}{M}\int_0^t \mathcal{E}_{t-s}^\lambda BB^\top(\mathcal{E}_{t-s}^\lambda)^\top\mathrm{d}s$$

$$+\int_0^t \mathcal{E}^\lambda_{t-s}[\lambda(I-\mathcal{L}^M_s C)(\phi_s(\mathcal{P}_0)-\varphi^\omega_s(\mathcal{Q}_0))(I-\mathcal{L}^M_s C)^\top](\mathcal{E}^\lambda_{t-s})^\top \mathrm{d}s. \quad (45)$$

The following statement about the monotonicity of flows is, therefore, a direct consequence of the inequality (44):

Lemma 6.6. *If $\mathcal{P}_0 \geq \mathcal{Q}_0$ then*

$$\phi_t(\mathcal{P}_0) - \varphi^\omega_t(\mathcal{Q}_0) \geq 0.$$

Using these developments, we now proceed to the proof of Theorem 6.5 which is carried out in two steps. In the first case, we compare $\phi_t(\mathcal{P}_0)$ with $\varphi^\omega_t(\mathcal{Q}_0)$, and then in the second step we compare $\varphi^\omega_t(\mathcal{Q}_0)$ with $\varphi_t(\mathcal{Q}_0)$.

6.2. Ensemble Filters with Deterministic Correction: `NPDC-EnF` *and* `DeT-EnF` *Cases*

In (43), setting $\mathbf{1}_v = 0$ corresponds to the case of `NPDC-EnF` and `DeT-EnF`. The following statement provides a bound on the difference between $\varphi^\omega_t(\mathcal{Q}_0)$ and the optimal expected error covariance $\mathcal{P}_t$, and could therefore be seen as a result of independent interest.

Proposition 6.7. *Consider the solution $\mathcal{P}_t$ of the equation* (35) *with initial condition $\mathcal{P}_0$ and the process $\mathcal{Q}^\omega_t = \varphi^\omega_t(\mathcal{Q}_0)$ solving* (33) *with initial condition $\mathcal{Q}_0$ and $\mathbf{1}_v = 0$. Suppose that Assumption 6.1 and Assumption 6.2 hold, so that $\mathcal{Q}^\omega_t \leq \overline{\mathcal{Q}}_0$, for all $t \geq 0$, and let α and β_0 be defined as in Theorem 6.5. If $\mathcal{P}_0 \geq \mathcal{Q}_0$, then*

$$\|\mathcal{P}_t - \mathcal{Q}^\omega_t\| \leq \exp(-\alpha t)\,(1+\beta_0 t)\,\|\mathcal{P}_0 - \mathcal{Q}_0\| + \frac{\beta_0 \|BB^\top\| \mathbf{1}_\omega}{\alpha^2 M}.$$

Proof. Recall that the flows of $\mathcal{P}_t$ and $\mathcal{Q}^\omega_t$ are described by $\phi_t(\mathcal{P}_0)$ and $\varphi^\omega_t(\mathcal{Q}_0)$, as we set $\mathbf{1}_v = 0$. Using the assumption $\mathcal{P}_0 \geq \mathcal{Q}_0$ and Lemma 6.6, we immediately have that $\phi_t(\mathcal{P}_0) - \varphi^\omega_t(\mathcal{Q}_0) \geq 0$. On the other hand, inequality (45) can be rewritten as

$$\phi_t(\mathcal{P}_0) - \varphi^\omega_t(\mathcal{Q}_0) \leq \mathcal{E}^\lambda_t(\mathcal{P}_0 - \mathcal{Q}_0)(\mathcal{E}^\lambda_t)^\top + \frac{\mathbf{1}_\omega}{M}\int_0^t \mathcal{E}^\lambda_{t-s} BB^\top (\mathcal{E}^\lambda_{t-s})^\top \mathrm{d}s$$
$$+ \lambda \int_0^t \mathcal{E}^\lambda_{t-s} K_s(\phi_s(\mathcal{P}_0) - \varphi^\omega_s(\mathcal{Q}_0))K_s^\top (\mathcal{E}^\lambda_{t-s})^\top \mathrm{d}s \quad (46)$$

where $K_s = I - \mathcal{L}^\omega_s C$ and $\mathcal{L}^\omega_s := \mathcal{Q}^\omega_s C^\top (C\mathcal{Q}^\omega_s C^\top + V)^{-1}$. Introducing norm on both sides, the above inequality results in,

$$\|\phi_t(\mathcal{P}_0) - \varphi^\omega_t(\mathcal{Q}_0)\| \leq \left\|\mathcal{E}^\lambda_t\right\|^2 \|\mathcal{P}_0 - \mathcal{Q}_0\| + \frac{\mathbf{1}_\omega}{M}\int_0^t \|\mathcal{E}^\lambda_{t-s}\|^2 \|BB^\top\| \mathrm{d}s$$
$$+ \lambda \int_0^t \left\|\mathcal{E}^\lambda_{t-s}\right\|^2 \|K_s\|^2 \left\|\phi_s(\mathcal{P}_0) - \varphi^\omega_s(\mathcal{Q}_0)\right\| \mathrm{d}s. \quad (47)$$

From Lemma 6.4, we have that $\|K_s\| \leq \kappa_{\overline{\mathcal{Q}}_0}$, for $t \in [0,\infty[$. By taking $\alpha = \lambda - \mu_{\max}(A + A^\top) > 0$, and recalling that $\left\|\mathcal{E}^\lambda_t\right\| \leq \exp\left(\left(\mu_{\max}(\frac{A+A^\top}{2}) - \frac{\lambda}{2}\right)t\right)$, we hence get,

$$\|\phi_t(\mathcal{P}_0) - \varphi^\omega_t(\mathcal{Q}_0)\| \leq \exp(-\alpha t)\,\|\mathcal{P}_0 - \mathcal{Q}_0\| + \frac{1-\exp(-\alpha t)}{\alpha M}\mathbf{1}_\omega \|BB^\top\|$$
$$+ \lambda \kappa^2_{\overline{\mathcal{Q}}_0} \int_0^t \exp(-\alpha(t-s))\,\|\phi_s(\mathcal{P}_0) - \varphi^\omega_s(\mathcal{Q}_0)\|\,\mathrm{d}s. \quad (48)$$

Applying Grönwall's inequality, and more specifically Lemma D.1 from the Appendix, we arrive at the result stated in Proposition 6.7. □

As the next step in the proof of Theorem 6.5, and as a result of independent interest as well, we now compare the two cases of (33) which are obtained from two different values of $\mathbf{1}_v$. Note that $\mathbf{1}_v \equiv 1$ corresponds to the case of DPNC-EnF(with $\mathbf{1}_\omega = 0$) and Vanilla-EnF(with $\mathbf{1}_\omega = 1$).

Proposition 6.8. *Consider the solution $\mathcal{Q}_t^M = \varphi_t(\mathcal{Q}_0^M)$ of the equation (33) with initial condition $\mathcal{Q}_0^M$ and some choice of $\mathbf{1}_\omega$, $\mathbf{1}_v$. Also, consider the process $\mathcal{Q}_t^\omega = \varphi_t^\omega(\mathcal{Q}_0)$ solving (33) with initial condition $\mathcal{Q}_0$ and $\mathbf{1}_v = 0$, while keeping all the other variables to be same. Suppose that Assumption 6.1 and Assumption 6.2 hold, so that $\mathcal{Q}_t^\omega \leq \mathcal{Q}_t^M \leq \overline{\mathcal{Q}}_0$, for all $t \geq 0$, and let α, β_0 and $\bar{V}_\lambda$ be defined as in Theorem 6.5. If $\mathcal{Q}_0^M \geq \mathcal{Q}_0$, then*

$$\left\| \mathcal{Q}_t^M - \mathcal{Q}_t^\omega \right\| \leq \exp(-\alpha t)\,(1+\beta_0 t)\left\| \mathcal{Q}_0^M - \mathcal{Q}_0 \right\| + \frac{\beta_0 \|\bar{V}_\lambda\|}{\alpha^2 M}\mathbf{1}_v.$$

