



Convergence Analysis of Ensemble Filters for Linear Stochastic Systems with Poisson-Sampled Observations

Aneel Tanwani, Olga Yufereva

► To cite this version:

Aneel Tanwani, Olga Yufereva. Convergence Analysis of Ensemble Filters for Linear Stochastic Systems with Poisson-Sampled Observations. International Symposium on Mathematical Theory of Networks and Systems, Aug 2024, Cambridge, United Kingdom. hal-04616321

HAL Id: hal-04616321

<https://laas.hal.science/hal-04616321>

Submitted on 19 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Convergence Analysis of Ensemble Filters for Linear Stochastic Systems with Poisson-Sampled Observations*

Aneel Tanwani^a, Olga Yufereva^{a,b}

^a*LAAS – CNRS, University of Toulouse, CNRS, Toulouse France.*
Email: aneel.tanwani@cnrs.fr

^b*Krasovskii Institute of Mathematics and Mechanics, Ural Branch of RAS, Yekaterinburg, Russia.* Email: oiufereva@laas.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. We propose a continuous-discrete McKean–Vlasov type diffusion process with additive Gaussian noise in observation model, which is used to describe the evolution of the individual particles in the ensemble. These particles are coupled through the empirical covariance and require less computations for implementation than the optimal ones based on solving Riccati differential equations. Using appropriate analysis tools, we show that the empirical mean and the sample covariance of the ensemble filter converges to the mean and covariance of the optimal filter if the mean sampling rate of the observation process satisfies certain bounds 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

A fundamental problem of interest in modern engineering systems is to develop techniques for extracting information about the unknown elements of the

*The work of A. Tanwani is partly supported by the project CyPHAI, financed by ANR-JST CREST program with grant number ANR-20-JSTM-0001. The work of O. Yufereva was performed as a part of research conducted in the Ural Mathematical Center with the financial support of the Ministry of Science and Higher Education of the Russian Federation (Agreement number 075-02-2024-1377).

system from the data measurements. Looking at such problems from the viewpoint of dynamical systems, we can naturally study them using filtering algorithms. The early developments in the area of filtering can be traced back to (Kalman and Bucy, 1961), where the authors provided an elegant recursive closed form solution for optimal state estimation in linear stochastic systems. Since this pioneering work, the field has evolved enormously to develop algorithms for nonlinear dynamical systems, address different information structures, and provide computationally efficient methods for large-scale systems. An excellent compilation of different developments on the topic of filtering appears in (Crisan and Rozovskii, 2011).

The basic problem in optimal filtering is to compute the distribution of the unknown state process conditioned upon the observations; see (Anderson and Moore, 1979) for an overview. Among the existing techniques for filtering, the use of Monte Carlo integration methods for approximating the optimal distribution have gained significant interest in the literature (Doucet et al., 2001; Ristic et al., 2003). In the same spirit, Evensen (1994) introduced the technique of ensemble Kalman filters, or feedback particle filters, to develop filtering methods for large scale applications related to geophysical sciences, so that the error covariance is computed from a *collection* of state estimators rather than from a single (Riccati) differential equation. Since then, the use of ensemble Kalman filters has had a notable impact where estimation with noisy data is required in large-scale models. This approach is based on 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 (Chen, 2003; Calvello et al., 2022) and the books (Evensen, 2009) 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.

Over the past decade, the 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 (Bishop and Del Moral, 2023; Taghvaei and Mehta, 2023). 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 (Del Moral and Tugaut, 2018), or as a non-diffusion equation that is optimal in the measure transportation sense (Taghvaei and Mehta, 2020).

Another important research direction is to study the filtering 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 ob-

servations arrive at some random time instants (Sinopoli et al., 2004). It is of interest to compute the conditional distribution of the state process conditioned upon this discrete observation process (Jazwinski, 2007), and results in continuous-discrete filters. 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, (Tanwani and Yufereva, 2020) proposes a continuous-discrete filter and analyze the boundedness of error covariance as a function of the mean sampling rate. A similar performance is studied for a collection of interacting filters over a network in (Tanwani, 2022).

Our primary objective in this article is to develop 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 (Bergemann and Reich, 2010) 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 (Yang et al., 2013) 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. In our recent work (Yufereva and Tanwani, 2023), we developed the transport-inspired particle filters with randomly sampled observations.

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 (Yufereva and Tanwani, 2023), our objective here is to study a different class of ensemble filters which is basically a generalization of the so-called *vanilla filters*: It involves noise terms in the description of the mean-field process. 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. However, the presence of noise terms makes the analysis more challenging, which are tackled using appropriate tools. In essence, we show that for appropriate sampling rate, if the number of the particles in the ensemble is large enough then the empirical mean and the empirical covariance converge to the optimal mean and the covariance.

2. 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

$$dx_t = Ax_t dt + G d\omega_t \quad (1)$$

where $(x_t)_{t \geq 0}$ is an \mathbb{R}^n -valued diffusion process describing the state. Let $(\Omega, \mathcal{F}, \mathbb{P})$ 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 adapted to the filtration $\mathcal{F}_t \subset \mathcal{F}$, with the property that $E[d\omega_t d\omega_t^\top] = I_m dt$, for each $t \geq 0$. The matrices $A \in \mathbb{R}^{n \times n}$ and $G \in \mathbb{R}^{n \times m}$ are taken as constant with (A, G) controllable, and the process $(\omega_t)_{t \geq 0}$ does not depend on the state. The solutions of the stochastic differential equation (1) are interpreted in the sense of Itô stochastic integral.

