2024 IEEE 63rd Conference on Decision and Control (CDC)
December 16-19, 2024. MiCo, Milan, Italy

On Analyzing Filters with Bayesian Parameter Inference and Poisson-Sampled Observations

Aneel Tanwani

Abstract— **The problem of state estimation in continuous-time linear stochastic systems is considered with several constraints on the available information. It is stipulated that the model of the system contains several unknown parameters and the observation process is randomly time-sampled. The classical solution due to Kalman-Bucy cannot be implemented in that case, and we revisit the idea of partitioning the set of unknown parameters, and consider a multiple filters corresponding to each possible value of the unknown parameter. The posterior distribution of the unknown parameters conditioned upon available observations is computed from Bayes' rule. The resulting state estimate is a weighted sum of the state estimates generated by multiple Kalman filters, where the weights are determined by the posterior distribution of the unknown parameters. We analyze the performance of the algorithm by looking at its asymptotic behavior and establishing boundedness of the error covariance matrix.**

Index Terms— **Filtering; partitioning; Poisson-sampled observations; stability; performance analysis.**

I. Introduction

Estimation theory in stochastic dynamical systems is a problem of great interest because of its relevance in several engineering related disciplines. Over the decades, we have seen development of several important concepts and elegant results in this field, stemming from the pioneering work of Kalman and Bucy [1]. In its simplest form, Kalman filter provides a rather intuitive way of weighing the information contained in the model and the measurements depending upon the covariance of the noise driving the state process and the covariance of the measurement noise. The accurate information of the model and noise statistics is therefore very important for optimal state estimation, and if this information is not available accurately, the filter can even generate diverging state estimates [2]. Moreover, with frequent use of communication channels for transmission of observations in several applications, the measurements are available for estimation at only some random time instants. Thus, the lack of information about the system parameters and random sampling of the observation process provide an interesting framework for the estimation problem which is explored in this paper.

For systems with unknown parameters, the state estimation for stochastic systems has been classically studied under the framework of adaptive state estimation [3]. Several approaches have been developed in this regard and several monographs, such as [4], [5], and the survey article [6] provide an overview of the developments in this area. Quite naturally, one possible way to study this problem is to compute the estimates of the unknown parameters from the noisy measurements and use them to update the filtering equations. One possible approach relies on inference using Bayes' rule to compute the posterior distribution of the unknown parameters [7], [8]. The approach has the drawback that computing this posterior, despite being in recursive form, is computationally heavy. Another approach (seen as a continuation of the Bayes' estimation) relates to maximizing different likelihood functions [4, Chapter 10]. This involves computing a tractable form for the likelihood function and setting the gradient with respect to parameters equal to zero to find the values which maximize the likelihood function. Analytical solutions can be derived in certain restrictive cases for steady state operation. Some of these techniques appear in the book [9], while providing an overview of Bayesian methods in filtering problem. A rather comprehensive collection of statistical algorithms used for parameter inference can be found in [10]. The recent work [11] provides one possible solution which relies on computing gradients backwards to speed up certain computations.

Most of the aforementioned references deal with systems in discrete-time and provide *only* the algorithms for optimal state estimation with unknown parameters, based on tools from information theory and control systems. There are relatively few works that establish theoretical guarantees of such algorithms, see for example [12, Chapter 10] and [13], which in turn builds on computing the variance of Bayes' estimates [14] and proving the consistency of maximum likelihood methods [15]. The book [5] collects some of these earlier developments in a succinct form with different assumptions on the set of unknown parameters. The problem of optimal state estimation under limited information continues to attract attention but lately, for systems with unknown dynamics, this problem has been studied using different tools. The paper [16] uses the form of optimal Kalman filter to compute the injection gain directly from the data using neural networks, and several other works also try to compute the optimal gain using different techniques [17]. However, more often than not, these works work under very restrictive cases that the full state measurements are available as ground truth which essentially violates the spirit of state estimation problems from system-theoretic viewpoint. Some preliminary work based on the use of stochastic gradient descent methods

A. Tanwani is affiliated with CNRS – LAAS, University of Toulouse, CNRS, 31400 Toulouse, France (e-mail: aneel.tanwani@cnrs.fr).

This work is sponsored by the project CyPhAI, financed by ANR-JST CREST program with grant number ANR-20-JSTM-0001.

to compute linear gains minimizing the square norm of the residual has been presented in [18].