Proof. For the processes $\mathcal{Q}_t^M$ and $\mathcal{Q}_t^\omega$, let us denote the corresponding injection gain by $\mathcal{L}_t^M$ and $\mathcal{L}_t^\omega$, so that, $\mathcal{L}_t^M = \mathcal{Q}_t^M C^\top (C\mathcal{Q}_t^M C^\top + V)^{-1}$ and $\mathcal{L}_t^\omega = \mathcal{Q}_t^\omega C^\top (C\mathcal{Q}_t^\omega C^\top + V)^{-1}$. Hence, by Lemma B.1 and (43), it follows that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}(\mathcal{Q}_t^M - \mathcal{Q}_t^\omega) &= A_\lambda(\mathcal{Q}_t^M - \mathcal{Q}_t^\omega) + (\mathcal{Q}_t^M - \mathcal{Q}_t^\omega)A_\lambda^\top \\ &\quad + \lambda(I - \mathcal{L}_t^M C)(\mathcal{Q}_t^M - \mathcal{Q}_t^\omega)(I - \mathcal{L}_t^\omega C)^\top + \frac{\lambda}{M}\mathcal{L}_t^M V (\mathcal{L}_t^M)^\top \mathbf{1}_v.\end{aligned}$$

Using Lemma B.2, we get

$$(I - \mathcal{L}_t^M C)(\mathcal{Q}_t^M - \mathcal{Q}_t^\omega)(I - \mathcal{L}_t^\omega C)^\top \leq (I - \mathcal{L}_t^\omega C)(\mathcal{Q}_t^M - \mathcal{Q}_t^\omega)(I - \mathcal{L}_t^\omega C)^\top.$$

Moreover, by Lemma 6.3, and $\bar{V}_\lambda$ introduced in Theorem 6.5, we get

$$\lambda \mathcal{L}_t^M V (\mathcal{L}_t^M)^\top \leq \lambda \mu_{\max}(V)\gamma_{\overline{\mathcal{Q}}_o}^2 I = \bar{V}_\lambda.$$

Using these two inequalities, it follows that

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathcal{Q}_t^M - \mathcal{Q}_t^\omega) \leq A_\lambda(\mathcal{Q}_t^M - \mathcal{Q}_t^\omega) + (\mathcal{Q}_t^M - \mathcal{Q}_t^\omega)A^\top + \lambda(I - \mathcal{L}_t^\omega C)(\mathcal{Q}_t^M - \mathcal{Q}_t^\omega)(I - \mathcal{L}_t^\omega C)^\top + \frac{\lambda}{M}\bar{V}_\lambda \mathbf{1}_v.$$

Integrating this differential inequality, we get

$$\begin{aligned}\varphi_t(\mathcal{Q}_0^M) - \varphi_t^\omega(\mathcal{Q}_0) &\leq \mathcal{E}_t^\lambda(\mathcal{Q}_0^M - \mathcal{Q}_0)(\mathcal{E}_t^\lambda)^\top + \frac{\mathbf{1}_v}{M}\int_0^t \mathcal{E}_{t-s}^\lambda \bar{V}_\lambda (\mathcal{E}_{t-s}^\lambda)^\top \mathrm{d}s \\ &\quad + \lambda \int_0^t \mathcal{E}_{t-s}^\lambda K_s(\varphi_s(\mathcal{Q}_0^M) - \varphi_s^\omega(\mathcal{Q}_0))K_s^\top (\mathcal{E}_{t-s}^\lambda)^\top \mathrm{d}s, \quad (49)\end{aligned}$$

where $K_s = I - \mathcal{L}_s^\omega C$. We thus have $\mathcal{Q}_t^M - \mathcal{Q}_t^\omega \geq 0$, for all $t \geq 0$. Taking the norm on both sides of the last inequality, we get

$$\|\mathcal{Q}_t^M - \mathcal{Q}_t^\omega\| \leq \left\|\mathcal{E}_t^\lambda\right\|^2 \left\|\mathcal{Q}_0^M - \mathcal{Q}_0\right\| + \frac{1}{M}\int_0^t \|\mathcal{E}_{t-s}^\lambda\|^2 \|\bar{V}_\lambda\| \mathrm{d}s + \lambda \int_0^t \|\mathcal{E}_{t-s}^\lambda\|^2 \|K_s\|^2 \|\varphi_s(\mathcal{Q}_0^M) - \varphi_s^\omega(\mathcal{Q}_0)\| \mathrm{d}s.$$

By Lemma 6.4, it is possible to find a constant scalar $\kappa_{\overline{\mathcal{Q}}_0}$ such that $\|K_s\| \leq \kappa_{\overline{\mathcal{Q}}_0}$, for $t \in [0, \infty)$. Using the scalar α as before, we know that $\left\|\mathcal{E}_t^\lambda\right\| \leq \exp(-0.5\alpha t)$, and this yields

$$\|\mathcal{Q}_t^M - \mathcal{Q}_t^\omega\| \leq \exp(-\alpha t)\left\|\mathcal{Q}_0^M - \mathcal{Q}_0\right\| + \frac{1 - \exp(-\alpha t)}{aM}\|\bar{V}_\lambda\| + \lambda \kappa_{\overline{\mathcal{Q}}_0}^2 \int_0^t \exp(-\alpha(t-s))\left\|\varphi_s(\mathcal{Q}_0^M) - \varphi_s^\omega(\mathcal{Q}_0)\right\| \mathrm{d}s$$

Applying Lemma D.1 from the Appendix, we arrive at the result stated in Proposition 6.8. □

6.4. Generic Case

Using the result of Proposition 6.7 and Proposition 6.8, we can now complete the proof of Theorem 6.5. **Proof of Theorem 6.5.** Let us consider the processes $\mathcal{P}_t = \phi_t(\mathcal{P}_0)$ and $\mathcal{Q}_t^M = \varphi_t(\mathcal{Q}_0)$ as defined in the statement of Theorem 6.5. We consider an auxilliary process $\mathcal{Q}_t^\omega = \varphi_t^\omega(\mathcal{Q}_0)$ that is obtained by solving (33) with $\mathbf{1}_v = 0$ while keeping all the other variables, and the initial condition, as in $\mathcal{Q}_t^M$. The use of triangle inequality allows us to write

$$\|\mathcal{P}_t - \mathcal{Q}_t^M\| \leq \|\phi_t(\mathcal{P}_0) - \varphi_t^\omega(\mathcal{Q}_0)\| + \|\varphi_t^\omega(\mathcal{Q}_0) - \varphi_t(\mathcal{Q}_0)\|. \tag{50}$$

In the above inequality, the upper bound on the first term is obtained from Proposition 6.7, that is,

$$\|\phi_t(\mathcal{P}_0) - \varphi_t^\omega(\mathcal{Q}_0)\| \leq \exp(-\alpha t)\,(1+\beta_0 t)\,\|\mathcal{P}_0 - \mathcal{Q}_0\| + \frac{\beta_0 \|BB^\top\| \mathbf{1}_\omega}{\alpha^2 M}. \tag{51}$$

For the second term on the right-hand side of (50), due to the same initial condition, Proposition 6.8 yields

$$\|\varphi_t^\omega(\mathcal{Q}_0) - \varphi_t(\mathcal{Q}_0)\| \leq \frac{\beta_0 \|\bar{V}_\lambda\|}{\alpha^2 M} \mathbf{1}_v \tag{52}$$

for the same choice of α and β_0 as in (51). Plugging the bounds from (51) and (52) into (50) leads to the bound given in (40). □

7. First Moment Asymptotics

After analyzing the convergence of empirical covariance towards the optimal covariance in the previous section, we now study the convergence of empirical mean in this section. In contrast to the previous section, where the convergence is achieved as the time increases *and* the number of particles increases, the convergence of the empirical mean towards the optimal mean of the Kalman filter is established asymptotically with respect to time, while the number of particles may be fixed.