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 monotone nondecreasing sequence $(\tau_n)_{n \in \mathbb{N}}$ taking values in $\mathbb{R}_{\geq 0}$ which denote the time instants at which the measurements are available for estimation. We introduce the process N_t defined as

$$N_t := \sup\{n \in \mathbb{N} \mid \tau_n \leq t\} \quad \text{for } t \in \mathbb{R}, \quad (2)$$

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 (Suhov and Kelbert, 2008, Theorem 2.3.2) that the Poisson process of intensity $\lambda > 0$ is a continuous-time random process $(N_t)_{t \geq 0}$ taking values in $\mathbb{N}^* := \mathbb{N} \cup \{0\}$, with $N_0 = 0$, for every $n \in \mathbb{N}^*$ and $0 =: t_0 < t_1 < \dots < t_n < +\infty$, 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., (Suhov and Kelbert, 2008, §2.3)) show that it is memoryless and Markovian.

The discretized, and noisy, observation process is thus defined as

$$y_{\tau_{N_t}} = Cx(\tau_{N_t}) + \nu_{\tau_{N_t}}, \quad t \geq 0, \quad (3)$$

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 $dz = Cxdt + d\eta$ for a Wiener process $(\eta_t)_{t \geq 0}$ is formally equivalent to $y_t = Cx_t + \nu_t$, with the identifications $y_t \sim \frac{dz_t}{dt}$ and $\nu_t \sim \frac{d\eta_t}{dt}$, so that ν_t is a Gaussian process; see (Jazwinski, 2007, Chapter 4) for further details. Our goal is to construct the estimate \hat{x}_t , which minimizes the mean square estimation error, using the observations $\mathcal{Y}_t := \{y_{\tau_k} \mid k \leq N_t\}$.

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 (Jazwinski, 2007, 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:

$$\dot{\hat{X}}_t = A\hat{X}_t dt, \quad t \in [\tau_{N_t}, \tau_{1+N_t}[\quad (4a)$$

$$\hat{X}_t^+ = \hat{X}_t + K_t(y_t - C\hat{X}_t), \quad t = \tau_{N_t} \quad (4b)$$

where the injection gain $K_t = P_t C^\top (CP_t C^\top + V_t)^{-1}$, and the error covariance process $(P_t)_{t \geq 0}$ is described as

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

$$\hat{P}_t^+ = P_t - P_t C^\top (CP_t C^\top + V_t)^{-1} CP_t, \quad t = \tau_{N_t}. \quad (5b)$$

To write things more compactly 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 :

$$d\hat{X}_t = A\hat{X}_t dt + K_t(y_t - C\hat{X}_t)dN_t. \quad (6)$$

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

$$dP_t = (AP_t + P_tA^\top + GG^\top)dt - P_t C^\top (CP_t C^\top + V_t)^{-1} CP_t dN_t. \quad (7)$$

In the foregoing discussion, one makes the observation that the optimal conditional distribution 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 (Tanwani and Yufereva, 2020), we showed that the expectation of piecewise deterministic process P_t , denoted by \mathcal{P}_t , is described by the following differential equation:

$$\dot{\mathcal{P}}_t = AP_t + P_tA^\top + GG^\top - \lambda P_t C^\top (CP_t C^\top + V)^{-1} CP_t. \quad (8)$$

We can now 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, G) and (A, C) that guarantee boundedness of \mathcal{P}_t . The boundedness is also important for asymptotic analysis of the first moment of the error process $(x_t - \hat{X}_t)_{t \geq 0}$.

3. Diffusion Process and Ensemble Filters

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 $(\hat{X}_t)_{t \geq 0}$. 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 equation which are coupled to each other through joint statistics of the population.

3.1. Description of McKean-Vlasov type process

For the definition of particles in the ensemble, we first describe a diffusion process of McKean-Vlasov type differential equation. This process is, in particular, continuous-discrete because our observation process is discrete.

$$dS_t := AS_t dt + Gd\bar{\omega}_t + Q_t C^\top (CQC^\top + V_t)^{-1} [y_t - (CS_t + \bar{\nu}_t)] dN_t \quad (9a)$$

$$Q_t := \mathbb{E}[(S_t - \hat{S}_t)(S_t - \hat{S}_t)^\top | \mathcal{Y}_t] \quad (9b)$$

$$\hat{S}_t := \mathbb{E}[S_t | \mathcal{Y}_t], \quad (9c)$$

where the processes $\bar{\omega}_t$ and $\bar{\nu}_t$ are independent copies of the processes ω_t and ν_t respectively, that were introduced in (1) and (3). The motivation of working with this particular diffusion process comes from the fact that the conditional mean \hat{S}_t and the conditional covariance Q_t satisfy the same differential equations as the ones obtained for their optimal counterparts, that is, it can be checked that

$$d\hat{S}_t = A\hat{S}_t dt + Q_t C^\top (CQ_t C^\top + V_t)^{-1} (y_t - C\hat{S}_t) dN_t \quad (10)$$

and

$$dQ_t = (AQ_t + Q_tA^\top + GG^\top)dt - Q_t C^\top (CQ_t C^\top + V_t)^{-1} CQ_t dN_t, \quad (11)$$

so that, one sees the resemblance with the processes described in (6) and (7), respectively.