Stochastic control under information constraints is another relevant direction of research which has grown considerably in the past two decades due to common use of communication channels between different components of the system. In such cases, it becomes important to adapt the policies to achieve optimality under given information structure and analyze the performance of such policies [19], [20], [21]. For our purposes, we are particularly interested in situations where the information structure is abstractly modeled by random sampling of the observation process. In this regard, we have seen some results developed for the continuous-time stochastic optimal control problems subject to randomly sampled state measurements [22]. Motivated by the idea of implementing filters subject to observations transmitted over networks through some communication protocols, it is natural to stipulate that the observations arrive at some random time instants [23]. For continuous-time systems, and observation process being discrete, it is of interest to compute the distribution of the state process conditioned upon the discrete observations [24]. In our previous work, we have studied the performance of filters for continuous-time systems using Poisson-sampled observations in [25], and analyze the boundedness of error covariance as a function of the mean sampling rate.

The problem of state estimation for stochastic systems with unknown parameters, despite being classic, continues to remain relevant as we seek algorithms which require less information with theoretical guarantees about their long-term behavior. This paper takes a step in this direction by studying the filter design for linear stochastic systems with unknown parameters, and the observations process being time sampled by a Poisson counter. We work under the assumption that all the unknown parameters belong to a discrete set. We construct an optimal filter for each value of the parameter which basically resembles Kalman filter but is driven by randomly-sampled observations. The optimal state estimate minimizing the variance of the estimation error is then described by a weighted sum of the estimates generated by individual filters, where the weights are determined by the posterior probability density function of the unknown parameters conditioned upon the observation process. For the discrete parameter set, we propose a recursive algorithm which allows for computing these conditional posterior distributions using the Bayes' rule. While such methods are computationally heavy, there are some consistency results in the literature that advocate their broad applicability. Once the algorithm is designed, we analyze the performance by looking at the asymptotic value of the expectation of the total error variance. We show that, if the underlying Bayes' inference rule is consistent, then the error covariance converges to the optimal value that arises in the case with completely known parameters. For sufficiently large mean sampling rate, one can also prove that the asymptotic error covariance matrix remains bounded.

II. Model with Unknown Parameters

We consider the problem of designing a filter for a class of linear stochastic systems under some constraints on what information is available for designing the filters, and the measurement process. In this section, we formally describe the system class and the associated measurement process.

In what follows, we will denote the unknowns by a vector θ, which belongs to a set Θ in some Euclidean space. These unknowns may appear in the system data or the noise statistics that are used later on for designing optimal filters. We consider the dynamical systems modeled by linear stochastic differential equations of the form:

$$\mathrm{d}X_t = A(\theta)X_t\,\mathrm{d}t + G(\theta)\,\mathrm{d}\omega_t \tag{1}$$

where $(X_t)_{t\geq 0}$ is an $\mathbb{R}^n$-valued diffusion process describing the state. The matrices $A(\theta) \in \mathbb{R}^{n\times n}$ and $G(\theta) \in \mathbb{R}^{n\times m}$ are constant but may depend on the unknown parameter $\theta \in \Theta$. 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 adapted to the filtration $\mathcal{F}_t \subset \mathcal{F}$, 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 process $(\omega(t))_{t\geq 0}$ does not depend on the state, and since we allow G to be unknown, there is no loss of generality in assuming that the statistics of random process $\mathrm{d}\omega$ are known. The solutions of the stochastic differential equation (1) are interpreted in the sense of Itô stochastic integral.

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

$$N_t := \sup\big\{n \in \mathbb{N} \;\big|\; \tau_n \leq t\big\} \quad \text{for } t \in \mathbb{R}, \tag{2}$$

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 [26, 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., [26, Section 2.3]) show that it is memoryless and Markovian.

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

$$Y_{\tau_{N_t}} = C(\theta)X(\tau_{N_t}) + \nu_{\tau_{N_t}}, \qquad t \geq 0. \tag{3}$$

where $C(\theta) \in \mathbb{R}^{p\times n}$ is a constant matrix, and ν_k is a sequence of i.i.d. Gaussian noise processes and $\nu_0 \sim \mathcal{N}(0, V(\theta))$. Equation (3) is motivated by the fact that a continuous observation process $\mathrm{d}z = Cx\mathrm{d}t + d\eta$ for a Wiener process η is formally equivalent to $y_t = Cx_t + \nu_t$, with the identifications $y_t \sim \frac{\mathrm{d}z_t}{\mathrm{d}t}$ and $\nu_t \sim \frac{\mathrm{d}\eta_t}{\mathrm{d}t}$, so that ν_t is a Gaussian process; see [24, Chapter 4] for further details. Our goal is to construct the estimate $\widehat{X}_t$, and compute a probabilistic estimate of the unknown parameter θ, which minimizes the mean square estimation error, using the observations $\mathcal{Y}_{N_t} := \{Y_{\tau_k} \,|\, k \leq N_t\}$.