Towards this end, let us begin with the processes $\widehat{X}_t$ and $\widehat{S}_t^M$ defined in (6) and (29); at this stage it is more convenient to rewrite their dynamics as follows:

$$\mathrm{d}\widehat{X}_t = A\widehat{X}_t \mathrm{d}t + L_t^{\mathrm{opt}}(CX_t + \nu_t - C\widehat{X}_t)\mathrm{d}N_t, \tag{53}$$

$$\mathrm{d}\widehat{S}_t^M = A\widehat{S}_t^M \mathrm{d}t + \frac{1}{\sqrt{M}}\mathbf{E}_t^M \mathrm{d}\widetilde{\omega}_t^M + \Big(L_t^M(CX_t + \nu_t - C\widehat{S}_t^M) + \frac{1}{\sqrt{M}}\mathbf{J}_t^M \widetilde{\nu}_t^M\Big)\mathrm{d}N_t. \tag{54}$$

We let $\widehat{\mathcal{S}}_t^M := \mathsf{E}[\widehat{S}_t^M]$, and $\widehat{\mathcal{X}}_t = \mathsf{E}[\widehat{X}_t]$. Applying Proposition A.3 to the system of $\widehat{X}_t$ and P_t, we first observe that

$$\dot{\widehat{\mathcal{X}}}_t = A\widehat{\mathcal{X}}_t + \lambda \mathcal{L}_t^{\mathrm{opt}} C(\mathsf{E}[X_t] - \widehat{\mathcal{X}}_t)$$

and similarly

$$\dot{\widehat{\mathcal{S}}}_t^M = A\widehat{\mathcal{S}}_t^M + \lambda \mathcal{L}_t^M C(\mathsf{E}[X_t] - \widehat{\mathcal{S}}_t^M).$$

We now look at the asymptotic behavior of $\widehat{\chi}_t^M := \widehat{\mathcal{X}}_t - \widehat{\mathcal{S}}_t^M$, which satisfies

$$\begin{aligned}\dot{\widehat{\chi}}_t^M &= A\widehat{\chi}_t^M + \lambda \mathcal{L}_t^{\mathrm{opt}} C(\mathsf{E}[X_t] - \widehat{\mathcal{X}}_t) - \lambda \mathcal{L}_t^M C(\mathsf{E}[X_t] - \widehat{\mathcal{S}}_t^M) \\ &= (A - \lambda \mathcal{L}_t^M C)\widehat{\chi}_t^M + \lambda(\mathcal{L}_t^{\mathrm{opt}} - \mathcal{L}_t^M) C(\mathsf{E}[X_t] - \widehat{\mathcal{X}}_t).\end{aligned}$$

Our main result establishes the asymptotic convergence of $\widehat{\chi}_t^M$ with respect to time, under the assumptions similar to the ones used in the statement of Theorem 6.5.

Proposition 7.1. *Suppose that Assumption 6.1 and Assumption 6.2 hold. For any number of particles M, it holds that*

$$\|\widehat{\chi}_t^M\| \to 0 \text{ as } t \to \infty.$$

Proof. We analyze the asymptotic behavior of $\widehat{\chi}_t^M$ by calculating the derivative of the following weighted quadratic function:

$$\widehat{\chi}_t^M \mapsto \mathcal{W}(\widehat{\chi}_t^M) := \widehat{\chi}_t^M (\mathcal{Q}_t^M)^{-1} \widehat{\chi}_t^M.$$

We observe that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}(\mathcal{Q}_t^M)^{-1} &= -(\mathcal{Q}_t^M)^{-1}\dot{\mathcal{Q}}_t^M(\mathcal{Q}_t^M)^{-1}\\ &= -(\mathcal{Q}_t^M)^{-1}(A-\lambda\mathcal{L}_t^M C) - (A-\lambda\mathcal{L}_t^M C)^\top(\mathcal{Q}_t^M)^{-1} - \lambda C^\top(C\mathcal{Q}_t^M C^\top + V)^{-1}C\\ &\quad - (\mathcal{Q}_t^M)^{-1}\left(\left(1-\frac{\mathbf{1}_\omega}{M}\right)BB^\top + \frac{\lambda\mathbf{1}_v}{M}\mathcal{L}_t^M V(\mathcal{L}_t^M)^\top\right)(\mathcal{Q}_t^M)^{-1}.\end{aligned}$$

Consequently, this last expression yields

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}(\widehat{\chi}_t^M)^\top(\mathcal{Q}_t^M)^{-1}\widehat{\chi}_t^M &= (\widehat{\chi}_t^M)^\top\frac{\mathrm{d}}{\mathrm{d}t}(\mathcal{Q}_t^M)^{-1}\widehat{\chi}_t^M + (\widehat{\chi}_t^M)^\top(\mathcal{Q}_t^M)^{-1}\dot{\widehat{\chi}}_t^M + (\dot{\widehat{\chi}}_t^M)^\top(\mathcal{Q}_t^M)^{-1}\widehat{\chi}_t^M\\ &= -(\widehat{\chi}_t^M)^\top\Big[\lambda C^\top(C\mathcal{Q}_t^M C^\top + V)^{-1}C\\ &\quad + (\mathcal{Q}_t^M)^{-1}\left(\left(1-\frac{\mathbf{1}_\omega}{M}\right)BB^\top + \frac{\lambda\mathbf{1}_v}{M}\mathcal{L}_t^M V(\mathcal{L}_t^M)^\top\right)(\mathcal{Q}_t^M)^{-1}\Big]\widehat{\chi}_t^M + 2(\widehat{\chi}_t^M)^\top(\mathcal{Q}_t^M)^{-1}u_t,\end{aligned}$$

where for brevity, we used the notation $u_t := \lambda(\mathcal{L}_t^{\text{opt}} - \mathcal{L}_t^M)C(\mathsf{E}[X_t] - \widehat{\mathcal{X}}_t)$.

Fix an arbitrary $\tau > 0$. For each $t \geq \tau$, we can write

$$\int_{t-\tau}^t \frac{\mathrm{d}}{\mathrm{d}s}\mathcal{W}(\widehat{\chi}_s^M)\mathrm{d}s \leq \int_{t-\tau}^t -(\widehat{\chi}_s^M)^\top C^\top(C\bar{\mathcal{Q}}_0C^\top + V)^{-1}C\widehat{\chi}_s^M\mathrm{d}s + 2\int_{t-\tau}^t(\widehat{\chi}_s^M)^\top(\mathcal{Q}_s^M)^{-1}u_s\mathrm{d}s \tag{55}$$

In the last display equation, the right-hand side can be seen as a sum of two terms: in the first term, the integrand is negative definite and is quadratic in $\widehat{\chi}_t^M$; and in the second term, the integrand is linear in $\widehat{\chi}_t^M$, but contains $(\mathcal{L}_t^{\text{opt}} - \mathcal{L}_t^M)$ and $(\mathsf{E}[X_t] - \widehat{\mathcal{X}}_t)$, which have nice convergence properties.

To analyze the right-hand side of (55), we observe that for each $s \in [t-\tau, t]$, we have

$$\widehat{\chi}_s = \Phi(s, t-\tau)\widehat{\chi}_{t-\tau} + \int_{t-\tau}^s \Phi(s,r)u_r\mathrm{d}r \tag{56}$$

where $\Phi(t,t_0) = \exp\left(\int_{t_0}^t(A-\lambda\mathcal{L}_s^M C)\mathrm{d}s\right)$. The first term on the right-hand side of (55) then satisfies

$$\begin{aligned}&-\int_{t-\tau}^t(\widehat{\chi}_s^M)^\top C^\top(C\overline{\mathcal{Q}}_0C^\top + V)^{-1}C\widehat{\chi}_s^M\mathrm{d}s\\ &\quad= -\int_{t-\tau}^t\left(\Phi(s,t-\tau)\widehat{\chi}_{t-\tau}^M + \int_{t-\tau}^s\Phi(s,r)u_r\mathrm{d}r\right)^\top C^\top(C\overline{\mathcal{Q}}_0C^\top + V)^{-1}C\\ &\qquad\qquad\left(\Phi(s,t-\tau)\widehat{\chi}_{t-\tau}^M + \int_{t-\tau}^s\Phi(s,r)u_r\mathrm{d}r\right)\mathrm{d}s\\ &\quad\leq -(\widehat{\chi}_{t-\tau}^M)^\top\overline{\mathcal{O}}(t,t-\tau)\widehat{\chi}_{t-\tau}^M - 2(\widehat{\chi}_{t-\tau}^M)^\top\int_{t-\tau}^t\Phi(s,t-\tau)^\top C^\top(C\overline{\mathcal{Q}}_0C^\top + V)^{-1}C\int_{t-\tau}^s\Phi(s,r)u_r\,\mathrm{d}r\,\mathrm{d}s\end{aligned} \tag{57}$$

where $\overline{\mathcal{O}}(t,t-\tau) = \int_{t-\tau}^t\Phi(s,t-\tau)C^\top(C\overline{\mathcal{Q}}_0C^\top + V)^{-1}C\Phi(s,t-\tau)\mathrm{d}s$, and the last negative-definite term was dropped in writing the inequality.

