# Auxiliary Particle Filtering with Variational Inference for Jump Markov Systems with Unknown Measurement Noise Covariance


Cheng Cheng
School of Astronautics
Northwestern Polytechnical University
Xi'an, China
Email: cheng.cheng@nwpu.edu.cn

Sinan Yıldırım
Faculty of Engineering and Natural Sciences
Sabancı University
Istanbul, Turkey
Email: sinanyildirim@sabanciuniv.edu

Jean-Yves Tourneret
ENSEEIHT-IRIT-TéSA
University of Toulouse
Toulouse, France
Email: jean-yves.tourneret@enseeiht.fr



***Abstract*—This paper studies an auxiliary particle filter with variational inference for jointly estimating the system mode, the state and the measurement noise covariance matrix of jump Markov systems. The joint posterior distribution of the system mode, the state and the noise covariance matrix is marginalized out with respect to the system mode. The marginalized posterior distribution of the mode is then approximated by using an auxiliary particle filter, and the state and noise covariance matrix conditionally on each particle of the mode variable are updated using variational Bayesian inference. A simulation study is conducted to compare the proposed method with state-of-the-art approaches for a target tracking scenario.**


## I. Introduction

Jump Markov systems (JMSs) have been widely investigated in the literature, especially for state-space models that are conditionally linear Gaussian models, i.e., jump Markov linear systems (JMLSs). In these systems, a finite-state Markov chain switches between different modes with an appropriate transition probability matrix (TPM). Many algorithms have been proposed to solve this state estimation problem, mainly based on Gaussian mixture approximations, such as the generalized pseudo-Bayes algorithm [1], the interacting multiple model (IMM) algorithm [2], [3], and the sequential Monte Carlo (SMC) algorithm [4]–[6]. In general, these approaches assume that the measurement noise covariance matrix is accurately known. However, in realistic contexts, changing environments result in outliers corrupting the measurements or different sensor functioning conditions. Thus the measurement noise covariance matrix can be unknown and evolve slowly during a long time interval, leading to degradations in state estimates for jump Markov systems.

Regarding the state estimation problem of JMSs with unknown measurement noise parameters, the IMM with variational Bayesian (VB) inference was studied for the joint estimation of the state and measurement noise covariance matrix of jump Markov systems, where VB inference was employed to derive mode-conditioned estimates and mode-likelihood functions in the framework of IMM [7]–[9]. An improved version of the IMM algorithm with variational inference was proposed in [10] to calculate all unknown variables according to the weighted Kullback-Leibler (KL) average of mode-conditional estimates.

This paper studies an auxiliary particle filter (APF) with VB inference (APF-VB) for jointly estimating the system mode variable, the state and the measurement noise covariance matrix of a JMLS. In order to reduce the dimension of the sampling space, the proposed approach marginalizes out the state and the noise covariance matrix from the joint posterior distribution of interest. The posterior distribution of the JMLS mode sequence is then approximated by using an APF. The joint posterior distribution of the state and the noise covariance matrix conditionally on each particle of this APF is finally handled by using VB inference.

The paper is organized as follows: Section II formulates the joint estimation problem for JMLS. Section III studies the APF with VB inference proposed to estimate the JMLS unknown parameters. The performance of the proposed approach is evaluated in Section IV using a target tracking scenario. Conclusions are finally reported in Section V.

## II. Problem Formulation

In this paper, we consider a discrete-time JMLS involving a hidden state process $\{\boldsymbol{x}_t \in \mathbb{R}^{d_x}\}_{t\geq 1}$ and the measurement process $\{\boldsymbol{y}_t \in \mathbb{R}^{d_y}\}_{t\geq 1}$ for some $d_x > 0$ and $d_y > 0$. In a JMLS, the evolution of the state and the measurement variables depend on the system mode and are defined by the following state and measurement equations:

$$\boldsymbol{x}_0|r_1 \sim f_{r_1}(\boldsymbol{x}_0),\ \boldsymbol{x}_t|(\boldsymbol{x}_{t-1}, r_t) \sim f_{r_{t-1},r_t}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}), \quad (1a)$$

$$\boldsymbol{y}_t|(\boldsymbol{x}_t, r_t) \sim g_{r_t}(\boldsymbol{y}_t|\boldsymbol{x}_t), \quad (1b)$$

for $t \geq 1$, where $r_t \in \mathcal{K} = \{1, \ldots, K\}$ is a discrete variable indicating the system mode at time $t$. The discrete-time process $\{r_t\}_{t\geq 1}$ is generally assumed to be a first-order homogeneous finite state Markov chain, with initial and transition probabilities defined by

$$\pi_{r_{t-1}r_t} := \mathbb{P}(r_t = k|r_{t-1} = l),$$

ISBN: 978-9-4645-9361-7

for all $k, l \in \mathcal{K}$ and $t > 1$. The state and measurement equations of the JMLS studied in this paper are:

$$\boldsymbol{x}_t = \boldsymbol{A}(r_t)\,\boldsymbol{x}_{t-1} + \boldsymbol{w}_{t-1}, \tag{2a}$$

$$\boldsymbol{y}_t = \boldsymbol{H}(r_t)\,\boldsymbol{x}_t + \boldsymbol{\nu}_t, \tag{2b}$$