III. Filtering Algorithm

The estimate of the state process which minimizes the mean square estimation error 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, $\mathcal{Y}_{N_t}$. In the current setting, our task is complicated by the fact that the system description contains some unknown parameters which need to be estimated as well.

Using standard results in filtering theory, see for example [27], it follows that the minimum mean square error (MMSE) estimate $\widehat{X}_t$ subject to the the information $\mathcal{Y}_{N_t}$, is given by,

$$\begin{aligned}
\widehat{X}_t = \mathsf{E}[X_t \,|\, \mathcal{Y}_{N_t}] &= \int X_t p(X_t \,|\, \mathcal{Y}_{N_t}) \mathrm{d}X_t, \\
&= \int X_t \int_\Theta p(X_t \,|\, \theta, \mathcal{Y}_{N_t}) p(\theta \,|\, \mathcal{Y}_{N_t}) \mathrm{d}\theta \, \mathrm{d}X_t, \\
&= \int_\Theta \widehat{x}_t(\theta) p(\theta \,|\, \mathcal{Y}_{N_t}) \mathrm{d}\theta,
\end{aligned} \tag{4}$$

where $\widehat{x}_t(\theta) := \mathsf{E}[X_t \,|\, \theta, \mathcal{Y}_{N_t}]$ is the minimum mean square estimate with θ given, and $p(\theta \,|\, \mathcal{Y}_{N_t})$ denotes the probability density of θ conditioned upon the observations received up to time t.

In the sequel, it will be assumed that Θ is a discrete set, and is described as $\Theta := \{\theta_1, \theta_2, \cdots, \theta_K\}$ with K being the total number of unknown parameters. We also stipulate that the pair $(A(\theta), C(\theta))$ is observable, and $(A(\theta), G(\theta))$ is controllable, for each $\theta \in \Theta$. In this case, the foregoing equation results in

$$\widehat{X}_t := \sum_{i=1}^K \widehat{x}_t(\theta_i) p(\theta_i \,|\, \mathcal{Y}_{N_t}). \tag{5}$$

That is, $\widehat{X}_t$ is obtained by computing $\widehat{x}_t(\theta_i)$ for each value of $\theta_i \in \Theta$ and taking their weighted sum as determined by the conditional probability mass function $p(\theta_i \,|\, \mathcal{Y}_{N_t})$. The next two subsections describe how to compute each of the two terms on the right-hand side of (5) for a fixed value of $\theta_i \in \Theta$.

A. *State estimate for given parameters*

Note that the conditional density $p(X_t \,|\, \theta, \mathcal{Y}_{N_t})$ is Gaussian and $\widehat{x}_t(\theta_i)$ represents its first moment. It is described by a differential equation 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 [24, 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:

$$\dot{\widehat{x}}_t(\theta) = A(\theta)\widehat{x}_t(\theta)\mathrm{d}t, \quad t \in [\tau_{N_t}, \tau_{1+N_t}[, \tag{6a}$$

$$\widehat{x}_t^+(\theta) = \widehat{x}_t(\theta) + K_t(\theta)(y_t - C(\theta)\widehat{x}_t(\theta)), \quad t = \tau_{N_t}, \tag{6b}$$

where the injection gain $K_t(\theta)$ is defined as

$$K_t(\theta) = P_t(\theta)C(\theta)^\top M(\theta)^{-1}, \tag{7}$$

$$M(\theta) := C(\theta)P_t(\theta)C^\top(\theta) + V_t(\theta), \tag{8}$$

and the error covariance process $(P_t(\theta))_{t\geq 0}$ is described by the following equations: for $t \in [\tau_{N_t}, \tau_{1+N_t}[$, we solve

$$\dot{P}_t(\theta) = (A(\theta)P_t(\theta) + P_t(\theta)A(\theta)^\top + G(\theta)G(\theta)^\top) \tag{9a}$$

and for $t = \tau_{N_t}$, we solve

$$\widehat{P}_t^+(\theta) = P_t(\theta) - P_t(\theta)C(\theta)^\top M(\theta)^{-1}C(\theta)P_t(\theta). \tag{9b}$$

To write things more compactly later on, we adopt the formalism of writing the continuous-discrete equations (6a) and (6b) together in a single differential equation, when the jumps are driven by a Poisson counter N_t :

$$\mathrm{d}\widehat{x}_t(\theta) = A(\theta)\widehat{x}_t(\theta)\mathrm{d}t + K_t(\theta)(y_t - C(\theta)\widehat{x}_t(\theta))\mathrm{d}N_t. \tag{10}$$

