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

Asynchronous Variational-Bayes Kalman Filtering

Marcus Greiff, Karl Berntorp*

***Abstract*— We consider the joint state and measurement-noise parameter estimation problem for nonlinear state-space models with asynchronous, variable-rate, and independent measurement sources. We approach the problem using variational Bayes Kalman filters (VB-KFs). By leveraging that the measurements from different sources are independent, we develop an asynchronous VB-KF (AVB-KF), which processes measurements from different sources sequentially and at a variable rate. Hence, in the measurement update step, we only update the noise parameters of measurements that have been processed at a particular time step. This results in faster computations, especially as the measurement dimension and the number of sensors grow. We validate the approach on a realistic application of autonomous mobile-robot platooning, where we perform fusion of multiple sensor modalities with time-varying noise characteristics. The results indicate more than a factor of two improvements measured as a time-averaged absolute error compared to a nonadaptive implementation.**

I. Introduction

Variational Bayes Kalman filters (VB-KFs) are noise-adaptive KFs that estimate the noise parameters jointly with the state of a dynamical system subject to Gaussian-assumed process and measurement noise recursively over time using the VB framework [1]–[3]. VB-KFs approximate the joint posterior distribution of the state and the noise variances by a factorized free-form distribution, where at each time step the state and the noise covariance are updated via a fixed-point iteration of the KF measurement update.

In this paper, we consider the joint recursive estimation of the state $\boldsymbol{x}_k \in \mathbb{R}^{n_x}$ and the measurement-noise covariance matrix $\boldsymbol{\Sigma}_k \in \mathbb{R}^{n_y \times n_y}$ at time step k, based on the measurement vector $\boldsymbol{y}_k \in \mathbb{R}^{n_y}$ and the input signal $\boldsymbol{u}_k \in \mathbb{R}^{n_u}$ related by a nonlinear state-space model

$$\boldsymbol{x}_{k+1} = \boldsymbol{f}(\boldsymbol{x}_k, \boldsymbol{u}_k, h_k) + \boldsymbol{w}_k, \tag{1a}$$

$$\boldsymbol{y}_k = \boldsymbol{h}(\boldsymbol{x}_k) + \boldsymbol{e}_k. \tag{1b}$$

The state $\boldsymbol{x}_k$ at each time step k is observed through the measurement $\boldsymbol{y}_k$. The nonlinear functions $\boldsymbol{f} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{\geq 0} \mapsto \mathbb{R}^{n_x}$ and $\boldsymbol{h} : \mathbb{R}^{n_x} \mapsto \mathbb{R}^{n_y}$ are known. The process noise $\boldsymbol{w}_k$ and measurement noise $\boldsymbol{e}_k$ are Gaussian distributed with covariance $\boldsymbol{Q}(h_k)$ and $\boldsymbol{\Sigma}_k$ according to $\boldsymbol{w}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}(h_k))$ and $\boldsymbol{e}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_k)$, respectively, and $h_k = t_k - t_{k-1}$ is the sampling period between time t_k and t_{k-1}.

Unlike prior work on VB-KFs, not all measurements are available at each time step k, and the time duration between two time steps k and $k+1$ varies. Specifically, there are N multi-dimensional independent measurement channels constituting $\boldsymbol{y}_k^1, \boldsymbol{y}_k^2, \ldots, \boldsymbol{y}_k^N$, and these are sampled at (possibly) differing rates. We propose an extension to the VB-KFs in [4] for multi-dimensional and independent measurement channels, using a special factorization of the posterior $p(\boldsymbol{x}_k, \boldsymbol{\Sigma}_k | \boldsymbol{y}_{0:k})$. The resulting method bears resemblance to the sequential-update KF [5], since it updates submatrices in a blockdiagonal $\boldsymbol{\Sigma}_k$. The case of differing measurement update rates has several applications, for example, in late fusion of measurements originating from different sensor modalities. We demonstrate the applicability of our approach to late fusion with autonomous mobile robots (AMRs), where measurements from different sensing modalities arrive at different rates, and where the measurement reliability varies with time.

One of the early works on adaptive KF for noise identification is [6]. Approaches based on variational approximations and the Student-t distribution can be found in [7], [8]. Some closely aligned work is [3], which is a VB-KF for linear systems, and [4], which extends [3] to nonlinear and multivariate systems. Several noise-adaptive particle-filter (PF) based solutions have been proposed. Notable contributions include [9], which is based on augmentation of the state with the parameters to be estimated, and [10], which considers static parameter estimation using the marginalized PF. To overcome the path degeneracy of estimating static quantities in PFs, [11] considers the role of exponential forgetting in an adaptive marginalized PF. This work has been extended to dependent noise sources in [12] and consideration of partially unknown state-space models [13], [14].