where $t = 1, \ldots, T$ denotes the $t$th time instant, $\boldsymbol{w}_t$ and $\boldsymbol{\nu}_t$ are the process and measurement noises distributed according to zero-mean Gaussian distributions with covariance matrices $\boldsymbol{Q}_t$ and $\boldsymbol{\Sigma}_t$, i.e., $\boldsymbol{w}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_t)$ and $\boldsymbol{\nu}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_t)$, $\boldsymbol{A}(r_t)$ and $\boldsymbol{H}(r_t)$ are the system and measurement matrices associated with the $r_t$th system mode that are assumed to be known. Accordingly, $f_{r_{t-1},r_t}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$ and $g_{r_t}(\boldsymbol{y}_t|\boldsymbol{x}_t)$ in (1) can be expressed for $t \geq 1$ as:

$$f_{r_{t-1},r_t}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) := \mathcal{N}(\boldsymbol{x}_t; \boldsymbol{A}(r_t)\,\boldsymbol{x}_{t-1}, \boldsymbol{Q}_t), \tag{3a}$$

$$g_{r_t}(\boldsymbol{y}_t|\boldsymbol{x}_t) := \mathcal{N}(\boldsymbol{y}_t; \boldsymbol{H}(r_t)\,\boldsymbol{x}_t, \boldsymbol{\Sigma}_t). \tag{3b}$$

In some practical applications, a changing environment results in outliers corrupting the measurements or in different sensor functioning conditions. In order to take into account these changing environments, this paper proposes to consider an unknown measurement noise covariance matrix $\boldsymbol{\Sigma}_t$. The measurement noise covariance matrix $\boldsymbol{\Sigma}_t$ is supposed to be a diagonal matrix independent of the system mode $r_t$, denoted as $\boldsymbol{\Sigma}_t = \operatorname{diag}\left[\sigma^2_{t,1}, \ldots, \sigma^2_{t,d_y}\right]$. The diagonal elements of this matrix are assigned independent inverse gamma prior distributions leading to:

$$p(\sigma^2_{t,1}, \ldots, \sigma^2_{t,d_y}) = \prod_{j=1}^{d_y} \mathcal{IG}\left(\sigma^2_{t,j}; \alpha_{t,j}, \beta_{t,j}\right), \tag{4}$$

where $\alpha_{t,j} > 1$ and $\beta_{t,j} > 0$ are the unknown shape and scale parameters of these inverse gamma distributions. Since it is not straightforward to choose a dynamical model $p(\boldsymbol{\Sigma}_t|\boldsymbol{\Sigma}_{t-1})$, we propose to assume the following dynamics for the shape and scale parameters:

$$\alpha^-_{t,j} = \rho\alpha_{t-1,j}, \;\; \beta^-_{t,j} = \rho\beta_{t-1,j}, \tag{5}$$

where $\alpha_{t-1,j}$ and $\beta_{t-1,j}$ (for $j = 1, \ldots, d_y$) are the parameters of the inverse gamma distributions at time $t-1$ and $\rho \in (0, 1]$ is a forgetting factor. This strategy was used successfully in [11] for recursive noise adaptive Kalman filtering. Assuming that the distribution of the measurement noise covariance matrix elements is maintained for different time instants, we specifically consider the following predictive distribution for the noise covariance matrix elements:

$$p\left(\sigma^2_{t,1}, \ldots, \sigma^2_{t,d_y}|\boldsymbol{y}_{1:t-1}\right) = \prod_{j=1}^{d_y} \mathcal{IG}\left(\sigma^2_{t,j}; \alpha^-_{t,j}, \beta^-_{t,j}\right). \tag{6}$$

The aim of this work is to infer the hidden state vector $\boldsymbol{x}_t$, the unknown measurement noise covariance matrix $\boldsymbol{\Sigma}_t$ and the mode sequence $r_{1:t}$ of the JMLS, leading to the joint posterior distribution of all unknown variables given the measurement sequence $\boldsymbol{y}_{1:t} = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_t\}$, denoted as $p(\boldsymbol{x}_t, \boldsymbol{\Sigma}_t, r_{1:t}|\boldsymbol{y}_{1:t})$, which is studied in the next section.

## III. An auxiliary particle filter with VB inference for JMLS with unknown measurement noise covariance matrix

Following the concept of the marginalized particle filter [12], the joint posterior distribution of the state, the unknown measurement noise covariance and system mode variables can be factorized as follows:

$$p(\boldsymbol{x}_t, \boldsymbol{\Sigma}_t, r_{1:t}|\boldsymbol{y}_{1:t}) = p(\boldsymbol{x}_t, \boldsymbol{\Sigma}_t|r_{1:t}, \boldsymbol{y}_{1:t})\,p(r_{1:t}|\boldsymbol{y}_{1:t}), \tag{7}$$

where $\boldsymbol{x}_t$ and $\boldsymbol{\Sigma}_t$ have been marginalized out in the second term of the right hand side of (7). This paper proposes to approximate $p(r_{1:t}|\boldsymbol{y}_{1:t})$ by using an empirical density following the principle of particle filters

$$p(r_{1:t}|\boldsymbol{y}_{1:t}) \approx \sum_{i=1}^{N} \omega_t^{(i)} \delta\left(r_{1:t} - r_{1:t}^{(i)}\right), \tag{8}$$

where $N$ is the number of particles, $\delta(\cdot)$ is the Dirac delta function, $r_{1:t}^{(i)}$ is the $i$th particle path and $\omega_t^{(i)}$ is the corresponding weight at the $t$th time instant. Substituting (8) into (7), the following result is obtained:

$$p(\boldsymbol{x}_t, \boldsymbol{\Sigma}_t, r_{1:t}, |\boldsymbol{y}_{1:t}) \approx \sum_{i=1}^{N} \omega_t^i p\left(\boldsymbol{x}_t, \boldsymbol{\Sigma}_t|r_{1:t}^{(i)}, \boldsymbol{y}_{1:t}\right) \delta\left(r_{1:t} - r_{1:t}^{(i)}\right), \tag{9}$$

where $p\left(\boldsymbol{x}_t, \boldsymbol{\Sigma}_t|r_{1:t}^{(i)}, \boldsymbol{y}_{1:t}\right)$ is the joint posterior distribution of $\boldsymbol{x}_t$ and $\boldsymbol{\Sigma}_t$ for the $i$th particle path of the mode variable.