Similarly, using this formalism, (9a) and (9b) can be written in a combined form as,

$$\begin{aligned}
\mathrm{d}P_t(\theta) = (A(\theta)P_t(\theta) + P_t(\theta)A(\theta)^\top + G(\theta)G(\theta)^\top)\mathrm{d}t \\
- P_t(\theta)C(\theta)^\top M(\theta)^{-1}C(\theta)P_t(\theta)\mathrm{d}N_t.
\end{aligned} \tag{11}$$

In the foregoing discussion, one makes the observation that the optimal conditional distribution with given θ 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.

As a performance metric, we now 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 [25], when the system parameters are known, we have computed the expected covariance for Poisson sampling. That result can be applied in the current setting to get the expectation of $P_t(\theta)$, for a known value of θ. In particular, we let $\mathcal{P}_t(\theta) = \mathsf{E}[P_t(\theta)]$ and a direct application of [25, Proposition 4.1] yields the following result:

$$\begin{aligned}
\dot{\mathcal{P}}_t(\theta) = A(\theta)\mathcal{P}_t(\theta) + \mathcal{P}_t(\theta)A(\theta)^\top + G(\theta)G(\theta)^\top \\
- \lambda\mathcal{P}_t(\theta)C(\theta)^\top M(\theta)^{-1}C(\theta)\mathcal{P}_t(\theta).
\end{aligned} \tag{12}$$

where $M(\theta) := (C(\theta)\mathcal{P}_t(\theta)C^\top(\theta) + V(\theta))$. We can now provide conditions in terms of the lower bounds on the mean sampling rate $\lambda > 0$ and the structural assumptions on controllability and observability of the pairs $(A(\theta), G(\theta))$ and $(A(\theta), C(\theta))$ that guarantee boundedness of $\mathcal{P}_t(\theta)$. The boundedness is also important for asymptotic analysis of the first moment of the error process $(X_t - \widehat{X}_t)_{t\geq 0}$.

B. Recursive expression for parameter estimation

The next element required for the computation of (5) is the posterior distribution $p(\theta \,|\, \mathcal{Y}_{N_t})$ for each $\theta \in \Theta$. A recursive formula for the computation of the conditional density $p(\theta \,|\, \mathcal{Y}_{N_t})$ is obtained from the Bayes' rule as follows:

$$p(\theta \,|\, \mathcal{Y}_{N_t}) = p(\theta \,|\, Y_{N_t}, \mathcal{Y}_{N_t-1}) \tag{13a}$$

$$= \frac{p(Y_{N_t} \,|\, \theta, \mathcal{Y}_{N_t-1}) p(\theta \,|\, \mathcal{Y}_{N_t-1})}{\sum_{i=1}^K p(Y_{N_t} \,|\, \theta_i, \mathcal{Y}_{N_t-1}) p(\theta_i \,|\, \mathcal{Y}_{N_t-1})} \tag{13b}$$

We next find the expression for $p(Y_{N_t} \,|\, \theta, \mathcal{Y}_{N_t-1})$ which corresponds to predicting the probability density of Y_{N_t} using the past measurement $\mathcal{Y}_{N_t-1}$ for a fixed value of θ. Let $\widehat{y}_{N_t}(\theta) := H(\theta)\widehat{x}_t(\theta)$ and $\widetilde{y}_{N_t}(\theta) := y_{N_t} - \widehat{y}_{N_t}(\theta)$. When the signal is generated by the parameter θ, we denote the covariance due to innovation by $\Omega_t(\theta) := \mathsf{E}[\widetilde{y}_{N_t(\theta)} \widetilde{y}_{N_t}^\top(\theta)]$. For each $\theta_i \in \Theta$, the expression for $\Omega_t(\theta)$ can be computed offline and equals $\Omega_t(\theta) = C(\theta)P_t(\theta)C^\top(\theta) + V(\theta)$. It can be readily checked that $p(Y_{N_t} \,|\, \theta_i, \mathcal{Y}_{N_t-1})$ is Gaussian with mean $H(\theta_i)\widehat{x}_t(\theta_i)$ and covariance $\Omega_t(\theta_i)$, so that $p(Y_{N_t} \,|\, \theta_i, \mathcal{Y}_{N_t-1}) = (2\pi)^{-p/2} |\Omega_t^{-1}(\theta_i)|^{1/2} e^{-\frac{1}{2}\widetilde{y}_{N_t}^\top(\theta_i)\Omega_t(\theta_i)\widetilde{y}(\theta_i)}$. Substituting this last expression in (13b), we calculate $p(\theta_i \,|\, \mathcal{Y}_{N_t})$ recursively as

$$p(\theta_i \,|\, \mathcal{Y}_{N_t}) = c |\Omega_t^{-1}(\theta_i)|^{1/2} e^{-\frac{1}{2}\widetilde{y}_{N_t}^\top(\theta_i)\Omega_t(\theta_i)\widetilde{y}(\theta_i)} p(\theta_i \,|\, \mathcal{Y}_{N_t-1}) \tag{14}$$

where c is a normalizing constant.