3.2. Ensemble Filters

From the mean-field process (9), we describe an ensemble of M particles. For $\ell = 1, \dots, M$, the particle S_t^ℓ is described by the following stochastic differential

equation conditioned upon the observation process and the Poisson sampling process:

$$dS_t^\ell = AS_t^\ell dt + Gd\omega_t^\ell + K_t^M(y_t - CS_t^\ell - \nu_t^\ell) dN_t, \quad (12)$$

where, for each ℓ , ω_t^ℓ and ν_t^ℓ represent the independent copies of the processes ω_t and ν_t respectively. The term K_t^M denotes the empirical gain computed at time instants $t = \tau_{N_t}$ as follows:

$$K_t^M = Q_t^M C^\top (CQ_t^M C^\top + V)^{-1} \quad (13)$$

in which Q_t^M denotes the empirical covariance, which is computed using the empirical mean \widehat{S}_t^M as follows:

$$Q_t^M = \frac{1}{M} \sum_{\ell=1}^M (S_t^\ell - \widehat{S}_t^M)(S_t^\ell - \widehat{S}_t^M)^\top \quad (14)$$

$$\widehat{S}_t^M = \frac{1}{M} \sum_{\ell=1}^M S_t^\ell. \quad (15)$$

One clearly sees that the computational effort in implementing the ensemble filters (12) involves solving Mn differential equations. For computing the gain, the empirical mean and the sample covariance only needs to be updated at discrete times using relatively simpler arithmetic operations.

3.3. Convergence result

The obvious question concerning the proposed filters in (12) is whether they approximate the optimal filter (6)–(7) reasonably well. In this regard, our main result shows that if the mean sampling rate satisfies certain bounds, then the difference (in expectation) between the empirical covariance of the particles (14) and the optimal covariance (15) is $O(\frac{1}{M})$. Similarly, we can show that, as the number of particles increase, the expectation of the empirical mean of the particles approaches the expected value of the optimal estimate \widehat{X} .

Our main result formalizing these observations is stated next. To avoid technicalities in the proof, we present the convergence result for the scalar case only, where A is a real number and G, C are nonzero scalars. This allows us to give explicit bounds on the approximation error, and the mean sampling rate required for establishing these bounds.

Theorem 1. Consider system (1) with optimal estimator (6) and the ensemble of M filters described by (12). Let $\bar{\gamma} := \max \left\{ \frac{V^2}{(\min\{P_0, Q_0^M\}C^2 + V)^2}, \frac{V^2(\lambda - 2A)^2}{((1 - 1/M)G^2C^2 + V(\lambda - 2A))^2} \right\}$. If the mean sampling rate satisfies the inequality

$$\lambda > 2A/(1 - \bar{\gamma}^2) \quad (16)$$

then it holds that,

$$\limsup_{t \rightarrow \infty} \|\mathbb{E}[P_t] - \mathbb{E}[Q_t^M]\| \leq \frac{G^2 + \lambda V/C^2}{M(\lambda(1 - \bar{\gamma}^2) - 2A)} \quad (17)$$

and moreover,

$$\|\mathbb{E}[\hat{S}_t^M] - \mathbb{E}[\hat{X}_t]\| \rightarrow 0, \quad \text{as } M \rightarrow \infty, \text{ and } t \rightarrow \infty.$$

The proof of this result is carried out in Section 4. The constant $\bar{\gamma}$ appearing in the theorem statement actually corresponds to the lower bound on the covariance matrices P_t and Q_t^M . Nonzero value of G is important for establishing this lower bound. Because of nonlinear dependence of $\bar{\gamma}$ on λ , inequality (16) provides the range of values for the mean sampling rate for which the result in Theorem 1 holds. For the measurement covariance V small enough, this inequality is indeed feasible. Also, if $A \leq 0$, (24) trivially holds with any $\lambda > 0$. The ultimate bound on approximation of covariance matrix in (17) increases with the increase in the covariance of process and sensor noises, and decreases as number of particles M gets large.

3.4. Simulation of an academic example

To illustrate the simulation of our proposed ensemble filter in (12) and the result reported in Theorem 1, we consider an academic example where the system is described by the equations:

$$dx_t = Ax_t dt + Gd\omega_t \tag{18a}$$

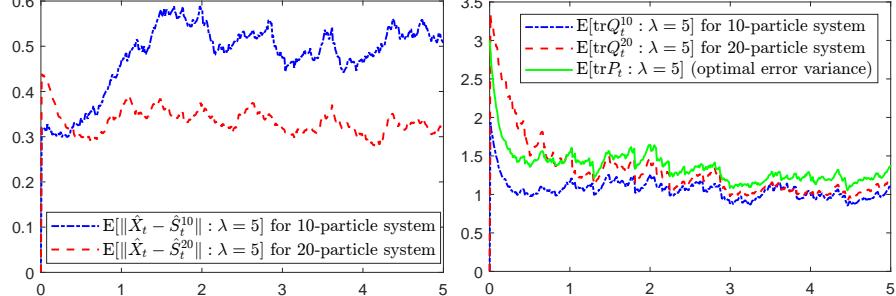
$$y_{\tau_k} = Cx_{\tau_k} + \nu_{\tau_k}, \tag{18b}$$

with $A = \begin{bmatrix} 0 & 3 & 1 \\ 2 & -2 & 1 \\ -2 & 1 & -3 \end{bmatrix}$, $C = \begin{bmatrix} 1 & -1 & 2 \\ 1 & 0 & 1 \end{bmatrix}$, $G = [0.5 \ 0.5 \ 0.5]^\top$, and for each $k \in \mathbb{N}^*$, ν_{τ_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, and look at the entities that are being compared in Theorem 1 for different values of mean sampling rate. The simulation results reported in Figure 1 and Figure 2 are consistent with the results in Theorem 1.