### A. An auxiliary particle filtering for the mode variable sequence $r_{1:t}$

According to Bayes' rule, $p(r_{1:t}|\boldsymbol{y}_{1:t})$ satisfies the following recursion:

$$\begin{aligned} p(r_{1:t}|\boldsymbol{y}_{1:t}) &= \frac{p(\boldsymbol{y}_t|r_t, \boldsymbol{y}_{1:t-1})\,p(r_{1:t}|\boldsymbol{y}_{1:t-1})}{p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1})} \\ &= \frac{p(\boldsymbol{y}_t|r_t, \boldsymbol{y}_{1:t-1})\,\pi_{r_{t-1}r_t}}{p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1})} p(r_{1:t-1}|\boldsymbol{y}_{1:t-1}), \end{aligned} \tag{10}$$

where $p(r_{1:t}|\boldsymbol{y}_{1:t-1}) = \pi_{r_{t-1}r_t} p(r_{1:t-1}|\boldsymbol{y}_{1:t-1})$. According to the sequential importance sampling resampling (SISR) method, the particle weights satisfy:

$$\omega_t^{(i)} \propto \omega_{t-1}^{(i)} \frac{p\left(\boldsymbol{y}_t|r_t^{(i)}, \boldsymbol{y}_{1:t-1}\right)\pi_{r_{t-1}^{(i)}r_t^{(i)}}}{q\left(r_t^{(i)}|r_{1:t-1}^{(i)}, \boldsymbol{y}_{1:t}\right)}, \tag{11}$$

where $i = 1, \ldots, N$, $q\left(r_t|r_{1:t-1}^{(i)}, \boldsymbol{y}_{1:t}\right)$ denotes the optimal importance function introduced in [13]:

$$q\left(r_t|r_{1:t-1}^{(i)}, \boldsymbol{y}_{1:t}\right) = \frac{p(\boldsymbol{y}_t|r_t, \boldsymbol{y}_{1:t-1})\,\pi_{r_{t-1}^{(i)}r_t}}{p\left(\boldsymbol{y}_t|r_{1:t-1}^{(i)}, \boldsymbol{y}_{1:t-1}\right)}, \tag{12}$$

and the associated importance weight for the $i$th particle is proportional to the predictive likelihood $\omega_t^{(i)} \propto \omega_{t-1}^{(i)} p(\boldsymbol{y}_t|r_{1:t-1}^{(i)}, \boldsymbol{y}_{1:t-1})$. Note that the importance weight $\omega_t$ does not depend on $r_t$. In the APF, the resampling/selection can be performed before extending trajectories, thus selecting the most promising trajectories before extension. Accordingly, the importance weight in (11) can be rewritten as:

$$\omega_t^{(i)} \propto \omega_{t-1}^{(i)} p(\boldsymbol{y}_t|r_{1:t-1}^{(i)}, \boldsymbol{y}_{1:t-1}) \times \frac{p\left(\boldsymbol{y}_t|r_t^{(i)}, \boldsymbol{y}_{1:t-1}\right) \pi_{r_{t-1}^{(i)} r_t^{(i)}}}{p\left(\boldsymbol{y}_t|r_{1:t-1}^{(i)}, \boldsymbol{y}_{1:t-1}\right) q\left(r_t^{(i)}|r_{1:t-1}^{(i)}, \boldsymbol{y}_{1:t}\right)}. \tag{13}$$

Furthermore, for $i = 1, \ldots, N$, $\lambda_t^{(i)}$ can be calculated for implementing an auxiliary variable resampling, i.e.,

$$\lambda_t^{(i)} \propto \omega_{t-1}^{(i)} p(\boldsymbol{y}_t|r_{1:t-1}^{(i)}, \boldsymbol{y}_{1:t-1}), \tag{14}$$

with $\sum_{i=1}^N \lambda_t^{(i)} = 1$. The resampling procedure is implemented by multiplying/discarding particles $r_{1:t-1}^{(1:N)}$ with respect to high/low weights $\lambda_t^{(1:N)}$, obtaining $N$ new particles $\widetilde{r}_{1:t-1}^{(1:N)}$. Then, $\widetilde{r}_t^{(i)}$ can be generated according to $q(r_t|\widetilde{r}_{1:t-1}^{(i)}, \boldsymbol{y}_{1:t})$.