For the second term on the right-hand side of (55), using (56), we get,

$$\int_{t-\tau}^t(\widehat{\chi}_s^M)^\top(\mathcal{Q}_s^M)^{-1}u_s\mathrm{d}s \leq (\widehat{\chi}_{t-\tau}^M)^\top\int_{t-\tau}^t\Phi(s,t-\tau)^\top(\mathcal{Q}_s^M)^{-1}u_s\mathrm{d}s + \int_{t-\tau}^t u_s^\top(\mathcal{Q}_s^M)^{-1}\int_{t-\tau}^s\Phi(s,r)u_r\mathrm{d}r\mathrm{d}s. \tag{58}$$

Before substituting (57) and (58) in (55), we recall some bounds on the variables appearing in these expressions.

1. Using the upper and lower bound on $\mathcal{Q}_t^M$, that is, $\rho I \leq \mathcal{Q}_t^M \leq \overline{\mathcal{Q}}_0$, we get

$$\mu_{\min}(\overline{\mathcal{Q}}_0^{-1})\|\widehat{\chi}_t^M\|^2 \leq (\widehat{\chi}_t^M)^\top \overline{\mathcal{Q}}_0^{-1} \widehat{\chi}_t^M \leq (\widehat{\chi}_t^M)^\top (\mathcal{Q}_t^M)^{-1} \widehat{\chi}_t^M \leq \rho^{-1}\|\widehat{\chi}_t^M\|^2.$$

2. Due to the observability assumption, there exists a positive constant η

$$0 < \eta I \leq \overline{\mathcal{O}}(t, t-\tau) \qquad \forall t > \tau.$$

   Hence, we get,

$$-(\widehat{\chi}_{t-\tau}^M)^\top \overline{\mathcal{O}}(t, t-\tau)\widehat{\chi}_{t-\tau}^M \leq -\eta(\widehat{\chi}_{t-\tau}^M)^\top \widehat{\chi}_{t-\tau}^M = -\eta\|\widehat{\chi}_{t-\tau}^M\|^2 \leq -\eta\rho(\widehat{\chi}_{t-\tau}^M)^\top (\mathcal{Q}_{t-\tau}^M)^{-1}\widehat{\chi}_{t-\tau}^M.$$

3. We choose $\overline{c}$ such that $\|C^\top (C\overline{\mathcal{Q}}_0 C^\top + V)^{-1}\| \leq \overline{c}$.

4. Note that the gain $\mathcal{L}_t^M$ is uniformly bounded due to Lemma 6.3 and the boundedness of $\mathcal{Q}_t^M$. This, consequently, provides a bound on the flow matrix $\Phi(s,t)$. We introduce the constant $a_{\mathrm{cl}} := \max_{s \geq 0} \|A - \lambda \mathcal{L}_s^M C\|$, so that one possible choice for a_{cl} is $\|A\| + \lambda\|\overline{\mathcal{Q}}_0\| \cdot \|C^\top V^{-1} C\|$. Then, it follows that

$$\|\Phi(s,t)\| \leq a_{\mathrm{cl}}(t-s).$$

Plugging the expressions (57) and (58) in (55), and using the aforementioned bounds on individual terms, we get

$$\mathcal{W}(\widehat{\chi}_t) - \mathcal{W}(\widehat{\chi}_{t-\tau}) \leq -\eta\rho\mathcal{W}(\widehat{\chi}_{t-\tau}) + 2\|\widehat{\chi}_{t-\tau}^M\|\tau^2 a_{\mathrm{cl}}(\tau^2 a_{\mathrm{cl}} + \rho)\|u_{[t-\tau,t]}\|_\infty + 2\rho\tau^4 a_{\mathrm{cl}}\|u_{[t-\tau,t]}\|_\infty^2. \tag{59}$$

For the second term on the right-hand side of (59), we apply Young's inequality as follows:

$$\begin{aligned}
&2\left(\|\widehat{\chi}_{t-\tau}^M\| \cdot \sqrt{\frac{\eta\rho}{2\mu_{\max}\overline{\mathcal{Q}}}}\right) \cdot \left(\sqrt{\frac{2\mu_{\max}\overline{\mathcal{Q}}}{\eta\rho}} \cdot \tau^2 a_{\mathrm{cl}}(\tau^2 a_{\mathrm{cl}} + \rho)\|u_{[t-\tau,t]}\|_\infty\right) \\
&\leq \|\widehat{\chi}_{t-\tau}^M\|^2 \cdot \frac{\eta\rho}{2\mu_{\max}\overline{\mathcal{Q}}} + \frac{2\mu_{\max}\overline{\mathcal{Q}}}{\eta\rho}\tau^4 a_{\mathrm{cl}}^2(\tau^2 a_{\mathrm{cl}} + \rho)^2\|u_{[t-\tau,t]}\|_\infty^2 \\
&\leq \frac{\eta\rho}{2}\mathcal{W}(\widehat{\chi}_{t-\tau}) + \frac{2\mu_{\max}\overline{\mathcal{Q}}}{\eta\rho}\tau^4 a_{\mathrm{cl}}^2(\tau^2 a_{\mathrm{cl}} + \rho)^2\|u_{[t-\tau,t]}\|_\infty^2.
\end{aligned}$$

Taking $\widehat{\rho} := 1 - \frac{\eta\rho}{2}$ and $\sigma(\tau) = \frac{2\mu_{\max}\overline{\mathcal{Q}}}{\eta\rho}\tau^4 a_{\mathrm{cl}}^2(\tau^2 a_{\mathrm{cl}} + \rho)^2 + 2\rho\tau^4 a_{\mathrm{cl}}$, (59) results in

$$\mathcal{W}(\widehat{\chi}_t) \leq \widehat{\rho}\mathcal{W}(\widehat{\chi}_{t-\tau}) + \sigma(\tau)\|u_{[t-\tau,t]}\|_\infty^2. \tag{60}$$

We next show that $\|u_t\|$ converges to zero as $t \to \infty$, which in turn implies that $\|u_{[t-\tau,t]}\|_\infty^2$ converges to zero, as $t \to \infty$. Towards this end, we observe that[1]

$$\begin{aligned}
\|u_t\| &\leq \lambda\|(I - \mathcal{L}_t^M C)\| \cdot \|\mathcal{P}_t - \mathcal{Q}_t^M\| \cdot \|C^\top\| \cdot \|(C\mathcal{P}_t C^\top + V)^{-1}\| \cdot \|C\| \cdot \|\mathsf{E}[X_t] - \widehat{\mathcal{X}}_t\| \\
&\leq \lambda\kappa_{\overline{\mathcal{Q}}_0}\|\mathcal{P}_t - \mathcal{Q}_t^M\| \cdot \|C\|^2 \cdot \|V^{-1}\| \cdot \|\mathsf{E}[X_t] - \widehat{\mathcal{X}}_t\|,
\end{aligned} \tag{61}$$

where $\kappa_{\overline{\mathcal{Q}}_0}$ is obtained from Lemma 6.4. By Theorem 6.5, $\|\mathcal{P}_t - \mathcal{Q}_t^M\|$ is bounded. Since $\frac{\mathrm{d}}{\mathrm{d}t}(\mathsf{E}[X_t] - \widehat{\mathcal{X}}_t) = (A - \lambda\mathcal{L}_t^{\mathrm{opt}} C)(\mathsf{E}[X_t] - \widehat{\mathcal{X}}_t)$, it follows from [7, Theorem 3] that for some $\overline{t} > 0$, and $t \geq \overline{t}$

$$\begin{aligned}
\|\mathsf{E}[X_t] - \widehat{\mathcal{X}}_t\| &\leq \|\mathsf{E}[X_{\overline{t}}] - \widehat{\mathcal{X}}_{\overline{t}}\| \exp\left(\int_0^t A - \lambda\mathcal{L}_s C \mathrm{d}s\right) \\
&\leq \|\mathsf{E}[X_{\overline{t}}] - \widehat{\mathcal{X}}_{\overline{t}}\| \frac{\underline{\rho}_C \underline{\rho}_O}{\left(1 + \underline{\rho}_C \overline{\rho}_O\right)\left(1 + \underline{\rho}_O \overline{\rho}_C\right)} \exp\left(\frac{-\zeta\underline{\rho}_C t}{1 + \underline{\rho}_C \overline{\rho}_O}\right)
\end{aligned}$$

for some $\zeta > 0$ and positive constants $\underline{\rho}_C, \overline{\rho}_C, \underline{\rho}_O, \overline{\rho}_O$. In particular, this estimate establishes the convergence of $\|u_t\|$ to 0, as t grows.