All of the aforementioned works have considered a fixed sampling period among measurements. For a known noise covariance, there exist several prior works that can handle differing measurement rates, such as [15] for Gaussian filters and [16] for PFs. A related set of works is those that concern out-of-sequence measurement processing, which handle measurements that arrive out of order, with or without known time stamps [17]–[20]. However, this line of research also assumes known noise covariances.

Notation: Vectors are denoted by $\boldsymbol{x} \in \mathbb{R}^n$, with $[\boldsymbol{x}]_i$ being the i^{th} element of $\boldsymbol{x}$. Matrices are indicated in bold, $\boldsymbol{X}$, and the element on row i and column j of $\boldsymbol{X}$ is $[\boldsymbol{X}]_{ij}$. We write the block-diagonal matrix of the set of matrices $\{\boldsymbol{X}_c\}_{c\in\mathcal{C}}$ as $\text{blkdiag}(\{\boldsymbol{X}_c\}_{c\in\mathcal{C}})$, and $[\{\boldsymbol{X}_c\}_{c\in\mathcal{C}}]$ denotes the vertical stacking of such set of matrices. The notation $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{x}|\boldsymbol{m}, \boldsymbol{P})$ indicates that $\boldsymbol{x}$ is a Gaussian distributed random variable with mean $\boldsymbol{m}$ and covariance $\boldsymbol{\Sigma}$ and $|\boldsymbol{P}|$ is the determinant of $\boldsymbol{P}$. Similarly, $\boldsymbol{\Sigma} \sim \mathcal{IW}(\boldsymbol{\Sigma}|\nu, \boldsymbol{V})$ means that $\boldsymbol{\Sigma}$ is Inverse-Wishart (IW) distributed, with ν degrees of freedom, scale matrix $\boldsymbol{V}$, and where $\text{Tr}(\cdot)$ is

Mitsubishi Electric Research Labs (MERL), Cambridge, MA, USA.
*Karl Berntorp, karl.o.berntorp@ieee.org.

the trace operator. Given the set of measurements $\boldsymbol{y}_{0:k} = \{\boldsymbol{y}_0, \ldots, \boldsymbol{y}_k\}$, $p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k})$ is the marginal filtering posterior of the state $\boldsymbol{x}_k$ up until time step k, corresponding to time t_k and $\boldsymbol{x}_k^-$ is the one-step prediction using measurements up until time step $k-1$. We write the expectation of $h(\boldsymbol{x})$ when $\boldsymbol{x} \sim p(\boldsymbol{x})$, $\mathbb{E}_{\boldsymbol{x}\sim p(\boldsymbol{x})}[h(\boldsymbol{x})] = \int h(\boldsymbol{x})p(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$ compactly as $\mathbb{E}_p[h(\boldsymbol{x})]$. In this notation, the Kullback-Leibler (KL) divergence of two densities $p(\boldsymbol{x})$ and $q(\boldsymbol{x})$ is $\mathrm{KL}(p(\boldsymbol{x})||q(\boldsymbol{x})) = \mathbb{E}_p[\log(p(\boldsymbol{x})/q(\boldsymbol{x}))]$. Finally, we write $(\boldsymbol{a})(\star)^\top$ for $(\boldsymbol{a})(\boldsymbol{a})^\top$.

II. Variational Bayes Kalman Filters

The VB-KF methods [3], [4] are recursive filtering techniques for joint inference of the state $\boldsymbol{x}_k$ and the noise covariance matrix $\boldsymbol{\Sigma}_k$ given the estimation model (1), and rely on the concept of conjugate priors, see Definition 1.

Definition 1 *Given a likelihood, the conjugate prior is the prior distribution such that the prior and posterior are in the same family of distributions.*

Thus, for a conjugate prior, the prior and posteriors are of the same type, and the estimation problem simplifies to updating the hyperparameters—that is, the parameters of the distribution, which can be done analytically. Lemma 1 provides an explicit expression of the conjugate prior for Gaussian likelihoods.