The conditional density $p(\boldsymbol{y}_t|r_{1:t-1}^{(i)}, \boldsymbol{y}_{1:t-1})$ (14) can be obtained by marginalization:

$$p(\boldsymbol{y}_t|r_{1:t-1}^{(i)}, \boldsymbol{y}_{1:t-1}) = \sum_{r_t \in \mathcal{K}} \pi_{r_{t-1}^{(i)} r_t} \iint g_{r_t}(\boldsymbol{y}_t|\boldsymbol{x}_t, \boldsymbol{\Sigma}_t) \times p(\boldsymbol{x}_t|i, r_t, \boldsymbol{y}_{1:t-1}) p(\boldsymbol{\Sigma}_t|i, \boldsymbol{y}_{1:t-1}) \mathrm{d}\boldsymbol{x}_t \mathrm{d}\boldsymbol{\Sigma}_t, \tag{15}$$

where $p(\boldsymbol{x}_t|i, r_t, \boldsymbol{y}_{1:t-1})$ is the predicted probability density function (pdf) of the state $\boldsymbol{x}_t$ (associated with the $r_t$th mode) for the $i$th particle at time $t$ defined by:

$$p(\boldsymbol{x}_t|i, r_t, \boldsymbol{y}_{1:t-1}) = \mathcal{N}\left(\boldsymbol{x}_t; \boldsymbol{m}_{t,r_t}^{-,(i)}, \boldsymbol{P}_{t,r_t}^{-,(i)}\right), \tag{16}$$

with

$$\begin{aligned} \boldsymbol{m}_{t,r_t}^{-,(i)} &= \boldsymbol{A}(r_t) \boldsymbol{m}_{t-1}^{(i)}, \\ \boldsymbol{P}_{t,r_t}^{-,(i)} &= \boldsymbol{A}(r_t) \boldsymbol{P}_{t-1}^{(i)} \boldsymbol{A}^{\mathrm{T}}(r_t) + \boldsymbol{Q}_{t-1}, \end{aligned} \tag{17}$$

where $\boldsymbol{m}_{t-1}^{(i)}$ and $\boldsymbol{P}_{t-1}^{(i)}$ (for $i = 1, \ldots, N$) are the mean and covariance of the state at time $t-1$. By replacing (16) into (15), the integral in (15) can be written as:

$$\int \mathcal{N}(\boldsymbol{y}_t; \boldsymbol{H}(r_t) \boldsymbol{m}_{t,r_t}^{-,(i)}, H(r_t) \boldsymbol{P}_{t,r_t}^{-,(i)} H^{\mathrm{T}}(r_t) + \boldsymbol{\Sigma}_t) \times p(\boldsymbol{\Sigma}_t|i, \boldsymbol{y}_{1:t-1}) \mathrm{d}\boldsymbol{\Sigma}_t. \tag{18}$$

It is difficult to obtain an analytic solution for the above integral. Using the unimodality of the inverse gamma distribution, we propose to approximate $p(\boldsymbol{\Sigma}_t|i, \boldsymbol{y}_{1:t-1})$ by using its first-moment, i.e.,

$$p(\boldsymbol{\Sigma}_t|i, \boldsymbol{y}_{1:t-1}) \approx \delta\left(\boldsymbol{\Sigma}_t - \langle \boldsymbol{\Sigma}_t^{-,(i)} \rangle\right) \tag{19}$$

with

$$\begin{aligned} \langle \boldsymbol{\Sigma}_t^{-,(i)} \rangle_{\boldsymbol{\Sigma}_t} &= \int \boldsymbol{\Sigma}_t \, p(\boldsymbol{\Sigma}_t|i, \boldsymbol{y}_{1:t-1}) \mathrm{d}\boldsymbol{\Sigma}_t \\ &= \operatorname{diag}\left[\beta_{t,1}^{-,(i)}/(\alpha_{t,1}^{-,(i)} - 1), \ldots, \beta_{t,d_y}^{-,(i)}/(\alpha_{t,d_y}^{-,(i)} - 1)\right]. \end{aligned} \tag{20}$$

where $\langle \cdot \rangle_a = \int (\cdot) \, q(a) \mathrm{d}a$ denotes the expectation with respect to the distribution of $a$. After substituting (19) into (18), one obtains:

$$\psi_{r_t}(\boldsymbol{y}_t; r_{t-1}^{(i)}, \boldsymbol{y}_{1:t-1}) := \mathcal{N}(\boldsymbol{y}_t; H(r_t) \boldsymbol{m}_{t,r_t}^{-,(i)}, H(r_t) \boldsymbol{P}_{t,r_t}^{-,(i)} H^{\mathrm{T}}(r_t) + \langle \boldsymbol{\Sigma}_t^{-,(i)} \rangle), \tag{21}$$

which leads to the following approximation:

$$p(\boldsymbol{y}_t|r_{1:t-1}^{(i)}, \boldsymbol{y}_{1:t-1}) \approx \sum_{r_t \in \mathcal{K}} \pi_{r_{t-1}^{(i)} r_t} \psi_{r_t}(\boldsymbol{y}_t; r_{t-1}^{(i)}, \boldsymbol{y}_{1:t-1}). \tag{22}$$

In addition to yielding accurate auxiliary weights, the approximation of the predictive likelihood in (21) also provides an importance function for $r_t$, i.e.,

$$q(r_t|\widetilde{r}_{1:t-1}^{(i)}, \boldsymbol{y}_{1:t}) \propto \pi_{\widetilde{r}_{t-1}^{(i)} r_t} \psi_{r_t}(\boldsymbol{y}_t; \widetilde{r}_{t-1}^{(i)}, \boldsymbol{y}_{1:t-1}). \tag{23}$$

Accordingly, the importance weight in (13) can be updated as:

$$\omega_t^{(i)} \propto \frac{\pi_{\widetilde{r}_{t-1}^{(i)} \widetilde{r}_t^{(i)}} \psi_{r_t}(\boldsymbol{y}_t; \widetilde{r}_{t-1}^{(i)}, \boldsymbol{y}_{1:t-1})}{\sum_{r_t \in \mathcal{K}} \pi_{r_{t-1}^{(i)} r_t} \psi_{r_t}(\boldsymbol{y}_t; r_{t-1}^{(i)}, \boldsymbol{y}_{1:t-1})}, \tag{24}$$

where $\widetilde{r}_t^{(i)} \sim \pi_{\widetilde{r}_{t-1}^{(i)} r_t} \psi_{r_t}(\boldsymbol{y}_t; \widetilde{r}_{t-1}^{(i)}, \boldsymbol{y}_{1:t-1})$ for $i = 1, \ldots, N$.