The proof of Proposition 7.1 can now be completed by checking that the inequality (60) results in $\mathcal{W}(\widehat{\chi}_t^M)$ converging to zero whenever $\|u_{[t-\tau,t]}\|_\infty$ converges to zero. □

[1] Recall that $\mathcal{L}_t^{\mathrm{opt}} - \mathcal{L}_t^M = (I - \mathcal{L}_t^M C)(\mathcal{P}_t - \mathcal{Q}_t^M)C^\top (C\mathcal{P}_t C^\top + V)^{-1}$ by Lemma 6.3.

8. Conclusions

We considered the design of ensemble filters for linear dynamical systems with time-sampled observations. These classes of filters are described by diffusion processes which are coupled via empirical mean and variance. Our primary objective was to analyze the asymptotic behavior of these filters and compare it with their optimal counterparts. Taking the expectation with respect to Poisson sampling process leads to a new class of differential equations for the variance, and we provide a rigorous treatment of its asymptotic behavior.

Several interesting questions arise from the study carried out in this paper. Firstly, it would be interesting to see if Assumption 6.1 could be removed and the analysis techniques could be adapted to unify the study of convergence towards a steady state together with developing comparisons with the optimal solution. We adopted this route in the scalar setting [30] and the advantage of that approach is that it gives explicit bounds on the mean sampling that characterize the asymptotic performance. The second direction that could be adopted is to take a step towards nonlinear ensemble filters. The design methodology proposed here provides a template for such filters in nonlinear setting: the prediction is based on propagating system dynamics and the gain for the correction is obtained via empirical moments. However, theoretical analysis in this case is expected to be more challenging and provides an interesting direction for further research.

References

[1] K. Bergemann and S. Reich. A mollified ensemble Kalman filter. *Q.J. Royal Meteorological Society*, 136(651):1636–1643, 2010.

[2] A. N. Bishop and P. Del Moral. On the mathematical theory of ensemble (linear-Gaussian) Kalman–Bucy filtering. *Math. Control, Signals and Systems*, 35:835–903, 2023.

[3] A.N. Bishop and P. Del Moral. On the stability of matrix-valued Riccati diffusions. *Elec. J. of Probability*, 24(84):1–40, 2019.

[4] E. Calvello, S. Reich, and A.M. Stuart. Ensemble Kalman methods: A mean field perspective. *Acta Numerica*, 34:123–291, 2025.

[5] Z. Chen. Bayesian filtering: From Kalman filters to particle filters, and beyond. *Statistics*, 182(1):1–69, 2003.

[6] D. Crisan and B. Rozovskii. *The Oxford Handbook of Nonlinear Filtering*. Oxford University Press, 2011.

[7] A. Dasgupta and A. Tanwani. Steady state analysis for Kalman filters with Poisson-sampled observations. *Systems & Control Letters*, 204:106209, 2025.

[8] J. de Wiljes, S. Reich, and W. Stannat. Long-time stability and accuracy of the ensemble Kalman–Bucy filter for fully observed processes and small measurement noise. *SIAM Journal on Applied Dynamical Systems*, 17(2):1152–1181, 2018.

[9] P. Del Moral and J. Tugaut. On the stability and the uniform propagation of chaos properties of ensemble Kalman–Bucy filters. *The Annals of Applied Probability*, 28(2):790–850, 2018.

[10] A. Doucet, N. De Freitas, and N. Gordon, editors. *Sequential Monte Carlo methods in practice*. Springer, 1 edition, 2001.

[11] G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99:10143–10162, 1994.

[12] G. Evensen. *Data Assimilation: The Ensemble Kalman Filter*. Springer Science & Business Media, 2nd edition, 2009.

[13] J. P. Hespanha and A. R. Teel. Stochastic impulsive systems driven by renewal processes. In *Proc. 17th International Symposium on Mathematical Theory of Networked Systems*, pages 606–618. 2006.

[14] O. Iufereva. *Filtering algorithms with Poisson-distributed observations*. PhD thesis, Université de Toulouse; L'université fédérale de l'Oural, 2024.

[15] A.H. Jazwinski. *Stochastic processes and filtering theory*. Dover Publications, Inc., New York, 2007.

[16] M. Lambert, S. Bonnabel, and F. Bach. The continuous-discrete variational Kalman filter (CD-VKF). In *2022 IEEE 61st Conference on Decision and Control (CDC)*, pages 6632–6639, 2022.

[17] F. Le Gland, V. Monbet, and V.-D. Tran. Large sample asymptotics for the ensemble Kalman filter. *The Oxford Handbook of Nonlinear Filtering*, pages 598–631, 2011.

[18] B. Øksendal. *Stochastic differential equations*. Springer, 2003.

[19] P. E. Protter. *Stochastic Integration and Differential Equations*. Springer, 2nd edition, 2005.

[20] S. Reich and C.J. Cotter. Ensemble filter techniques for intermittent data assimilation. In M. Cullen, M.A. Freitag, S. Kindermann, and R. Scheichl, editors, *Large Scale Inverse Problems: Computational Methods and Applications in the Earth Sciences*, pages 91–134. De Gruyter, Berlin, Boston, 2013.

[21] B. Ristic, S. Arulampalam, and N. Gordon. *Beyond the Kalman filter: Particle filters for tracking applications*. Artech house, 2003.

[22] S. Särkkä and L. Svensson. *Bayesian Filtering and Smoothing*. Cambridge University Press, second edition, 2023.

[23] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. I. Jordan, and S. S. Sastry. Kalman filtering with intermittent observations. *IEEE Transactions on Automatic Control*, 49(9):1453–1464, 2004.

[24] Y. Suhov and M. Kelbert. *Probability and Statistics by Example. II.* Cambridge University Press, Cambridge, 2008. Markov chains: a primer in random processes and their applications.

[25] A.-S. Sznitman. Topics in propagation of chaos. *Ecole d'été de probabilités de Saint-Flour XIX–1989*, 1464:165–251, 1991.

[26] A. Taghvaei and P. G. Mehta. A survey of feedback particle filter and related controlled interacting particle systems (CIPS). *Annual Reviews in Control*, 55:356–378, 2023.

[27] A. Taghvaei and P.G. Mehta. An optimal transport formulation of the ensemble Kalman filter. *IEEE Transactions on Automatic Control*, 66:3052–3067, 2020.

[28] A. Tanwani, D. Chatterjee, and D. Liberzon. Stabilization of continuous-time deterministic systems under random sampling: Overview and recent developments. In T. Başar, editor, *Uncertainty in Complex Networked Systems*, pages 209–246. Springer Nature, 2018.

[29] A. Tanwani and O. Yufereva. Error covariance bounds for suboptimal filters with Lipschitzian drift and Poisson-sampled measurements. *Automatica*, 122, 2020.

[30] A. Tanwani and O. Yufereva. Convergence analysis of ensemble filters for linear stochastic systems with Poisson-sampled observations. *IFAC-PapersOnLine*, 58(17):202–207, 2024. Proceedings of 25th Intl. Symp. on Mathematical Theory of Networks and Systems.

[31] T. Yang, P. G. Mehta, and S. P. Meyn. Feedback particle filter. *IEEE Transactions on Automatic Control*, 58(10):2465–2480, 2013.

[32] O. Yufereva and A. Tanwani. Transport inspired particle filters with Poisson-sampled observations in Gaussian setting. In *Proceedings of 62nd IEEE Conference on Decision and Control (CDC)*, pages 7695–7700, 2023.

A. Tools for Analysis

Consider the following differential equation for a stochastic process $(x_t)_{t\geq 0}$ evolving in $\mathbb{R}^n$:

$$\mathrm{d}x_t = f(x_t)\mathrm{d}t + g(x_t)\mathrm{d}\omega + h(x_t,\nu_t)\mathrm{d}N_t \tag{A.1}$$

where N_t is a Poisson process with intensity $\lambda > 0$, ω is a standard Wiener process, and $\nu_t := \nu_{N_t}$ for the sake of brevity with ν_{N_t} being a sequence of i.i.d. random variables with probability law ξ.