Lemma 1 (See [21]) *For zero-mean multivariate Gaussian data $\boldsymbol{w} \in \mathbb{R}^d$ with unknown covariance $\boldsymbol{\Sigma}$, an IW distribution defines the conjugate prior.*

VB-KFs typically approximate the joint posterior using

$$p(\boldsymbol{x}_k, \boldsymbol{\Sigma}_k|\boldsymbol{y}_{0:k}) \approx q_{\boldsymbol{x}}(\boldsymbol{x}_k)q_{\boldsymbol{\Sigma}}(\boldsymbol{\Sigma}_k), \tag{2}$$

where the state and covariance posterior are assumed independent. Using (2), VB-KFs minimize the KL divergence between the true and approximate posterior,

$$\min_{q_{\boldsymbol{x}}, q_{\boldsymbol{\Sigma}}} \mathrm{KL}(q_{\boldsymbol{x}}(\boldsymbol{x}_k)q_{\boldsymbol{\Sigma}}(\boldsymbol{\Sigma}_k)||p(\boldsymbol{x}_k, \boldsymbol{\Sigma}_k|\boldsymbol{y}_{0:k})), \tag{3}$$

where, using Lemma 1, $q_{\boldsymbol{x}}(\boldsymbol{x}_k) = \mathcal{N}(\boldsymbol{x}_k|\boldsymbol{m}_k, \boldsymbol{P}_k)$ and $q_{\boldsymbol{\Sigma}}(\boldsymbol{\Sigma}_k) = \mathcal{IW}(\boldsymbol{\Sigma}_k|\nu_k, \boldsymbol{V}_k)$. By variational calculus [1], the solutions are known to proportionality,

$$q_{\boldsymbol{x}}(\boldsymbol{x}_k) \propto \exp(\mathbb{E}_{\boldsymbol{q}_{\boldsymbol{\Sigma}}}[p(\boldsymbol{x}_k, \boldsymbol{\Sigma}_k, \boldsymbol{y}_k|\boldsymbol{y}_{0:k-1})]), \tag{4a}$$

$$q_{\boldsymbol{\Sigma}}(\boldsymbol{\Sigma}_k) \propto \exp(\mathbb{E}_{\boldsymbol{q}_{\boldsymbol{x}}}[p(\boldsymbol{x}_k, \boldsymbol{\Sigma}_k, \boldsymbol{y}_k|\boldsymbol{y}_{0:k-1})]). \tag{4b}$$

Both of the right-hand sides in (4) can be evaluated to a local maxima by fixed-point iterations, as the expectations in (4) are known in closed form given the assumed form of the posterior in (2). The VB-KFs tend to differ in how the density $q_{\boldsymbol{\Sigma}}(\boldsymbol{\Sigma}_k)$ is factored, with [3] proposing a diagonal noise matrix and a product of Inverse-Gamma priors, whereas [4] uses a dense IW prior. Prior works assume that all of the measurements are sampled at the same time instant.

III. Asynchronous Variational Bayes Kalman Filters

In this section we present our proposed method for VB Kalman filtering with variable rate, possibly independent, measurements, denoted asynchronous VB-KF (AVB-KF).

A. Preliminaries

We consider N measurement channels with, possibly different, fixed dimensions,

$$\boldsymbol{y}_k^c \in \mathbb{R}^{n_c} \text{ on channel } c \in \mathbb{N}_{[1,N]}. \tag{5}$$

Measurements on these channels can be sampled at different rates, and we define the following indicator set.

Definition 2 *Let $\mathcal{C}_k \subseteq \mathbb{N}_{[1,N]}$, such that $c \in \mathcal{C}_k$ if and only if the measurement $\boldsymbol{y}_k^c$ is sampled at a time step k.*

We consider inputs in (1a). Therefore, it is possible that $\mathcal{C}_k = \emptyset$, as we may have a situation in which only an input is sampled at a time t_k. In contrast, the methods [3], [4] assume that $\mathcal{C}_k = \mathbb{N}_{[1,N]}$ for all k. To derive a VB-KF in this setting, we first specify the assumptions on the measurement model.