IV. Performance Analysis

For the algorithm proposed in the previous section, we now analyze the performance and in particular the asymptotic behavior of the error covariance matrices and *a posterior* probability density function of the unknown parameters conditioned upon the observation process.

A. Derivation of error covariance

We recall that the state estimate $(\widehat{X}_t)_{t \geq 0}$ is defined by equation (5). The following proposition characterizes the error covariance resulting from this estimate:

Proposition 1. *Let us consider the error covariance matrix,*

$$P_t := \mathsf{E}[(X_t - \widehat{X}_t) \cdot (X_t - \widehat{X}_t)^\top \,|\, \mathcal{Y}_{N_t}] \tag{15}$$

Then, it holds that

$$P_t = \sum_{i=1}^K \big\{ P_t(\theta_i) + P_t^{\mathrm{m}}(\theta_i) \big\} p(\theta_i \,|\, \mathcal{Y}_{N_t}), \tag{16}$$

where for each $\theta_i \in \Theta$, $P_t(\theta_i)$ is obtained by the equation (11)*, and $P_t^{\mathrm{m}}(\theta_i)$ is defined as,*

$$P_t^{\mathrm{m}}(\theta_i) := (\widehat{X}_t - \widehat{x}_t(\theta_i)) \cdot (\widehat{X}_t - \widehat{x}_t(\theta_i))^\top. \tag{17}$$

Proof. By definition, we have

$$\begin{aligned} P_t &= \mathsf{E}[(X_t - \widehat{X}_t) \cdot (X_t - \widehat{X}_t)^\top \,|\, \mathcal{Y}_{N_t}] \\ &= \sum_{i=1}^K P_t^{\mathrm{a}}(\theta_i) p(\theta_i \,|\, \mathcal{Y}_{N_t}), \end{aligned}$$

where $P_t^{\mathrm{a}}(\theta)$ is obtained as follows:

$$\begin{aligned} P_t^{\mathrm{a}}(\theta) &:= \mathsf{E}[(X_t - \widehat{X}_t) \cdot (X_t - \widehat{X}_t)^\top \,|\, \theta, \mathcal{Y}_{N_t}] \\ &= \mathsf{E}[X_t X_t^\top \,|\, \theta, \mathcal{Y}_{N_t}] + \widehat{X}_t \widehat{X}_t^\top - \widehat{X}_t \mathsf{E}[X_t^\top \,|\, \theta, \mathcal{Y}_{N_t}] \\ &\quad - \mathsf{E}[X_t \,|\, \theta, \mathcal{Y}_{N_t}] \widehat{X}_t^\top \\ &= \mathsf{E}[X_t X_t^\top \,|\, \theta, \mathcal{Y}_{N_t}] + \widehat{X}_t \widehat{X}_t^\top - \widehat{X}_t \widehat{x}_t^\top(\theta) - \widehat{x}_t(\theta) \widehat{X}_t^\top \\ &= \mathsf{E}[(X_t - \widehat{x}_t(\theta)) \cdot (X_t - \widehat{x}_t(\theta))^\top \,|\, \theta, \mathcal{Y}_{N_t}] \\ &\quad + (\widehat{X}_t - \widehat{x}_t(\theta)) \cdot (\widehat{X}_t - \widehat{x}_t(\theta))^\top \\ &= P_t(\theta) + P_t^{\mathrm{m}}(\theta), \end{aligned}$$

which proves the desired statement. □

In the statement of Proposition 1, P_t is introduced as a performance metric for the state estimation process with unknown parameters. Equation (16) provides a decomposition of this quantity in two terms which we can compute and the weights of these terms are determined by posterior distribution of the parameters. The first of these terms $P_t(\theta_i)$ is obtained from the Kalman-like filter by setting $\theta = \theta_i$ in the system equations. The second term $P_t^{\mathrm{m}}(\theta_i)$ denotes the cost of simulating multiple Kalman-like filters with mismatch in the parameters of the true system. Note that, for each $i = 1, \dots, K$, both terms are multiplied by the posterior distribution of θ_i.