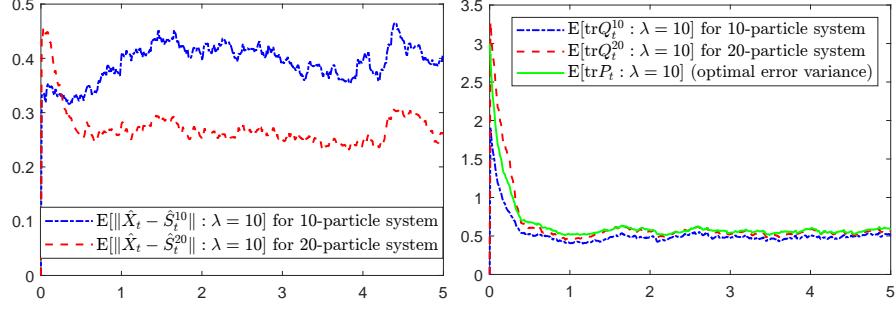
4. Convergence of Ensemble Filters

This section essentially carries out the proof of Theorem 1. We first derive the differential equations for empirical mean and covariance using Ito's chain rule, using Proposition 3 in Appendix A). We then compute the expectations of the processes using the generator equation stated in Proposition 5, which are then used for studying convergence results stated in Theorem 1. We emphasize that the developments in this section are carried out for the scalar case only.



(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 ensemble filters 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 ensemble filters with mean sampling rate $\lambda = 10$.

4.1. Evolution of empirical mean and covariance

We recall that the evolution of the particles is described by the differential equation (12) and the corresponding empirical mean is defined in (15). We now write down the evolution of the empirical mean of the particles in differential form as follows:

$$d\hat{S}_t^M = A\hat{S}_t^M + \frac{1}{\sqrt{M}}Gd\tilde{\omega}_t^M + K_t^M \left(y_t - C\hat{S}_t^M - \frac{1}{\sqrt{M}}\tilde{\nu}_t^M \right) dN_t$$

where we used the notation

$$\tilde{\omega}_t^M := \frac{1}{\sqrt{M}} \sum_{\ell=1}^M \omega_t^\ell \quad \text{so that} \quad E[d\tilde{\omega}_t^M d\tilde{\omega}_t^{M\top}] = dt$$

and

$$\tilde{\nu}_t^M := \frac{1}{\sqrt{M}} \sum_{\ell=1}^M \nu_t^\ell \quad \text{so that} \quad \tilde{\nu}_t^M \sim \nu.$$

Note that the process $\tilde{\nu}_t^M$ is also i.i.d. Gaussian with covariance equal to V .

Next we write the empirical covariance of the particles, defined in (14), in differential form. To do so, we let

$$q_t^\ell := S_t^\ell - \hat{S}_t^M.$$

It then follows that

$$dq_t^\ell = Aq_t^\ell dt + G \left(d\omega_t^\ell - \frac{1}{\sqrt{M}} d\tilde{\omega}_t^M \right) - K_t^M C q_t^\ell dN_t - K_t^M \left(\nu_t^\ell - \frac{1}{\sqrt{M}} \tilde{\nu}_t^M \right) dN_t. \quad (19)$$

Consequently, we can use Proposition 3 to get

$$\begin{aligned} d(q_t^\ell)^2 &= \left(2A(q_t^\ell)^2 + G^2 - \frac{1}{M} G^2 \right) dt + 2Gq_t^\ell \left(d\omega_t^\ell - \frac{1}{\sqrt{M}} d\tilde{\omega}_t^M \right) \\ &\quad + \left[(q_t^\ell - K_t^M C q_t^\ell - K_t^M \nu_t^\ell + \frac{1}{\sqrt{M}} K_t^M \tilde{\nu}_t^M)^2 - (q_t^\ell)^2 \right] dN_t. \end{aligned}$$

Straightforward simplification yields

$$\begin{aligned} d(q_t^\ell)^2 &= \left(2A(q_t^\ell)^2 + G^2 - \frac{1}{M} G^2 \right) dt + 2Gq_t^\ell \left(d\omega_t^\ell - \frac{1}{\sqrt{M}} d\tilde{\omega}_t^M \right) \\ &\quad + \left[(K_t^M)^2 C^2 (q_t^\ell)^2 - 2K_t^M C (q_t^\ell)^2 + (K_t^M)^2 (\nu_t^\ell)^2 \right. \\ &\quad \left. + \frac{1}{M} (K_t^M)^2 (\tilde{\nu}_t^M)^2 - \frac{2}{\sqrt{M}} (K_t^M)^2 \nu_t^\ell \tilde{\nu}_t^M \right. \\ &\quad \left. - 2(1 - K_t^M C) q_t^\ell K_t^M (\nu_t^\ell - \frac{1}{\sqrt{M}} K_t^M \tilde{\nu}_t^M) \right] dN_t. \end{aligned}$$