Assumption 1 *At each time step k, $\mathcal{C}_k$ is known and*

$$p(\boldsymbol{y}_k|\boldsymbol{x}_k, \boldsymbol{\Sigma}_k) = \prod_{c\in\mathcal{C}_k} \mathcal{N}(\boldsymbol{y}_k^c|\boldsymbol{h}^c(\boldsymbol{x}_k), \boldsymbol{\Sigma}_k^c). \tag{6}$$

In this formulation, the dimension of $\boldsymbol{y}_k$ is time varying, which affects the VB-KF derivations. We start by further factoring the posterior in (2):

$$q_{\boldsymbol{x}}(\boldsymbol{x}_k) = \mathcal{N}(\boldsymbol{x}_k|\boldsymbol{m}_k, \boldsymbol{P}_k), \tag{7a}$$

$$q_{\boldsymbol{\Sigma}}(\boldsymbol{\Sigma}_k) = \prod_{c\in\mathcal{C}_k} \mathcal{IW}(\boldsymbol{\Sigma}_k^c|\nu_k^c, \boldsymbol{V}_k^c). \tag{7b}$$

The factorization (7a) is common in the VB-KF literature (c.f. [4]). We take the factorization of $q_{\boldsymbol{\Sigma}}(\boldsymbol{\Sigma}_k)$ a step further in (7b), owing to Assumption 1. This leads to an efficient measurement-update step for independent measurements.

B. Time-Propagation

Due to the variable-rate filtering and asynchronous measurement updates, we need to adjust the time-propagation of the noise-covariance statistics. Standard VB-KFs use

$$p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) = \mathcal{N}(\boldsymbol{x}_k|\boldsymbol{f}(\boldsymbol{x}_{k-1}), \boldsymbol{Q}_k), \tag{8a}$$

$$p(\boldsymbol{\Sigma}_k|\boldsymbol{\Sigma}_{k-1}) = \mathcal{IW}(\boldsymbol{\Sigma}_k|\rho(\nu_{k-1} - \nu_r) + \nu_r, \boldsymbol{B}\boldsymbol{V}_k\boldsymbol{B}^\top), \tag{8b}$$

where $\nu_r = n_y + 1$. It is straightforward how the prediction model (8a) can be adapted based on the varying sampling period h_k. This is less clear for the prediction model (8b), which is based on heuristics [3], [4], often with $\boldsymbol{B} = \sqrt{\rho}\boldsymbol{I}$ for some $\rho \in (0, 1]$. To get a similar propagation of the IW statistics under variable sampling periods, consider two ordinary differential equations

$$(\mathrm{d}/\mathrm{d}t)\nu = -\frac{1}{\tau}(\nu - \nu_r), \tag{9a}$$

$$(\mathrm{d}/\mathrm{d}t)\boldsymbol{V} = -\frac{1}{\tau}\boldsymbol{V}. \tag{9b}$$

Discretizing (9) with sampling period h_k yields [22]

$$\nu_k = a_k\nu_{k-1} + b_k, \tag{10a}$$

$$\boldsymbol{V}_k = \boldsymbol{B}_k\boldsymbol{V}_{k-1}. \tag{10b}$$

where $a_k = \exp(-h_k\tau^{-1})$, $b_k = (1-a_k)\nu_r$, and $\boldsymbol{B}_k = \exp(-h_k\tau^{-1}/2)\boldsymbol{I}$. Eq. (10) allows the definition of prediction models emulating (8b) for variable sampling periods h_k,

$$p(\boldsymbol{\Sigma}_k|\boldsymbol{\Sigma}_{k-1}) = \mathcal{IW}(\boldsymbol{\Sigma}_k|a_k\nu_{k-1}+b_k, \boldsymbol{B}_k\boldsymbol{V}_k\boldsymbol{B}_k^\top). \quad (11)$$

Predicting the covariance using (11) results in a prediction that is independent on the sampling period h_k, and can be related directly to the original method (8b) run at a sampling period h by choosing $\tau = -h/\log(\rho)$.

C. Measurement Update

Another complexity when designing a VB-KF for asynchronous measurements is the measurement-update step. Since measurements from different channels may be collected at different times, we need to address how to update submatrices in the full covariance matrix. Unlike regular VB implementations, which operate under a model

$$\boldsymbol{y}_k = \boldsymbol{h}(\boldsymbol{x}_k) + \boldsymbol{e}_k, \qquad \boldsymbol{e}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_k),$$

we partition the measurements into *a set of models*

$$\boldsymbol{y}_k^c = \boldsymbol{h}^c(\boldsymbol{x}_k) + \boldsymbol{e}_k^c \in \mathbb{R}^{n_{y^c}}, \quad c \in \mathcal{C}_k, \quad (12)$$

according to (5). From Assumption 1, the measurement models are statistically independent, that is,

$$p(\boldsymbol{y}_k|\boldsymbol{x}_k, \boldsymbol{\Sigma}_k) = \prod_{c\in\mathcal{C}_k} \mathcal{N}(\boldsymbol{y}_k^c|\boldsymbol{h}^c(\boldsymbol{x}_k), \boldsymbol{\Sigma}_k^c).$$