### *B. VB approximation for the joint distribution of $\mathbf{x}_t$ and $\boldsymbol{\Sigma}_t$*

Since it is difficult to calculate the conditional distribution $p\left(\boldsymbol{x}_t, \boldsymbol{\Sigma}_t|r_t^{(i)}, \boldsymbol{y}_{1:t}\right)$ in closed form, we propose to use an approximation resulting from VB inference [14]. The target posterior density of $(\boldsymbol{x}_t, \boldsymbol{\Sigma}_t)$ associated with the $r_t$th mode at time $t$ is defined by $p(\boldsymbol{x}_t, \boldsymbol{\Sigma}_t|r_t, \boldsymbol{y}_{1:t})$ and the corresponding VB approximation density is denoted as $q_{\mathrm{VB}}(\boldsymbol{x}_t, \boldsymbol{\Sigma}_t)$. We propose to factorize $q_{\mathrm{VB}}$ using single-variable factors [15], i.e., $q_{\mathrm{VB}}(\boldsymbol{x}_t, \boldsymbol{\Sigma}_t) := q_1(\boldsymbol{x}_t) q_2(\boldsymbol{\Sigma}_t)$, where $q_1(\boldsymbol{x}_t) = \mathcal{N}(\boldsymbol{x}_t; \boldsymbol{m}_t, \boldsymbol{P}_t)$ and $q_2(\boldsymbol{\Sigma}_t) = \prod_{j=1}^{d_y} \mathcal{IG}(\sigma_{t,j}^2; \alpha_{t,j}, \beta_{t,j})$. According to VB inference, the logarithm of the marginal likelihood $\ln p(\boldsymbol{y}_t|r_t, \boldsymbol{y}_{1:t-1})$ can be defined by using the following identity

$$\ln p(\boldsymbol{y}_t|r_t, \boldsymbol{y}_{1:t-1}) = \mathcal{L} + \mathrm{KL}\left[q_{\mathrm{VB}}(\boldsymbol{x}_t, \boldsymbol{\Sigma}_t), p(\boldsymbol{x}_t, \boldsymbol{\Sigma}_t|r_t, \boldsymbol{y}_{1:t})\right] \tag{25}$$

with

$$\mathcal{L} = \iint q_{\mathrm{VB}}(\boldsymbol{x}_t, \boldsymbol{\Sigma}_t) \ln \frac{p(\boldsymbol{y}_t, \boldsymbol{x}_t, \boldsymbol{\Sigma}_t|r_t, \boldsymbol{y}_{1:t-1})}{q_{\mathrm{VB}}(\boldsymbol{x}_t, \boldsymbol{\Sigma}_t)} \mathrm{d}\boldsymbol{x}_t \mathrm{d}\boldsymbol{\Sigma}_t \tag{26}$$

and

$$\mathrm{KL}\left[q_{\mathrm{VB}}(\boldsymbol{x}_t, \boldsymbol{\Sigma}_t), p(\boldsymbol{x}_t, \boldsymbol{\Sigma}_t|r_t, \boldsymbol{y}_{1:t})\right] = \iint q_{\mathrm{VB}}(\boldsymbol{x}_t, \boldsymbol{\Sigma}_t) \ln \frac{q_{\mathrm{VB}}(\boldsymbol{x}_t, \boldsymbol{\Sigma}_t)}{p(\boldsymbol{x}_t, \boldsymbol{\Sigma}_t|r_t, \boldsymbol{y}_{1:t})} \mathrm{d}\boldsymbol{x}_t \mathrm{d}\boldsymbol{\Sigma}_t, \tag{27}$$

where $\mathcal{L}$ is a variational objective function used in variational inference, $\mathrm{KL}\left[q_{\mathrm{VB}}(\boldsymbol{x}_t, \boldsymbol{\Sigma}_t), p(\boldsymbol{x}_t, \boldsymbol{\Sigma}_t|r_t, \boldsymbol{y}_{1:t})\right]$ is the KL divergence between the true posterior and its approximation. Since the KL divergence is non-negative, minimizing the KL divergence can be achieved by maximizing the variational objective function $\mathcal{L}$ [16]. Accordingly, maximizing $\mathcal{L}$ can be achieved by computing expectations of

$\ln p\left(\boldsymbol{y}_t, \boldsymbol{x}_t, \boldsymbol{\Sigma}_t | r_t, \boldsymbol{y}_{1:t-1}\right)$ with respect to $q_1\left(\boldsymbol{x}_t\right)$ and $q_2\left(\boldsymbol{\Sigma}_t\right)$ in turn, i.e.,

$$\ln q_1\left(\boldsymbol{x}_t\right) \propto \int q_2\left(\boldsymbol{\Sigma}_t\right) \ln p\left(\boldsymbol{y}_t, \boldsymbol{x}_t, \boldsymbol{\Sigma}_t | r_t, \boldsymbol{y}_{1:t-1}\right) \mathrm{d}\boldsymbol{\Sigma}_t, \quad (28)$$

$$\ln q_2\left(\boldsymbol{\Sigma}_t\right) \propto \int q_1\left(\boldsymbol{x}_t\right) \ln p\left(\boldsymbol{y}_t, \boldsymbol{x}_t, \boldsymbol{\Sigma}_t | r_t, \boldsymbol{y}_{1:t-1}\right) \mathrm{d}\boldsymbol{x}_t. \quad (29)$$