Note that, we can write $Q_t^M = \sum_{\ell=1}^M (q_t^\ell)^2$. Furthermore, recalling the expression for empirical gain in (13) we can check that

$$(K_t^M)^2 C^2 Q_t^M - 2K_t^M C Q_t^M + (K_t^M)^2 V = -K_t^M C Q_t^M.$$

The last two identities allow us to write

$$\begin{aligned} dQ_t^M &= \left(2AQ_t^M + G^2 - \frac{1}{M} G^2 \right) dt + 2G \sum_{\ell=1}^M q_t^\ell \left(d\omega_t^\ell - \frac{1}{\sqrt{M}} d\tilde{\omega}_t^M \right) \\ &\quad + \left[-K_t^M C Q_t^M - (K_t^M)^2 V + \frac{1}{M} (K_t^M)^2 \sum_{\ell=1}^M (\nu_t^\ell)^2 \right. \\ &\quad \left. + \frac{1}{M} (K_t^M)^2 (\tilde{\nu}_t^M)^2 - \frac{2}{M \sqrt{M}} (K_t^M)^2 \sum_{\ell=1}^M \nu_t^\ell \tilde{\nu}_t^M \right. \\ &\quad \left. - \frac{2}{M} (1 - K_t^M C) K_t^M \sum_{\ell=1}^M q_t^\ell (\nu_t^\ell - \frac{1}{\sqrt{M}} K_t^M \tilde{\nu}_t^M) \right] dN_t. \end{aligned} \quad (20)$$

We next use this expression to compute the expectation of $\mathbb{E}[Q_t^M]$ with respect to the noise terms and, in addition, the sampling times of the Poisson counter.

4.2. Convergence of expected covariance

Using Proposition 5, we now turn our attention to computing the expectation of optimal covariance P_t (with respect to sampling times), which was described by a piecewise deterministic Markov process in (7). Similarly, we will use Proposition 5 to describe the expectation of the empirical covariance (with respect to noise in the ensemble filter and the underlying sampling process), which is described by a jump diffusion process in (20). In what follows, we use the notation $\mathcal{P}_t := \mathbb{E}[P_t]$, $\mathcal{Q}_t^M = \mathbb{E}[Q_t^M]$, and similarly, for the corresponding injection gains, we use $\mathcal{K}_t := \mathcal{P}_t C (\mathcal{P}_t C^2 + V)^{-1}$ and $\mathcal{K}_t^M := \mathcal{Q}_t^M C (\mathcal{Q}_t^M C^2 + V)^{-1}$. By applying Proposition 5 to the process (7), we first get

$$\begin{aligned}\dot{\mathcal{P}}_t &= 2A\mathcal{P}_t + G^2 - \lambda\mathcal{P}_t^2 C^2 (\mathcal{P}_t C^2 + V)^{-1} \\ &= (2A - \lambda)\mathcal{P}_t + G^2 + \lambda\mathcal{P}_t(1 - \mathcal{P}_t C^2 (\mathcal{P}_t C^2 + V)^{-1}) \\ &= (2A - \lambda)\mathcal{P}_t + G^2 + \lambda\mathcal{P}_t\phi(\mathcal{P}_t)\end{aligned}\quad (21)$$

where the mapping $\mathcal{P}_t \mapsto \phi(\mathcal{P}_t)$ is defined as

$$\phi(\mathcal{P}_t) := (1 - \mathcal{P}_t C^2 (\mathcal{P}_t C^2 + V)^{-1}).$$

On the other hand, using the fact that ν_t^ℓ and $\tilde{\nu}_t^M$ are mean zero processes, the application of Proposition 5 to the process Q_t^M in (20) yields,

$$\begin{aligned}\dot{\mathcal{Q}}_t^M &= 2A\mathcal{Q}_t^M + G^2 - \frac{1}{M}G^2 - \lambda\mathcal{Q}_t^M C^2 (\mathcal{Q}_t^M C^2 + V)^{-1} \mathcal{Q}_t^M + \frac{\lambda}{M}(\mathcal{K}_t^M)^2 V \\ &= (2A - \lambda)\mathcal{Q}_t^M + G^2 + \lambda\mathcal{Q}_t^M\phi(\mathcal{Q}_t^M) - \frac{1}{M}G^2 + \frac{\lambda}{M}(\mathcal{K}_t^M)^2 V.\end{aligned}\quad (22)$$

To analyze the difference between optimal and empirical covariance, let us introduce $\mathcal{E}_t^M := \mathcal{Q}_t^M - \mathcal{P}_t$, which satisfies the following equation:

$$\dot{\mathcal{E}}_t^M = (2A - \lambda)\mathcal{E}_t^M + \lambda\mathcal{Q}_t^M\phi(\mathcal{Q}_t^M) - \lambda\mathcal{P}_t\phi(\mathcal{P}_t) - \frac{1}{M}G^2 + \frac{\lambda}{M}(\mathcal{K}_t^M)^2 V. \quad (23)$$

The following lemma states some important properties that we need in the sequel. Its proof can be found in Appendix C.

Lemma 2. *For the equation (21) and equation (22), we have the following relations:*

1. *There exists $0 < \gamma \leq \bar{\gamma} < 1$ such that for each $t \geq 0$, we have*

$$\begin{aligned}1 - \gamma &\leq \|\mathcal{P}_t C^2 (\mathcal{P}_t C^2 + V)^{-1}\| < 1, \\ 1 - \gamma &\leq \|\mathcal{Q}_t^M C^2 (\mathcal{Q}_t^M C^2 + V)^{-1}\| < 1.\end{aligned}$$

Consequently, it holds that $\|\phi(\mathcal{P}_t)\| \leq \gamma$ and $\|\phi(\mathcal{Q}_t^M)\| \leq \gamma$.