Without loss of generality, we restrict $\boldsymbol{\Sigma}_k = \mathrm{blkdiag}(\cdots, \boldsymbol{\Sigma}_k^i, \boldsymbol{\Sigma}_k^j, \cdots)$ to have dense blocks ordered such that $i < j$. For simplicity, we write $\boldsymbol{\Sigma}_k = \mathrm{blkdiag}(\{\boldsymbol{\Sigma}_k^c\}_{c\in\mathcal{C}_k})$. Then, for a predicted density

$$p(\boldsymbol{x}_k, \boldsymbol{\Sigma}_k|\boldsymbol{y}_{0:k-1}) = \mathcal{N}(\boldsymbol{x}_k|\boldsymbol{m}_k^-, \boldsymbol{P}_k^-) \prod_{c\in\mathcal{C}_k} \mathcal{IW}(\boldsymbol{\Sigma}_k^c|(\nu_k^c)^-, (\boldsymbol{V}_k^c)^-),$$

we seek to approximate a posterior $p(\boldsymbol{x}_k, \boldsymbol{\Sigma}_k|\boldsymbol{y}_{0:k}) \approx q(\boldsymbol{x}_k, \boldsymbol{\Sigma}_k)$, factored as (see (7))

$$q(\boldsymbol{x}_k, \boldsymbol{\Sigma}_k) = q_{\boldsymbol{x}}(\boldsymbol{x}_k)q_{\boldsymbol{\Sigma}}(\boldsymbol{\Sigma}_k) = q_{\boldsymbol{x}}(\boldsymbol{x}_k) \prod_{c\in\mathcal{C}_k} q_{\boldsymbol{\Sigma}}^c(\boldsymbol{\Sigma}_k^c),$$

where $q_{\boldsymbol{x}}(\boldsymbol{x}_k) = \mathcal{N}(\boldsymbol{x}_k|\boldsymbol{m}_k, \boldsymbol{P}_k)$, $q_{\boldsymbol{\Sigma}}^c(\boldsymbol{\Sigma}_k) = \mathcal{IW}(\boldsymbol{\Sigma}_k^c|\nu_k^c, \boldsymbol{V}_k^c)$. Following the usual steps in the VB literature (e.g., [4]), we obtain the minimizers of

$$\{q_{\boldsymbol{x}}^\star, q_{\boldsymbol{\Sigma}}^{c\in\mathcal{C}_k,\star}\} = \underset{q_{\boldsymbol{x}}, \{q_{\boldsymbol{\Sigma}}^c\}_{c\in\mathcal{C}_k}}{\arg\min} \ \mathrm{KL}(q_{\boldsymbol{x}}(\boldsymbol{x}_k)q_{\boldsymbol{\Sigma}}(\boldsymbol{\Sigma}_k)||p(\boldsymbol{x}_k, \boldsymbol{\Sigma}_k|\boldsymbol{y}_{0:k})), \quad (13)$$

through the Euler-Lagrange (E-L) equations,

$$q_{\boldsymbol{x}}^\star(\boldsymbol{x}_k) \propto \exp\Big(\int p(\boldsymbol{y}_k, \boldsymbol{x}_k, \boldsymbol{\Sigma}_k|\boldsymbol{y}_{0:k-1}) q_{\boldsymbol{\Sigma}}(\boldsymbol{\Sigma}_k) \mathrm{d}\boldsymbol{\Sigma}_k\Big), \quad (14a)$$

$$q_{\boldsymbol{\Sigma}}^\star(\boldsymbol{\Sigma}_k) \propto \exp\Big(\int p(\boldsymbol{y}_k, \boldsymbol{x}_k, \boldsymbol{\Sigma}_k|\boldsymbol{y}_{0:k-1}) q_{\boldsymbol{x}}(\boldsymbol{x}_k) \mathrm{d}\boldsymbol{x}_k\Big). \quad (14b)$$

Theorem 1 states the main result.