B. Asymptotic optimality

We can now state the primary result which looks at the asymptotic behavior of expectation of P_t in (16) over sampling times (governed by a Poisson counter). To state our result, we need to introduce two main assumptions. The first of these assumptions corresponds to the convergence of posterior distribution for large t.

Assumption 1. (Consistency of Posterior) There exists $\theta^* \in \Theta$ such that

$$\lim_{t \to \infty} p(\theta^* \,|\, \mathcal{Y}_{N_t}) = 1, \quad \text{a.s.} \tag{18}$$

Although, not written explicitly, the condition (18) ensures that, for $\theta_i \neq \theta^*$, the posterior distribution $p(\theta_i \,|\, \mathcal{Y}_{N_t})$ goes to zero as t gets large.

Remark 1. Assumption 1 basically guarantees that the probability that our algorithm finds the true parameter eventually converges to 1 as we get more and more data. Such statements about the Bayesian update rules, or maximum likelihood technique are usually called *consistency results* in information theory. Some references do appear in the literature which develop conditions for checking consistency of the maximum likelihood method [15], Bayes' inference rule [13], [12, Chapter 10].

Under Assumption 1, we can now develop a qualitative result about the asymptotic behavior of the total error covariance P_t given in (16).

Theorem 2. *Consider the system* (1) *with observation process* (3) *and N_t a Poisson counter of intensity $\lambda > 0$. The optimal state estimate in mean square sense is given by $\widehat{X}_t$ in* (5)*, where $\widehat{x}_t(\theta_i)$ is obtained from* (6) *and $p(\theta_i \,|\, \mathcal{Y}_{N_t})$ is obtained from* (14)*. If Assumption 1 holds, and the mean sampling rate $\lambda > 0$ is sufficiently large, then the estimation error covariance $\mathsf{E}[P_t]$ defined in* (15) *converges to the solution of the following equation:*

$$\dot{\mathcal{P}}_t(\theta^*) = A(\theta^*)\mathcal{P}_t(\theta^*) + \mathcal{P}_t(\theta^*)A(\theta^*)^\top + G(\theta^*)G(\theta^*)^\top - \lambda\mathcal{P}_t(\theta^*)C(\theta^*)^\top M(\theta^*)^{-1}C(\theta^*)\mathcal{P}_t(\theta^*) \quad (19)$$

where $\mathcal{P}_t(\theta^*) = \mathsf{E}[P_t(\theta^*)]$, and θ^* satisfies (18).

The proof of Theorem 2 is deferred to Section IV-D, and before that we introduce the main theoretical tools that are used in establishing the result in Section IV-C.

C. Notion of extended generator

In this subsection, we discuss some analysis tools for stochastic differential equations that will be used for the proof of Theorem 2. To do so, we consider the following differential equation for a stochastic process $(x_t)_{t\geq 0}$ evolving in $\mathbb{R}^n$:

$$\mathrm{d}x_t = f(x_t)\mathrm{d}t + h(x_t, \nu_t)\mathrm{d}N_t \quad (20)$$

where N_t is a Poisson process with intensity $\lambda > 0$, and ν_{N_t} is a sequence of i.i.d. random processes 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 [28, 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 $\psi : \mathbb{R}^n \to \mathbb{R}$, it holds that*

$$\mathrm{d}\psi = \langle \nabla\psi(x), f(x)\rangle \mathrm{d}t + \Big[\psi(x+h(x,\nu)) - \psi(x)\Big]\mathrm{d}N_t. \quad (21)$$

Ito's chain rule describes the evolution of the function ψ evaluated along the solution of the stochastic differential equation (20). However, to study the evolution of expectation of $\psi(x_t)$, with x_t described by (20), we need to consider the extended generator as defined below:

Definition 1 (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 (20) is the linear operator $\psi \mapsto \mathcal{L}\psi$ defined by

$$\mathbb{R}^n \ni z \mapsto \mathcal{L}\psi(z) \in \mathbb{R}$$
$$\mathcal{L}\psi(z) := \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon}\Big(\mathsf{E}\big[\psi\big(x(t+\varepsilon)\big) \,\big|\, x(t) = z\big] - \psi(z)\Big). \quad (22)$$

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

$$\mathsf{E}\big[\psi(x(t))\big] = \mathsf{E}\big[\psi(x(0))\big] + \mathsf{E}\Bigg[\int_0^t \mathcal{L}\psi(x(s))\,\mathrm{d}s\Bigg]. \quad (23)$$

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 [29] for a derivation in the context of Wiener process driven differential equations with renewal processes. Here, we provide a statement specifically tailored for our purposes.