To study the evolution of a function of the random process $(x_t)_{t\geq 0}$, we make use of the Ito's chain rule. The reader may consult [19, Chapter II, Section 7] for detailed exposition on this topic. Here, the particular form we adopt is tailored for the differential equations appearing in earlier sections.

Proposition A.1 (Ito's chain rule)**.** *For a twice continuously differentiable function $\psi : \mathbb{R}^n \to \mathbb{R}$, it holds that*

$$\begin{aligned}\mathrm{d}\psi(x_t) = \Big[\langle \nabla\psi(x_t), f(x_t)\rangle + \frac{1}{2}\mathrm{tr}\Big(\frac{\partial^2\psi(x_t)}{\partial x} g(x_t)g(x_t)^\top\Big)\Big]\mathrm{d}t \\ + \langle \nabla\psi(x_t), g(x_t)\rangle\mathrm{d}\omega + \Big[\psi(x_t + h(x,\nu_t)) - \psi(x_t)\Big]\mathrm{d}N_t.\end{aligned} \tag{A.2}$$

Ito's chain rule describes the evolution of the function ψ evaluated along the solution of the stochastic differential equation (A.1). However, to describe the evolution of the expectation of ψ in differential form, we need to consider the extended generator as defined below:

Definition A.2 (Extended generator)**.** Given a real-valued function $\psi : \mathbb{R}^n \to \mathbb{R}$, the *extended generator* of the process $\big(x_t\big)_{t\geq 0}$ described by (A.1) is the linear operator $\psi \mapsto \mathcal{L}\psi$ defined by

$$\begin{gathered}\mathbb{R}^n \ni z \mapsto \mathcal{L}\psi(z) \in \mathbb{R} \\ \mathcal{L}\psi(z) := \lim_{\varepsilon\downarrow 0} \frac{1}{\varepsilon}\Big(\mathsf{E}\big[\psi\big(x(t+\varepsilon)\big) \,\big|\, x(t) = z\big] - \psi(z)\Big).\end{gathered} \tag{A.3}$$

We obtain the expected value of ψ by integrating the generator, which can be seen as a generalization of the classical Dynkin's formula:

$$\mathsf{E}\big[\psi(x(t))\big] = \mathsf{E}\big[\psi(x(0))\big] + \mathsf{E}\Big[\int_0^t \mathcal{L}\psi(x(s))\,\mathrm{d}s\Big]. \tag{A.4}$$

For our purposes, it is useful to compute an explicit expression of the generator which can then be analyzed for studying the qualitative behavior of $\mathsf{E}[\psi(x)]$. Several references in the literature derive generator equations for stochastic processes with jumps, see for example [13, Theorem 1] for a derivation in the context of Wiener process driven differential equations with renewal processes. The proof builds on the simpler case given in [28, Proposition 2.1] $g \equiv 0$ and h being independent of the noise at jump instants ν_t.

Proposition A.3. *If the sampling process $(N_t)_{t\geq 0}$ is Poisson with intensity $\lambda > 0$, then the process $\big(x(t)\big)_{t\geq 0}$ described in* (A.1) *is Markovian. Moreover, for any function $\mathbb{R}^n \ni z \mapsto \psi(z) \in \mathbb{R}$ with at most polynomial growth as $\|z\| \to +\infty$, we have*

$$\mathcal{L}\psi(z) = \langle \nabla\psi(z), f(z)\rangle + \frac{1}{2}\mathrm{tr}\Big(\frac{\partial^2\psi(z)}{\partial z^2} g(z)g(z)^\top\Big) + \lambda\Big[\int \psi\big(z + h(z,\nu)\big)\xi(\mathrm{d}\nu) - \psi(z)\Big]. \tag{A.5}$$

Proof. The fact that $\big(x(t)\big)_{t\geq 0}$ is Markovian follows from the observation that the future of $x(t)$ depends on $x(\tau_{N_t})$.

Since the system under consideration is well-posed, we have, for $\varepsilon > 0$ small,

$$\mathsf{E}\big[\psi\big(x(t+\varepsilon)\big)\,\big|\,x(t) = z\big] = \mathsf{E}\big[\psi\big(x(t+\varepsilon)\big)\big(1_{\{N_{t+\varepsilon}=N_t\}} + 1_{\{N_{t+\varepsilon}=1+N_t\}} + 1_{\{N_{t+\varepsilon}-N_t\geq 2\}}\big)\,\big|\,x(t)\big]. \tag{A.6}$$

We now compute the conditional probability distribution of $\big(x(t+\varepsilon)\big)$ for small $\varepsilon > 0$ given $x(t)$. Since the sampling process is independent of the state joint, by definition of the sampling (Poisson) process we have, for $\varepsilon \downarrow 0$,

$$\begin{cases} \mathsf{P}\big(N_{t+\varepsilon} - N_t = 0\,\big|\,N_t, x(t)\big) = 1 - \lambda\varepsilon + o(\varepsilon),\\ \mathsf{P}\big(N_{t+\varepsilon} - N_t = 1\,\big|\,N_t, x(t)\big) = \lambda\varepsilon + o(\varepsilon),\\ \mathsf{P}\big(N_{t+\varepsilon} - N_t \geq 2\,\big|\,N_t, x(t)\big) = o(\varepsilon). \end{cases}$$

Using these expressions, we develop (A.6) further for $\varepsilon \downarrow 0$ as

$$\begin{aligned} \mathsf{E}\big[\psi\big(x(t+\varepsilon)\big)\,\big|\,x(t) = z\big] &= \mathsf{E}\big[\psi\big(x(t+\varepsilon)\big)\big(1_{\{N_{t+\varepsilon}=N_t\}} + 1_{\{N_{t+\varepsilon}=1+N_t\}}\big)\,\big|\,x(t)\big] + o(\varepsilon)\\ &= \mathsf{E}\big[\psi\big(x(t+\varepsilon)\big)\,\big|\,x(t), N_{t+\varepsilon} = N_t\big]\cdot\big(1 - \lambda\varepsilon + o(\varepsilon)\big)\\ &\quad + \mathsf{E}\big[\psi\big(x(t+\varepsilon)\big)\,\big|\,x(t), N_{t+\varepsilon} = 1 + N_t\big]\big(\lambda\varepsilon\big) + o(\varepsilon). \end{aligned} \tag{A.7}$$

The two significant terms on the right-hand side of (A.7) are now computed separately. First, for $\varepsilon \downarrow 0$, we have

$$\mathsf{E}\big[\psi\big(x(t+\varepsilon)\big)\,\big|\,N_{t+\varepsilon} = N_t, x(t) = z\big] = \psi(z) + \varepsilon\,\big\langle\nabla\psi\big(z\big), f\big(z\big)\big\rangle + 0.5\varepsilon\bar{g}(z) + o(\varepsilon),$$

leading to the first term on the right-hand side of (A.7) having the estimate

$$\begin{aligned} &\mathsf{E}\big[\psi\big(x(t+\varepsilon)\big)\,\big|\,N_{t+\varepsilon} = N_t, x(t) = z\big]\cdot\big(1 - \lambda\varepsilon + o(\varepsilon)\big)\\ &\qquad\qquad = \psi\big(z\big) + \varepsilon\,\big\langle\nabla\psi\big(z\big), f\big(z\big)\big\rangle + \varepsilon\bar{g}(z) - (\lambda\varepsilon)\psi\big(z\big) + o(\varepsilon), \end{aligned}$$

where $\bar{g}(z) := 0.5\mathrm{tr}\big((\partial^2\psi(z)/\partial z)g(z)g(z)^\top\big)$. Concerning the second term on the right-hand side of (A.7), we observe that conditional on $N_{t+\varepsilon} = 1 + N_t$, the probability distribution of $\tau_{N_{t+\varepsilon}}$ is [24, Theorem 2.3.7] uniform over $[t, t+\varepsilon[$ by definition of the sampling (Poisson) process, i.e.,

$$\mathsf{P}\big(\tau_{N_{t+\varepsilon}} \in [s, s+s'[\,\big|\,N_{t+\varepsilon} = 1 + N_t\big) = \frac{1}{\varepsilon}s'$$

for $[s, s+s'[\, \subset [t, t+\varepsilon[$. Since the sampling process is independent of the state process, the preceding conditional probability is equal to

$$\mathsf{P}\big(\tau_{N_{t+\varepsilon}} \in [s, s+s'[\,\big|\,N_{t+\varepsilon} = 1 + N_t, x(t) = z\big).$$