By evaluating the integrals in (28) and (29), the parameters $\boldsymbol{m}_t$, $\boldsymbol{P}_t$, $\alpha_{t,j}$ and $\beta_{t,j}$ for $q_1(\boldsymbol{x}_t)$ and $q_2(\boldsymbol{\Sigma}_t)$ can be updated according to the following equations,

$$\boldsymbol{m}_t = \boldsymbol{m}_{t,r_t}^{-} + K_{t,r_t}\left(\boldsymbol{y}_t - \boldsymbol{H}(r_t)\boldsymbol{m}_{t,r_t}^{-}\right), \quad (30)$$

$$\boldsymbol{P}_t = \boldsymbol{P}_{t,r_t}^{-} - K_{t,r_t}\boldsymbol{H}(r_t)\boldsymbol{P}_{t,r_t}^{-}, \quad (31)$$

$$\alpha_{t,j} = \alpha_{t,j}^{-} + \frac{1}{2}, \quad (32)$$

$$\beta_{t,j} = \beta_{t,j}^{-} + \frac{1}{2}\left\{\left[\boldsymbol{y}_t - \boldsymbol{H}(r_t)\boldsymbol{m}_t\right]_j^2 + \left[\boldsymbol{H}(r_t)\boldsymbol{P}_t\boldsymbol{H}^{\mathrm{T}}(r_t)\right]_{jj}\right\}, \quad (33)$$

where $j = 1, \ldots, d_y$, $[\cdot]_j$ denotes the $j$th element of a vector, $[\cdot]_{jj}$ is the $j$th diagonal element of a matrix and $K_{t,r_t}$ is defined as follows:

$$K_{t,r_t} = \boldsymbol{P}_{t,r_t}^{-}\boldsymbol{H}^{\mathrm{T}}(r_t)\left(\boldsymbol{H}(r_t)\boldsymbol{P}_{t,r_t}^{-}\boldsymbol{H}^{\mathrm{T}}(r_t) + \widehat{\boldsymbol{\Sigma}}_t\right)^{-1} \quad (34)$$

with

$$\widehat{\boldsymbol{\Sigma}}_t = \operatorname{diag}\left[\beta_{t,1}/\alpha_{t,1}, \ldots, \beta_{t,d_y}/\alpha_{t,d_y}\right]. \quad (35)$$

Thus $p\left(\boldsymbol{x}_t, \boldsymbol{\Sigma}_t | r_{1:t}, \boldsymbol{y}_{1:t}\right)$ can be approximated by iteratively calculating (30)-(33) until an iteration stopping rule is satisfied. Finally, the proposed APF-VB algorithm for JMLS with unknown measurement noise covariance matrix is summarized in Alg. 1.

---

**Algorithm 1:** APF with VB inference for JMLS - a single time step

---

**Input:** Particles from time $t-1$: $r_{1:t-1}^{(1:N)}$; Variational inference parameters for the particles at time $t-1$: $\boldsymbol{m}_{t-1}^{(1:N)}$, $\boldsymbol{P}_{t-1}^{(1:N)}$, $\{\alpha_{t-1,j}^{(1:N)}, \beta_{t-1,j}^{(1:N)}\}_{j=1}^{d_y}$.

**Output:** Particles for time $t$: $\tilde{r}_t^{(1:N)}$; Variational inference for the particles at time $t$: $\boldsymbol{m}_t^{(1:N)}$, $\boldsymbol{P}_t^{(1:N)}$ $\{\alpha_{t,j}^{(1:N)}, \beta_{t,j}^{(1:N)}\}_{j=1}^{d_y}$.

**for** $i = 1, \ldots, N$ **do**

  Compute $\{\boldsymbol{m}_{t,k}^{-,(i)}, \boldsymbol{P}_{t,k}^{-,(i)}\}_{k=1}^{K}$, $\{\alpha_{t,j}^{-,(i)}, \beta_{t,j}^{-,(i)}\}_{j=1}^{d_y}$ by using (17) and (5), respectively.

  Compute $\lambda_t^{(i)}$ by using (22) and normalize $\widetilde{\lambda}_t^{(i)} = \lambda^{(i)} / \sum_{i=1}^{N}(\lambda^{(i)})$.

Multiply/discard particles $r_{1:t-1}^{(1:N)}$, $\boldsymbol{m}_{t-1}^{(1:N)}$, $\boldsymbol{P}_{t-1}^{(1:N)}$, $\{\alpha_{t-1,j}^{(1:N)}, \beta_{t-1,j}^{(1:N)}\}_{j=1}^{d_y}$ with respect to high/low weights $\widetilde{\lambda}_t^{(1:N)}$ to obtain $N$ particles $\widetilde{r}_{1:t-1}^{(1:N)}$, $\widetilde{\boldsymbol{m}}_{t-1}^{(1:N)}$, $\widetilde{\boldsymbol{P}}_{t-1}^{(1:N)}$, $\{\widetilde{\alpha}_{t-1,j}^{(1:N)}, \widetilde{\beta}_{t-1,j}^{(1:N)}\}_{j=1}^{d_y}$ with equal weights.

**for** $i = 1, \ldots, N$ **do**

  Sample $\widetilde{r}_t^{(i)}$ according to (23).

  **repeat**

    Update $\widetilde{\boldsymbol{m}}_t^{(i)}$, $\widetilde{\boldsymbol{P}}_t^{(i)}$ according to (30) and (31).

    Update $\{\widetilde{\alpha}_{t,j}^{(i)}, \widetilde{\beta}_{t,j}^{(i)}\}_{j=1}^{d_y}$ according to (32) and (33).