2. The function ϕ has the property that for each P_1 and P_2 , it satisfies

$$\phi(P_1)P_1 - \phi(P_2)P_2 \leq \phi(P_1)(P_1 - P_2)\phi(P_2).$$

One can now use the statements in this lemma to get a bound on the evolution of \mathcal{E}_t^M . This is done by looking at the derivative of $(\mathcal{E}_t^M)^2$, that is,

$$\frac{d}{dt}(\mathcal{E}_t^M)^2 \leq 2(2A - \lambda)(\mathcal{E}_t^M)^2 + 2\lambda\gamma^2(\mathcal{E}_t^M)^2 + \frac{2}{M}\|\mathcal{E}_t^M\|G^2 + \frac{2\lambda}{MC^2}V\|\mathcal{E}_t^M\|.$$

In the last inequality, the quadratic term in $(\mathcal{E}_t^M)^2$ is negative if $\lambda - 2A > \lambda\gamma^2$, or equivalently,

$$\lambda > \frac{2A}{1 - \gamma^2}. \quad (24)$$

The ultimate bound on \mathcal{E}_t^M now follows for every finite value of λ satisfying (24). In particular,

$$\limsup_{t \rightarrow \infty} \|\mathcal{E}_t^M\| \leq \frac{G^2 + \lambda V/C^2}{M(\lambda(1 - \gamma^2) - 2A)}$$

where we note that the denominator is strictly positive due to the bound on λ in (24).

4.3. Convergence of expected mean

The last step in the proof of Theorem 1 is to analyze the difference between the expected values of the optimal mean and empirical mean. In this regard, we let $\widehat{\mathcal{S}}_t^M := \mathbb{E}[\widehat{S}_t^M]$, and $\widehat{\mathcal{X}}_t = \mathbb{E}[\widehat{X}_t]$. Applying Proposition 5, we first observe that

$$\dot{\widehat{\mathcal{X}}}_t = A\widehat{\mathcal{X}}_t + \lambda\mathcal{K}_t(\mathbb{E}[y_t] - C\widehat{\mathcal{X}}_t)$$

and

$$\dot{\widehat{\mathcal{S}}}_t^M = A\widehat{\mathcal{S}}_t^M + \lambda\mathcal{K}_t^M(\mathbb{E}[y_t] - C\widehat{\mathcal{S}}_t^M)$$

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

$$\begin{aligned} \dot{\widehat{\chi}}_t^M &= A\widehat{\chi}_t^M + \lambda\mathcal{K}_t(y_t - C\widehat{\mathcal{X}}_t) - \lambda\mathcal{K}_t^M(\mathbb{E}[y_t] - C\widehat{\mathcal{S}}_t^M) \\ &= (A - \lambda\mathcal{K}_t^MC)\widehat{\chi}_t^M + \lambda(\mathcal{K}_t - \mathcal{K}_t^M)(\mathbb{E}[y_t] - C\widehat{\mathcal{X}}_t). \end{aligned}$$

The bounds on asymptotic value of $\widehat{\chi}_t^M$ can now be established by analyzing the derivative of the following weighted quadratic function:

$$\widehat{\chi}_t^M \mapsto (\mathcal{Q}_t^M)^{-1}(\widehat{\chi}_t^M)^2.$$

We observe that

$$\begin{aligned} \frac{d}{dt}(\mathcal{Q}_t^M)^{-1} &= -(\mathcal{Q}_t^M)^{-1}\dot{\mathcal{Q}}_t^M(\mathcal{Q}_t^M)^{-1} \\ &= -2(A - \lambda\mathcal{K}_t^MC)(\mathcal{Q}_t^M)^{-1} - \lambda C^2(\mathcal{Q}_t^M C^2 + V)^{-1} \\ &\quad - (\mathcal{Q}_t^M)^{-2} \left(G^2 + \frac{1}{M}G^2 + \frac{\lambda}{M}\mathcal{K}_t^M V \mathcal{K}_t^M \right). \end{aligned}$$

Consequently, this last expression yields

$$\begin{aligned} \frac{d}{dt} \hat{\chi}_t^M (\mathcal{Q}_t^M)^{-1} \hat{\chi}_t^M &= -(\hat{\chi}_t^M)^2 \left[\lambda C^2 (\mathcal{Q}_t^M C^2 + V)^{-1} \right. \\ &\quad \left. + (\mathcal{Q}_t^M)^{-2} \left(G^2 + \frac{1}{M} G^2 + \frac{\lambda}{M} \mathcal{K}_t^M V \mathcal{K}_t^M \right) \right] \\ &\quad + 2\lambda (\mathcal{Q}_t^M)^{-1} \hat{\chi}_t^M (\mathcal{K}_t - \mathcal{K}_t^M) (\mathbb{E}[y_t] - C \hat{\mathcal{X}}_t). \end{aligned}$$

The right-hand side can be seen as a sum of two terms: the first one is negative definite and is quadratic in $\hat{\chi}_t^M$; the second term is linear in $\hat{\chi}_t^M$ but contains $(\mathcal{K}_t - \mathcal{K}_t^M)$ which is converging to zero as M gets large. Using simple algebraic manipulations, and noting that \mathcal{Q}_t^M is positive definite, it can be shown that $\hat{\chi}_t^M$ converges to zero, as $t \rightarrow \infty$ and as M gets large. This completes the proof of Theorem 1.