We define $\theta \in [0, 1[$ such that $\tau_{N_{t+\varepsilon}} = t + \theta\varepsilon, x(t) = z$; then θ is uniformly distributed on $[0, 1[$ given $N_{t+\varepsilon} = 1 + N_t$. We also have, conditioned on the same event,

$$x(\tau_{N_{t+\varepsilon}}) = x(t+\theta\varepsilon) = x(t+\theta\varepsilon^-) + h(x(t+\theta\varepsilon^-), \nu),$$

and

$$x(t+\theta\varepsilon^-) = x(t) + \theta\varepsilon\bar{f}\big(x(t)\big) + o(\varepsilon),$$

where $\bar{f}(z)$ denotes the linear interpolation of the solution of (A.1) over an interval of length ε starting from z. The above expressions then lead to, conditioned on the event $N_{t+\varepsilon} = 1 + N_t, x(t) = z$ and for $\varepsilon \downarrow 0$,

$$x(t+\varepsilon) = x(t+\theta\varepsilon) + (1-\theta)\varepsilon\bar{f}\big(x(t+\theta\varepsilon)\big)\big) + o(\varepsilon)$$

$$= x(t) + \theta\varepsilon\bar{f}\big(x(t)\big) + h(x(t) + \theta\varepsilon\bar{f}\big(x(t)\big), \nu) + (1-\theta)\varepsilon\bar{f}\big(x(t+\theta\varepsilon)\big) + o(\varepsilon)$$
$$= x(t) + \theta\varepsilon\bar{f}\big(x(t)\big) + h(x(t), \nu) + (1-\theta)\varepsilon\bar{f}\big(x(t+\theta\varepsilon)\big) + \theta\varepsilon\nabla_x h(x(t), \nu)\cdot\bar{f}(x(t)) + o(\varepsilon).$$

Therefore, for $\varepsilon \downarrow 0$,

$$\begin{aligned}
&\mathsf{E}\big[\psi\big(x(t+\varepsilon)\big)\,\big|\,x(t) = z, N_{t+\varepsilon} = 1 + N_t\big]\cdot(\lambda\varepsilon)\\
&= \int_0^1 \psi\Big(z + \theta\varepsilon\bar{f}(z) + h(z,\nu) + (1-\theta)\varepsilon\bar{f}\big(x(t+\theta\varepsilon)\big) + \theta\varepsilon\nabla_z h(z,\nu)\cdot\bar{f}(z) + o(\varepsilon)\Big)\,\xi(\mathrm{d}\nu)\mathrm{d}\theta\cdot(\lambda\varepsilon)\\
&= \int_0^1 \Big(\psi\big(z + h(z,\nu)\big) + \varepsilon\langle\nabla\psi\big(z + h(z,\nu)\big), o(\varepsilon) + \theta\bar{f}(z)\\
&\quad + (1-\theta)\bar{f}(x(t+\theta\varepsilon)) + \theta\nabla_z h(z,\nu)\cdot\bar{f}(z)\rangle\Big)\,\xi(\mathrm{d}\nu)\mathrm{d}\theta\cdot(\lambda\varepsilon)\\
&= \Big(\int \psi\big(z + h(z,\nu)\xi(\mathrm{d}\nu) + O(\varepsilon)\Big)\cdot(\lambda\varepsilon)\\
&= (\lambda\varepsilon)\int \psi\big(z + h(z,\nu)\big)\xi(\mathrm{d}\nu) + o(\varepsilon).
\end{aligned}$$

Putting everything together, we arrive at

$$\begin{aligned}
\mathsf{E}\big[\psi\big(x(t+h)\big)\,\big|\,x(t) = z\big] = \psi(z) + \varepsilon\Big(\langle\nabla\psi(z), f(z)\rangle + \frac{1}{2}\mathrm{tr}\Big(\frac{\partial^2\psi(z)}{\partial z}g(z)g(z)^\top\Big)\Big)\\
- (\lambda\varepsilon)\big(\psi(z) - \int\psi(z + h(z,\nu))\xi(\mathrm{d}\nu)\big) + o(\varepsilon).
\end{aligned}$$

Substituting these expressions in (A.3), we see that for each $z \in \mathbb{R}^n$, we get the expression (A.5). □

B. Properties of Gain Matrix

In this part of the Appendix, we recall some useful properties of the injection gain that are useful for the analysis. In what follows, let $\mathcal{P}_1 \in \mathbb{R}^{n\times n}$ and $\mathcal{P}_2 \in \mathbb{R}^{n\times n}$ be positive semi-definite matrices, and $V \in \mathbb{R}^{p\times p}$ be positive definite. For any matrix $C \in \mathbb{R}^{p\times n}$, let $\mathcal{L}_1 := \mathcal{P}_1 C^\top(C\mathcal{P}_1C^\top + V)^{-1}$, $\mathcal{L}_2 := \mathcal{P}_2C^\top(C\mathcal{P}_2C^\top + V)^{-1}$.

Lemma B.1 ([7, Lemmata 3 & 4]). *The matrices $\mathcal{L}_1$ and $\mathcal{L}_2$ satisfy the following properties:*

1. $(I - \mathcal{L}_1C)P_1C^\top = \mathcal{L}_1V$
2. $(I - \mathcal{L}_1C)\mathcal{P}_1 - (I - \mathcal{L}_2C)\mathcal{P}_2 = (I - \mathcal{L}_1C)(\mathcal{P}_1 - \mathcal{P}_2)(I - \mathcal{L}_2C)^\top$

Lemma B.2. *The matrices $\mathcal{L}_1$ and $\mathcal{L}_2$ satisfy the following inequalities*

$$(I - \mathcal{L}_1C)(\mathcal{P}_1 - \mathcal{P}_2)(I - \mathcal{L}_2C)^\top \geq (I - \mathcal{L}_1C)(\mathcal{P}_1 - \mathcal{P}_2)(I - \mathcal{L}_1C)^\top$$
$$(I - \mathcal{L}_1C)(\mathcal{P}_1 - \mathcal{P}_2)(I - \mathcal{L}_2C)^\top \leq (I - \mathcal{L}_2C)(\mathcal{P}_1 - \mathcal{P}_2)(I - \mathcal{L}_2C)^\top$$

Proof. We keep the same proof that one can find in [7]; the first inequality follows from the chain below and the second one can be proven similarly.

$$\begin{aligned}
&(I - \mathcal{L}_1C)(\mathcal{P}_1 - \mathcal{P}_2)(I - \mathcal{L}_2C)^\top - (I - \mathcal{L}_1C)(\mathcal{P}_1 - \mathcal{P}_2)(I - \mathcal{L}_1C)^\top\\
&\quad = (I - \mathcal{L}_1C)(\mathcal{P}_1 - \mathcal{P}_2)C^\top(\mathcal{L}_1 - \mathcal{L}_2)^\top\\
&\quad = ((I - \mathcal{L}_1C)\mathcal{P}_1 - (I - \mathcal{L}_1C)\mathcal{P}_2)C^\top(\mathcal{L}_1 - \mathcal{L}_2)^\top\\
&\quad = (I - \mathcal{L}_1C)\mathcal{P}_1C^\top(\mathcal{L}_1 - \mathcal{L}_2)^\top - (I - \mathcal{L}_1C)\mathcal{P}_2C^\top(\mathcal{L}_1 - \mathcal{L}_2)^\top\\
&\quad \overset{\text{(a)}}{\geq} (I - \mathcal{L}_2C)\mathcal{P}_2C^\top(\mathcal{L}_1 - \mathcal{L}_2)^\top - (I - \mathcal{L}_1C)\mathcal{P}_2C^\top(\mathcal{L}_1 - \mathcal{L}_2)^\top\\
&\quad = ((I - \mathcal{L}_2C)\mathcal{P}_2 - (I - \mathcal{L}_1C)\mathcal{P}_2)C^\top(\mathcal{L}_1 - \mathcal{L}_2)^\top\\
&\quad = (\mathcal{L}_1 - \mathcal{L}_2)C\mathcal{P}_2C^\top(\mathcal{L}_1 - \mathcal{L}_2)^\top \geq 0,
\end{aligned}$$

where inequality (a) follows from Lemma B.1 since $(I - \mathcal{L}_1C)\mathcal{P}_1C^\top(\mathcal{L}_1 - \mathcal{L}_2)^\top - (I - \mathcal{L}_2C)\mathcal{P}_2C^\top(\mathcal{L}_1 - \mathcal{L}_2)^\top = \mathcal{L}_1V(\mathcal{L}_1 - \mathcal{L}_2)^\top - \mathcal{L}_2V(\mathcal{L}_1 - \mathcal{L}_2)^\top = (\mathcal{L}_1 - \mathcal{L}_2)V(\mathcal{L}_1 - \mathcal{L}_2)^\top \geq 0$. □