  **until** *convergence*;

Compute $\omega_t^{(i)}$ according to (24) and normalize $\widetilde{\omega}_t^{(i)} = \omega_t^{(i)} / \sum_{i=1}^{N} \omega_t^{(i)}$

---

## IV. Simulation Results

The proposed APF-VB algorithm is evaluated using a simplified illustrative example of a moving target with Markov switching acceleration considered in [17], [18]. The state and measurement equations of this example are defined by:

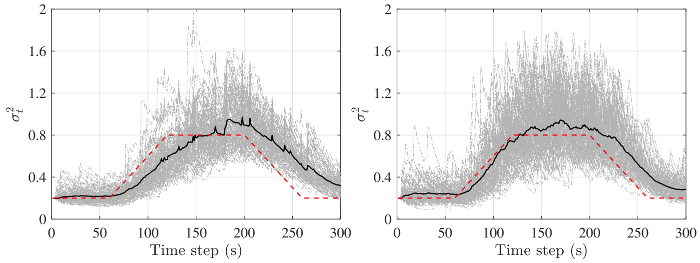
$$\begin{aligned} \boldsymbol{x}_t &= \boldsymbol{A}\boldsymbol{x}_{t-1} + \boldsymbol{\Gamma}(a_t + \omega_t), \\ y_t &= \boldsymbol{H}\boldsymbol{x}_t + \nu_t, \end{aligned} \quad (36)$$

with

$$\boldsymbol{A} = \begin{pmatrix} 1 & \Delta \\ 0 & 1 \end{pmatrix}, \ \boldsymbol{\Gamma} = \begin{pmatrix} \frac{\Delta^2}{2} \\ \Delta \end{pmatrix}, \ \boldsymbol{H} = [1, 0],$$

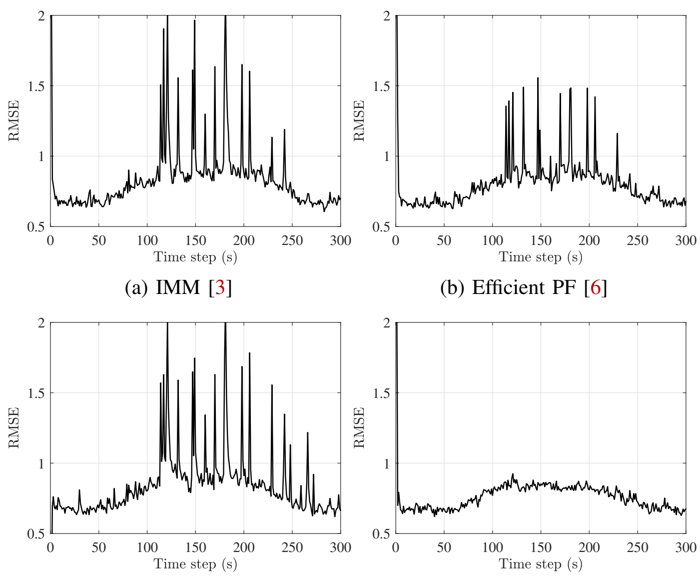
where $t = 1, \ldots, T$, the simulation time is $T = 300$, $\Delta = 1$ is the sampling period, the 2D state $\boldsymbol{x}_t = (p_t, v_t)^{\mathrm{T}}$ includes the target position in $m$ and the velocity in $m/s$, and $a_t$ denotes the acceleration in $m/s^2$. Three different values of the acceleration are considered, i.e., $a_1 = -10$, $a_2 = 0$ and $a_3 = 10$. The switching between the three models is governed by a first-order homogeneous Markov chain with known transition probabilities $\pi_{kk} = 0.9$ and $\pi_{kl} = 0.05$ $(k \neq l)$ for $k, l \in \{1, 2, 3\}$. The process and measurement noises are defined by $\omega_t \sim \mathcal{N}(0, 0.1)$ and $\nu_t \sim \mathcal{N}(0, \sigma_t^2)$ with a time-varying variance $\sigma_t^2$, whose evolution is displayed in Fig. 1. The proposed approach is compared to the IMM studied [3], the efficient particle filter studied in [6] and IMM with variational Bayes (IMM-VB) [9]. We assume here that the efficient particle filter and the IMM know the measurement noise variance, which corresponds to an ideal situation. The forgetting factor, maximum number of iterations and stopping threshold in VB inference for the APF-VB and the IMM-VB are set to 0.9, 3 and 0.01, respectively. The number of particles used in the particle filter is 200 and $N_{\mathrm{m}} = 100$ Monte Carlo runs have been generate to compute the root mean square errors (RMSEs) of the estimates defined by $\sqrt{(N_{\mathrm{m}})^{-1}\sum_{m=1}^{N_{\mathrm{m}}}\|\widehat{\boldsymbol{x}}_t(m) - \boldsymbol{x}_t\|^2}$, where $\widehat{\boldsymbol{x}}_t(m)$ is the $m$th state estimate.

Fig. 1 displays the estimates of the noise variances obtained with the proposed approach and IMM-VB. The means of the noise variance estimates were also computed, which are compared to the true values of the variances. The proposed approach APF-VB provides a better performance than IMM-VB for tracking the variance changes. Fig. 2 displays the RMSEs of state estimates obtained with different approaches. The proposed approach provides smoother estimates than the other methods, which results in better state estimation accuracy. It seems that a particle filter combined with VB iterations better estimates the state vector and the noise variance conditionally on the particles associated with the system mode $r_t$ when the system mode of the JMLS switches. Finally, the runtimes




(a) IMM-VB [9] (b) APF-VB

Fig. 1: Estimates of the measurement noise variances obtained with the IMM-VB [9] and the APF-VB. Noise variance estimates for 100 MC simulation runs: gray dash-dotted lines; Means of the noise variance estimates (100 MC simulation runs): black solid line; True values: red dashed line.