5. Conclusions

In this paper, we considered the analysis of vanilla ensemble filters for continuous-time linear Gaussian processes with Poisson-sampled discrete observations. The description of the particles is based on a continuous-discrete mean field process that contains flow equations with Wiener processes and jump equations with additive Gaussian noise in the modeled observation process. Using an extended generator, we compute the expectation of the empirical mean and empirical covariance of the particles with respect to the sampling process. The main result shows that for a sufficiently large value of the mean sampling rate, the empirical mean and covariance converge to the optimal mean and variance, respectively, as the number of particles tend to infinity.

The immediate extension of our work would be to generalize the proof of Theorem 1 to systems of higher dimension. A better understanding of the bounds on the mean sampling rate needs is required. It is also interesting to investigate whether the tools developed in this paper can be used to study broader class of particle filters with stochastic noise.

References

- Anderson, B., Moore, J., 1979. Optimal Filtering. Dover Publications.
- Bergemann, K., Reich, S., 2010. A mollified ensemble Kalman filter. Q.J. Royal Meteorological Society 136, 1636–1643.
- Bishop, A.N., Del Moral, P., 2023. On the mathematical theory of ensemble (linear-Gaussian) Kalman–Bucy filtering. Math. Control, Signals and Systems
- .

- Calvello, E., Reich, S., Stuart, A.M., 2022. Ensemble Kalman methods: A mean field perspective. Available online; arXiv:2209.11371v1.
- Chen, Z., 2003. Bayesian filtering: From Kalman filters to particle filters, and beyond. *Statistics* 182, 1–69.
- Crisan, D., Rozovskii, B., 2011. *The Oxford Handbook of Nonlinear Filtering*. Oxford University Press.
- Del Moral, P., Tugaut, J., 2018. On the stability and the uniform propagation of chaos properties of ensemble Kalman–Bucy filters. *The Annals of Applied Probability* 28, 790–850.
- Doucet, A., De Freitas, N., Gordon, N. (Eds.), 2001. *Sequential Monte Carlo methods in practice*. 1 ed., Springer.
- Evensen, G., 1994. 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.
- Evensen, G., 2009. *Data Assimilation: The Ensemble Kalman Filter*. 2nd ed., Springer Science & Business Media.
- Hespanha, J.P., Teel, A.R., 2006. Stochastic impulsive systems driven by renewal processes, in: Proc. 17th International Symposium on Mathematical Theory of Networked Systems, pp. 606–618.
- Jazwinski, A., 2007. *Stochastic processes and filtering theory*. Dover Publications, Inc., New York.
- Kalman, R., Bucy, R., 1961. New results in linear filtering and prediction theory. *Journal of Basic Engineering* 83, 95–108.
- Protter, P.E., 2005. *Stochastic Integration and Differential Equations*. 2nd ed., Springer.
- Ristic, B., Arulampalam, S., Gordon, N., 2003. *Beyond the Kalman filter: Particle filters for tracking applications*. Artech house.
- Sinopoli, B., Schenato, L., Franceschetti, M., Poola, K., Jordan, M.I., Sastry, S.S., 2004. Kalman filtering with intermittent observations. *IEEE Transactions on Automatic Control* 49, 1453–1464.
- Suhov, Y., Kelbert, M., 2008. *Probability and Statistics by Example. II. Cambridge University Press, Cambridge*. Markov chains: a primer in random processes and their applications.
- Taghvaei, A., Mehta, P., 2020. An optimal transport formulation of the ensemble Kalman filter. *IEEE Transactions on Automatic Control* 66, 3052–3067.

- Taghvaei, A., Mehta, P., 2023. A survey of feedback particle filter and related controlled interacting particle systems (CIPS). *Annual Reviews in Control* 55, 356–378.
- Tanwani, A., 2022. Suboptimal filtering over sensor networks with random communication. *IEEE Transactions on Automatic Control* 67, 5456–5463.
- Tanwani, A., Yuferova, O., 2020. Error covariance bounds for suboptimal filters with Lipschitzian drift and Poisson-sampled measurements. *Automatica* 122.
- Yang, T., Mehta, P.G., Meyn, S.P., 2013. Feedback particle filter. *IEEE Transactions on Automatic Control* 58, 2465–2480.
- Yuferova, O., Tanwani, A., 2023. Transport inspired particle filters with Poisson-sampled observations in Gaussian setting, in: Proceedings of 62nd IEEE Conference on Decision and Control (CDC), pp. 7695–7700.

A. Tools for Analysis

In this section, we discuss some analysis tools for stochastic differential equations that will be used in Section 4 for the proof of Theorem 1. To do so, we consider the following differential equation for a stochastic process $(x_t)_{t \geq 0}$ evolving in \mathbb{R}^n :

$$dx_t = f(x_t)dt + g(x_t)d\omega + h(x_t, \nu_{N_t})dN_t \quad (\text{A.1})$$

where N_t is a Poisson process with intensity $\lambda > 0$, ω is a standard Wiener process, and ν_{N_t} is a sequence of i.i.d. random variables with probability law μ .