Proposition 4. *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* (20) *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*

$$\mathcal{L}\psi(z) = \langle \nabla\psi(z), f(z)\rangle + \lambda\Big[\int \psi\big(z + h(z,\nu)\big)\mu(\mathrm{d}\nu) - \psi(z)\Big]. \quad (24)$$

The expression for extended generator in (24) has a rather simple interpretation: the first term $\langle \nabla\psi(z), f(z)\rangle$ corresponds to the time-derivative of ψ and relates to the continuous flow of x_t. The second term is due to the jumps in x_t, where the jump intensity λ multiplies the average (with respect to noise ν_t) difference in the value of ψ.

D. Proof of Theorem 2

For the asymptotic convergence of the error covariance, we only provide a sketch of proof. To compute the infinitesimal generator for the error covariance process $(P_t)_{t\geq 0}$, which would provide us the time evolution of $\mathsf{E}[P_t]$, where the expectation is taken with respect to Poisson counter N_t of intensity $\lambda > 0$. In what follows, let us use the notation $p_i(t) = p(\theta_i \,|\, \mathcal{Y}_{N_t})$, for $t \in [\tau_{N_t,1+\tau_{N_t}}[$. Using the expression in (16), we observe that

$$\mathsf{E}[P_t] = \sum_{i=1}^{K} \mathsf{E}[P_t(\theta_i)p_i(t)] + \mathsf{E}[P_t^{\mathrm{m}}(\theta_i)p_i(t)]. \quad (25)$$

While we already have evolution equations for $P_t(\theta_i)$ and p_i, we use Ito's chain rule from Proposition 3 to compute a differential equation for $P_t^{\mathrm{m}}(\theta_i)$. Due to the linearity of the operator $\psi \mapsto \mathcal{L}\psi$, these equations will allow us to compute the extended generator for each of the terms in the summation with θ_i being fixed.

The first term inside the summation in (25) is of the form $P_t(\theta_i)p_i(t)$, where we note that p_i changes its value only when the new observation arrives. The corresponding extended generator takes the form:

$$\mathcal{L}(P_t(\theta_i)p_i(t)) = \dot{P}_t(\theta_i)p_i(t) + \lambda(P_t^+(\theta_i)p_i(t)^+ - P_t^-(\theta_i)p_i^-)$$

Under the consistency hypothesis stated in Assumption 1, we see that the right-hand side converges to 0 for $\theta_i \neq \theta^*$, and for $\theta_i = \theta^*$, it converges to

$$\dot{P}_t(\theta^*) + \lambda(P_t^+(\theta^*) - P_t^-(\theta^*)).$$

One can use similar reasoning for analyzing the second term. We first compute $\mathrm{d}P_t^{\mathrm{m}}(\theta_i)$ and it can be shown that for $\lambda > 0$ sufficiently large, $P_t^{\mathrm{m}}(\theta_i)$ remains bounded. Under Assumption 1, one can then show that $\mathsf{E}[P_t^{\mathrm{m}} p_i(t)]$ converges to zero.

V. Conclusions and Perspectives

We considered the problem of joint state estimation and detection of unknown parameters for a class of linear stochastic systems subject to randomly sampled observation process. We use the techniques based on Bayesian inference to compute the posterior distribution of the unknown parameters conditioned upon the available observations. For each value of the unknown parameter, a standard Kalman-like filter is designed, and the overall estimate is the weighted sum of individual filters. We use the covariance of the estimation error between the true state and the computed estimate as a measure of performance. Expectation of this covariance matrix over the sampling process is computed using the infinitesimal generator, and this allows us to analyze the asymptotic behavior of our proposed algorithm.

Several interesting questions appear from this initial investigation. One needs to understand carefully the conditions under which the inference update rule for the unknown parameters can converge, and provide easy-to-check tests for such conditions. The other immediate direction of research is to study the case where the set of unknown parameters is continuous and not necessarily discrete. Some work on adaptive state estimation under such hypothesis involves quantization of the parameter space and showing that the inference rule converges to a value closest to the unknown parameter in some appropriate metric [30, Section 5]. At the same time, one could revisit the question of investigating other approaches for the search of unknown parameters. It could also be interesting to see whether the ideas presented in this paper could be merged with ongoing research on developing computationally efficient filtering algorithms [31], so that the proposed methodology could be applied to nonlinear systems as well.