(a) IMM [3] (b) Efficient PF [6]

(c) IMM-VB [9] (d) APF-VB

Fig. 2: RMSEs of the state estimates for different approaches.

TABLE I: Runtime of different approaches.

| Approaches | Runtime of 100 MC simulations (in s) |
|---|---|
| IMM | 5.32 |
| Efficient particle filter | 85.9 |
| IMM-VB | 10.2 |
| APF-VB | 145.5 |

of 100 MC simulations obtained with different approaches are reported in Table I. This table allows the additional computational complexity of APF-VB (with respect to IMM-VB) to be evaluated. Although variational inference is used to approximate the joint posterior distribution of $\boldsymbol{x}_t$ and $\boldsymbol{\Sigma}_t$ for both APF-VB and IMM-VB, the VB step in the proposed approach is implemented conditionally on the particles of the system mode $r_t$, which are generated according to the APF. This strategy allows a better estimation performance for APF-VB at the price of a higher computational cost.

## V. Conclusions

An auxiliary particle filter with variational inference was investigated for obtaining the joint posterior distribution of the system mode, the state and the measurement noise covariance matrix of a jump Markov linear system. The proposed approach was compared to three algorithms of the state-of-the-art, namely IMM, IMM with variational inference, and the efficient particle filter, providing accurate estimates of the state and measurement noise variance at the price of a higher computational cost. Our future work will be devoted to implementing the proposed approach for jump Markov systems with unknown process and measurement noise covariance matrices. Applying the proposed approach for jump Markov nonlinear systems is also under investigation.

## References


[1] Y. Bar-Shalom and X. Li, *Multitarget-Multisensor Tracking: Principles and Techniques*. Storrs: YBS Publishing, 1995.

[2] E. Mazor, A. Averbuch, Y. Bar-Shalom, and J. Dayan, “Interacting multiple model methods in target tracking: A survey,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 34, no. 1, pp. 103–123, 1998.

[3] Y. Bar-Shalom, X. Li, and T. Kirubarajan, *Estimation with Appplication to Tracking and Navigation: Theory, Algorithm and Software*. New York: John Weily & Sons, 2004, ch. The multiple model approach, pp. 441–475.

[4] A. Doucet, N. J. Gordon, and V. Krishnamurthy, “Particle filters for state estimation of jump Markov linear systems,” *IEEE Trans. Signal Process.*, vol. 49, no. 3, pp. 613–624, 2001.

[5] C. Andrieu, M. Davy, and A. Doucet, “Efficient particle filtering for jump Markov systems: Application to time-varying autoregressions,” *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1762–1770, 2003.

[6] Y. Cheng and T. Singh, “Efficient particle filtering for road-constrained target tracking,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 43, no. 4, pp. 1454–1469, 2007.

[7] W. Li and Y. Jia, “Adaptive filtering for jump Markov systems with unknown noise covariance,” *IET Contr. Theory & Appl.*, vol. 7, no. 13, pp. 1765–1772, 2013.

[8] K. Li, S. Zhao, C. K. Ahn, and F. Liu, “State estimation for jump Markov nonlinear systems of unknown measurement data covariance,” *J. Franklin Inst.*, vol. 358, no. 2, pp. 1673–1691, 2021.

[9] C. Shen and W. Huang, “A Kullback-Leibler-based IMM information filter for the jump Markov system with unknown noise,” *Circuits Syst. Signal Process.*, vol. 37, no. 9, pp. 4065–4081, 2018.

[10] G. Wang, X. Wang, and Y. Zhang, “Variational Bayesian IMM-filter for JMSs with unknown noise covariances,” *IEEE Trans. Aerosp. Electron. Syst,.*, vol. 56, no. 2, pp. 1652–1660, 2020.

[11] S. Särkkä and A. Nummenmaa, “Recursive noise adaptive Kalman filtering by variational Bayesian approximations,” *IEEE Trans. Automat. Contr.*, vol. 54, no. 3, pp. 596–600, 2009.

[12] T. Schön, F. Gustafsson, and P.-J. Nordlund, “Marginalized particle filters for mixed linear/nonlinear state-space models,” *IEEE. Trans. Signal Process.*, vol. 53, no. 7, pp. 2279–2289, 2005.

[13] O. Cappé, S. J. Godsill, and E. Moulines, “An overview of existing methods and recent advances in sequential Monte Carlo,” *Proc. of the IEEE*, vol. 95, no. 5, pp. 899–924, 2007.

[14] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, “The variational approximation for Bayesian inference,” *IEEE Signal Process. Mag.*, vol. 25, no. 6, pp. 131–146, 2008.

[15] C. W. Fox and S. J. Roberts, “A tutorial on variational Bayesian inference,” *Artif. Intell. Rev.*, vol. 38, pp. 85–95, 2012.

[16] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer Publishing, 2006, ch. Approximate Inference, pp. 461–517.

[17] V. P. Jilkov and X. R. Li, “Online Bayesian estimation of transition probabilities for Markovian jump systems,” *IEEE Trans. Signal Process.*, vol. 52, no. 6, pp. 1620–1630, 2004.

[18] G. Wang, “ML estimation of transition probabilities in jump Markov systems via convex optimization,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 46, no. 3, pp. 1492–1502, 2010.