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 (Protter, 2005, 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 3 (Ito's chain rule). *For a twice continuously differentiable function $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$, it holds that*

$$\begin{aligned} d\psi(x_t) = & \left[\langle \nabla \psi(x_t), f(x_t) \rangle + \frac{1}{2} \text{tr} \left(\frac{\partial^2 \psi(x_t)}{\partial x} g(x_t) g(x_t)^\top \right) \right] dt \\ & + \langle \nabla \psi(x_t), g(x_t) \rangle d\omega + \left[\psi(x_t + h(x, \nu_{N_t})) - \psi(x_t) \right] dN_t. \end{aligned} \quad (\text{A.2})$$

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

$$\begin{aligned} \mathbb{R}^n \ni z &\mapsto \mathcal{L}\psi(z) \in \mathbb{R} \\ \mathcal{L}\psi(z) &:= \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} \left(\mathbb{E}[\psi(x(t + \varepsilon)) \mid x(t) = z] - \psi(z) \right). \end{aligned} \quad (\text{A.3})$$

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

$$\mathbb{E}[\psi(x(t))] = \mathbb{E}[\psi(x(0))] + \mathbb{E} \left[\int_0^t \mathcal{L}\psi(x(s)) \, ds \right]. \quad (\text{A.4})$$

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 $\mathbb{E}[\psi(x)]$. Several references in the literature derive generator equations for stochastic processes with jumps, see for example (Hespanha and Teel, 2006) for a derivation in the context of Wiener process driven differential equations with renewal processes.

Proposition 5. *If the sampling process $(N_t)_{t \geq 0}$ is Poisson with intensity $\lambda > 0$, then the process $(x(t))_{t \geq 0}$ 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 $\|z\| \rightarrow +\infty$, we have*

$$\begin{aligned} \mathcal{L}\psi(z) &= \langle \nabla \psi(z), f(z) \rangle + \frac{1}{2} \text{tr} \left(\frac{\partial^2 \psi(z)}{\partial z^2} g(z) g(z)^\top \right) \\ &\quad + \lambda \left[\int \psi(z + h(z, \nu)) \mu(d\nu) - \psi(z) \right]. \end{aligned} \quad (\text{A.5})$$

B. Proof of Lemma 2

We provide a proof for the third item, which is independent of the first two items. To do so, we introduce the notation

$$K_1 := P_1 C (P_1 C^2 + V)^{-1}; \quad K_2 := P_2 C (P_2 C^2 + V)^{-1}.$$

It can be easily checked that

$$\phi(P_1) = 1 - K_1 C; \quad \phi(P_2) = 1 - K_2 C,$$

and moreover, we have the identities

$$\phi(P_1) P_1 C = K_1 V; \quad \phi(P_2) P_2 C = K_2 V.$$

The last two equations result in

$$\begin{aligned}
& \phi(P_1)P_1 - \phi(P_2)P_2 \\
&= \phi(P_1)P_1 + K_1VK_2 - K_1VK_2 - \phi(P_2)P_2 \\
&= \phi(P_1)P_1 + \phi(P_1)P_1CK_2 - K_1\phi(P_2)P_2C - \phi(P_2)P_2 \\
&= \phi(P_1)P_1(1 - K_2C) - (1 - K_1C)\phi(P_2)P_2 \\
&= \phi(P_1)P_1\phi(P_2) - \phi(P_1)\phi(P_2)P_2 \\
&= \phi(P_1)(P_1 - P_2)\phi(P_2)
\end{aligned}$$

which is the desired result.

C. Bounds on Sampling Rate

It can be seen that the constant γ used in the derivations actually depends on the mean sampling rate λ . Here we derive the bounds explicitly on λ . From the definition of γ , it follows that

$$\gamma \leq 1 - \frac{\mathcal{P}_t C^2}{\mathcal{P}_t C^2 + V} = \frac{V}{\mathcal{P}_t C^2 + V}.$$

On the other hand, it can be shown that

$$\mathcal{P}_t \geq \min \left\{ P_0, \frac{G^2}{\lambda - 2A} \right\}.$$

and, for $G_M := G\sqrt{(1 - 1/M)}$

$$\mathcal{Q}_t \geq \min \left\{ Q_0^M, \frac{G_M^2}{\lambda - 2A} \right\}.$$

Letting $\tilde{\lambda} := \lambda - 2A$, we get

$$\gamma \leq \max \left\{ \frac{V}{\min\{P_0, Q_0^M\}C^2 + V}, \frac{V\tilde{\lambda}}{G_M^2 C^2 + V\tilde{\lambda}} \right\}.$$

Consequently, the condition

$$\tilde{\lambda} > \lambda\gamma^2$$

holds if

$$\tilde{\lambda} > \max \left\{ \frac{V^2}{(\min\{P_0, Q_0\}C^2 + V)^2}, \frac{V^2\tilde{\lambda}^2}{(G_M^2 C^2 + V\tilde{\lambda})^2} \right\} (\tilde{\lambda} + 2A).$$

Using simple algebraic manipulations, the last inequality holds if and only if

$$2V(\lambda - 2A)(AV - G_M^2 C^2) < G_M^4 C^4,$$

and

$$\lambda > 2A \left(1 - \frac{V^2}{(\min\{P_0, Q_0\}C^2 + V)^2} \right)^{-1}.$$

The first inequality holds if $AV \leq G_M^2 C^2$, or the measurement noise covariance V is relatively small.