C. Lower Bound

We denote the error covariance flow associated with the standard Kalman–Bucy error covariance $\mathcal{R}^\tau$ for a continuous time system described by $(A, C, B, \tau, V, 1/M)$ and with initial error covariance Q as:

$$\dot{\mathcal{R}}_t^\tau = A\mathcal{R}_t^\tau + \mathcal{R}_t^\tau A^\top + (1 - 1/M)BB^\top - \frac{1}{\tau}\mathcal{R}_t^\tau C^\top V^{-1} C\mathcal{R}_t^\tau. \tag{C.1}$$

Proposition C.1. *Let $\mathcal{P}_t$ be the solution of* (35)*, and $\mathcal{Q}_t^M$ be the solution to* (33)*. For every symmetric positive definite Q satisfying $\mathcal{R}_0^\tau = \mathcal{P}_0 = \mathcal{Q}_0^M = Q$, and $\tau \leq \frac{1}{\lambda}$, it holds that*

$$\mathcal{P}_t \geq \mathcal{R}_t^\tau, \ \mathcal{Q}_t^M \geq \mathcal{R}_t^\tau, \quad \textit{for all} \quad t \geq 0.$$

In addition, there is a positive number ρ, such that

$$\mathcal{P}_t \geq \rho I, \ \mathcal{Q}_t^M \geq \rho I, \quad \textit{for all} \quad t \geq 0.$$

Proof. Notice that[2],

$$\begin{aligned}
\dot{\mathcal{P}}_t - \dot{\mathcal{R}}_t^\tau &\geq A\mathcal{P}_t + \mathcal{P}A^\top - \lambda\mathcal{P}_t C^\top(C\mathcal{P}_t C^\top + V)^{-1}C\mathcal{P}_t - A\mathcal{R}_t^\tau - \mathcal{R}_t^\tau A^\top + \frac{1}{\tau}\mathcal{R}_t^\tau C^\top V^{-1} C\mathcal{R}_t^\tau \\
&\geq A\mathcal{P}_t + \mathcal{P}_t A^\top - \lambda\mathcal{P}_t C^\top(V)^{-1}C\mathcal{P}_t - A\mathcal{R}_t^\tau - \mathcal{R}_t^\tau A^\top + \frac{1}{\tau}\mathcal{R}_t^\tau C^\top V^{-1} C\mathcal{R}_t^\tau \\
&\geq A\mathcal{P}_t + \mathcal{P}A^\top - \lambda\mathcal{P}C^\top V^{-1}C\mathcal{P}_t - A\mathcal{R}_t^\tau - \mathcal{R}_t^\tau A^\top + \lambda\mathcal{R}_t^\tau C^\top V^{-1} C\mathcal{R}_t^\tau \\
&= \left(A - \frac{\mathcal{P}_t + \mathcal{R}_t^\tau}{2}\lambda C^\top V^{-1} C\right)(\mathcal{P}_t - \mathcal{R}_t^\tau) + (\mathcal{P}_t - \mathcal{R}_t^\tau)\left(A - \frac{\mathcal{P}_t + \mathcal{R}_t^\tau}{2}\lambda C^\top V^{-1} C\right)^\top,
\end{aligned}$$

This differential inequality ensures that $\mathcal{P}_t \geq \mathcal{R}_t^\tau$ for all $t \geq 0$. A similar reasoning can be used to show that the empirical variance $\mathcal{Q}_t^M \geq \mathcal{R}_t^\tau$, for all cases.

We know that there is a uniform lower bound on $\mathcal{R}_t^\tau$, and this yields the uniform lower bounds on $\mathcal{P}_t$ and $\mathcal{Q}_t^M$. Let us define the controllability and observability Gramians associated with $\mathcal{R}_t^\tau$ as

$$\begin{aligned}
\mathcal{C}(t, t - t_0) &:= \int_{t-t_0}^{t} (1 - 1/M)e^{As}(BB^\top)e^{A^\top s}\, \mathrm{d}s, \\
\mathcal{O}^\tau(t, t - t_0) &:= \int_{t-t_0}^{t} e^{As}(C^\top(\tau V)^{-1}C)e^{A^\top s}\, \mathrm{d}s
\end{aligned}$$

respectively. If the continuous-time linear system defined by $(A, C, B, \tau, V, 1/M)$, $\tau \leq \frac{1}{\lambda}$ is uniformly completely controllable and uniformly completely observable, it means that there exists a $t_0 \geq 0$ such that for all $t \geq t_0$, we have

$$0 < \underline{\rho}_C I \leq \mathcal{C}(t, t - t_0) \leq \overline{\rho}_C I, \tag{C.2a}$$

$$0 < \underline{\rho}_O^\tau I \leq \mathcal{O}^\tau(t, t - t_0) \leq \overline{\rho}_O^\tau I, \tag{C.2b}$$

respectively. These inequalities lead to the following lower bound [15, Lemma 7.2]:

$$\mathcal{R}_t^\tau \geq \left(\mathcal{C}^{-1}(t, t - t_0) + \mathcal{O}^\tau(t, t - t_0)\right)^{-1} \geq \left(\frac{\underline{\rho}_C}{1 + \underline{\rho}_C \overline{\rho}_O^\tau}\right) I.$$

which consequently provide the lower bound on $\mathcal{P}_t$ and $\mathcal{Q}_t^M$, for each $t \geq 0$. □

[2] We use the trivial equality $MSM - NSN = \left(\frac{M+N}{2}S\right)(M - N) + (M - N)\left(S\frac{M+N}{2}\right)$, that holds for symmetric and positive semidefinite matrices M, N and S of appropriate dimension.

D. Application of Grönwall's Lemma

Lemma D.1. *Let $f : \mathbb{R}_+ \to \mathbb{R}_+$ be a given function, and $a > 0$ be a positive scalar, and b, c, d be some non-negative constants. If it holds that*

$$f(t) \leq \exp(-at)b + (1 - \exp(-at))c + d \int_0^t \exp(-a(t-s))f(s)\mathrm{d}s$$

then, with $\beta := d \exp\left(\frac{d}{a}\right)$, we have

$$f(t) \leq \exp(-at)(1 + \beta t)b + \frac{c\beta}{a}.$$

Proof. Applying Gronwall's inequality in the current setting, we get

$$\begin{aligned} f(t) \leq \exp(-at)b + (1 - \exp(-at))c + bd \exp(-at) \exp\left(\frac{d}{a}\right) \int_0^t \exp\left(\frac{-d \exp(-a(t-s))}{a}\right) \mathrm{d}s \\ + cd \int_0^t (1 - \exp(-as)) \exp(-a(t-s)) \exp\left(\frac{-d \exp(-a(t-s))}{a}\right) \end{aligned} \tag{D.1}$$

Note that $\exp\left(\frac{-d \exp(-a(t-s))}{a}\right) \leq 1$ as $t \geq s$. We can thus bound the second term in (D.1) as below:

$$bd \exp(-at) \exp\left(\frac{d}{a}\right) \int_0^t \exp\left(\frac{-d \exp(-a(t-s))}{a}\right) \mathrm{d}s \leq bd \exp(-at) \exp\frac{d}{a} =: b \exp(-at)\beta t.$$

It remains to find the bound on the last term in (D.1) only. For that, we observe that the integrand in the last term is less than $\exp(-a(t-s)) \exp\left(\frac{d}{a}\right)$, and thus, can be bounded from above by

$$cd \exp\left(\frac{d}{a}\right) \int_0^t \exp(-a(t-s))\mathrm{d}s \leq \frac{cd}{a} \exp\left(\frac{d}{a}\right) = \frac{c\beta}{a}$$

Using these bounds on the second and third term, (D.1) yields the desired result. □