References

[1] R. Kalman and R. Bucy, "New results in linear filtering and prediction theory," *Journal of Basic Engineering*, vol. 83, no. 1, pp. 95–108, 1961.

[2] R. Fitzgerald, "Divergence of the Kalman filter," *IEEE Transactions on Automatic Control*, vol. 16, no. 6, pp. 736–747, 1971.

[3] R. Mehra, "Approaches to adaptive filtering," *IEEE Transactions on Automatic Control*, vol. 17, no. 5, pp. 693–698, 1972.

[4] P. Maybeck, *Stochastic Models, Estimation, and Control*. Academic Press, 1982, vol. 1.

[5] K. Watanabe, *Adaptive Estimation and Control: Partitioning Approach*. Prentice Hall International (UK), 1992.

[6] L. Zhang, D. Sidoti, A. Bienkowski, K. R. Pattipati, Y. Bar-Shalom, and D. L. Kleinman, "On the identification of noise covariances and adaptive Kalman filtering: A new look at a 50 year-old problem," *IEEE Access*, vol. 8, pp. 59 362–59 388, 2020.

[7] D. Magill, "Optimal adaptive estimation of sampled stochastic processes," *IEEE Transactions on Automatic Control*, vol. 10, no. 4, pp. 434–439, 1965.

[8] D. Lainiotis, "Partitioning: A unifying framework for adaptive systems, I: Estimation," *Proceedings of the IEEE*, vol. 64, no. 8, pp. 1126–1143, 1976.

[9] S. Särkkä, *Bayesian Filtering and Smoothing*. Cambridge, UK: Cambridge University Press, 2013.

[10] O. Cappé, E. Moulines, and T. Rydén, *Inference in Hidden Markov Models*. Springer New York, NY, 2005.

[11] C. Parellier, A. Barrau, and S. Bonnabel, "Speeding-up backpropagation of gradients through the Kalman filter via closed-form expressions," *IEEE Transactions on Automatic Control*, vol. 68, no. 12, pp. 8171–8177, 2023.

[12] B. Anderson and J. Moore, *Optimal Filtering*. Englewood Cliffs. NJ., USA: Prentice-Hall, 1979.

[13] Y. Baram and N. Sandell, "Consistent estimation on finite parameter set with application to linear systems identification," *IEEE Transactions on Automatic Control*, vol. AC-23, no. 3, pp. 451–454, 1976.

[14] L. Liporace, "Variance of bayes estimates," *IEEE Transactions on Information Theory*, vol. 17, no. 6, pp. 665–669, 1971.

[15] P. Caines and J. Rissanen, "Maximum likelihood estimation of parameters in multivariate Gaussian stochastic processes," *IEEE Transactions on Information Theory*, vol. 20, no. 1, pp. 102–104, 1974.

[16] G. Revach, N. Shlezinger, X. Ni, A. Escoriza, R. van Sloun, and Y. Eldar, "KalmanNet: Neural network aided Kalman filtering for partially known dynamics," *IEEE Transactions on Signal Processing*, vol. 70, pp. 1532–1547, 2022.

[17] G. Gebhardt, A. Kupcsik, and G. Neumann, "The kernel Kalman rule," *Maching Learning*, vol. 108, pp. 2113–2157, 2019.

[18] S. Talebi, A. Taghvaei, and M. Mesbahi, "Data-driven optimal filtering for linear systems with unknown noise covariances," in *Poster at NeurIPS*, 2023.

[19] A. Mahajan and D. Teneketzis, "Optimal performance of networked control systems with nonclassical information structures," *SIAM Journal on Control and Optimization*, vol. 48, no. 3, pp. 1377–1404, 2009.

[20] S. Yüksel and T. Basar, *Stochastic Networked Control Systems*. New York, NY: Springer, 2013, vol. 10.

[21] T. Başar, J. Moon, J. Yong, and H. Zhang, "Nonstandard linear-quadratic decision making," in *Proceedings of 62nd IEEE Conference on Decision and Control (CDC)*, 2023, pp. 8901–8920.

[22] M. Adès, P. E. Caines, and R. P. Malhamé, "Stochastic optimal control under Poisson-distributed observations," *IEEE Transactions on Automatic Control*, vol. 45, no. 1, pp. 3–13, 2000.

[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*, vol. 49, no. 9, pp. 1453–1464, 2004.

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

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

[26] 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.

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

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

[29] 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*, 2006, pp. 606–618.

[30] R. Hawkes and J. Moore, "Performance bounds for adaptive estimation," *Proceedings of the IEEE*, vol. 64, no. 8, pp. 1143–1150, 1976.

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