Theorem 1 *The minimizers of* (13) *found through the E-L equations* (14) *for a measurement model*

$$\boldsymbol{y}_k = \boldsymbol{H}_k\boldsymbol{x}_k + \boldsymbol{e}_k, \quad \boldsymbol{e}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_k),$$

where $\boldsymbol{H}_k = [\{\boldsymbol{H}_k^c\}_{c\in\mathcal{C}_k}]$ *result in*

$$q_{\boldsymbol{x}}^\star(\boldsymbol{x}_k) = \mathcal{N}(\boldsymbol{x}_k|\boldsymbol{m}_k, \boldsymbol{P}_k),$$

where $\boldsymbol{m}_k$ *and* $\boldsymbol{P}_k$ *are found using a recursion over* $c \in \mathcal{C}_k$,

$$\boldsymbol{m}_k^c = \boldsymbol{m}_k^{c-} + \boldsymbol{K}_k^c(\boldsymbol{y}_k^c - \boldsymbol{H}_k^c\boldsymbol{m}_k^{c-}), \quad (15a)$$

$$\boldsymbol{P}_k^c = \boldsymbol{P}_k^{c-} - \boldsymbol{K}_k^c\boldsymbol{S}_k^c(\boldsymbol{K}_k^c)^\top, \quad (15b)$$

$$\boldsymbol{S}_k^c = \boldsymbol{H}_k^c\boldsymbol{P}_k^{c-}(\boldsymbol{H}_k^c)^\top + \hat{\boldsymbol{\Sigma}}_k^c, \quad (15c)$$

$$\boldsymbol{K}_k^c = \boldsymbol{P}_k^{c-}(\boldsymbol{H}_k^c)^\top(\boldsymbol{S}_k^c)^{-1}, \quad (15d)$$

$$\hat{\boldsymbol{\Sigma}}_k^c = \tfrac{1}{\nu_k^c - n_{y^c} - 1}\boldsymbol{V}_k^c. \quad (15e)$$

where $\{\boldsymbol{m}_k, \boldsymbol{P}_k\} = \{\boldsymbol{m}_k^c, \boldsymbol{P}_k^c\}$ *after the last recursion and where* $\{\boldsymbol{m}_k^c, \boldsymbol{P}_k^c\}$ *are initialized with* $\{\boldsymbol{m}_k^c, \boldsymbol{P}_k^c\} = \{\boldsymbol{m}_k^-, \boldsymbol{P}_k^-\}$, *with* $(\star)^{c-}$ *denoting the estimate at the previous iteration. Furthermore,*

$$q_{\boldsymbol{\Sigma}}^{c,\star}(\boldsymbol{\Sigma}_k^c) = \mathcal{IW}(\boldsymbol{\Sigma}_k^c|\nu_k^c, \boldsymbol{V}_k^c),$$

where

$$\nu_{k,i}^c = (\nu^c)_{k,i}^- + 1, \quad (16a)$$

$$\boldsymbol{V}_k^c = (\boldsymbol{V}_k^c)^- + (\boldsymbol{y}_k^c - \boldsymbol{H}_k^c\boldsymbol{m}_k)(\star)^\top + \boldsymbol{H}_k^c\boldsymbol{P}_k(\boldsymbol{H}_k^c)^\top. \quad (16b)$$

Remark 1 *The extension to the nonlinear measurement model* (12) *is trival, for example, using moment-matching techniques, see [4] for the case when* $\boldsymbol{\Sigma}_k$ *is dense, and is similar to the extension of KFs to nonlinear measurement models. Our validation in Sec. IV is nonlinear, both in prediction and measurement model.*

D. Algorithm Summary

The solution to (15) requires the solution to (16) and vice versa. To solve this, we follow the usual VB literature and employ a fixed-point iteration. Algorithm 1 gives the proposed method, denoted by AVB-KF. Note that if all measurements are sampled simultanously and all channels are one-dimensional, the proposed method collapses to [4]. With the recursion on Lines 14–21, the complexity is $\mathcal{O}(\max_{c\in\mathcal{C}_k}(n_{y^c})^3)$ instead of $\mathcal{O}(\sum_{c\in\mathcal{C}_k}(n_{y^c})^3)$ if updating the full covariance matrix in one step.

IV. Application to Late Fusion with Autonomous Mobile Robots

In this section we apply Algorithm 1 to late sensor fusion for autonomous mobile robots (AMRs). This is a setting where variable-rate measurement channels are pervasive, and where various sensing modalities, such as wheel odometry, distance measurements, and configuration measurements (position and orientation) are to be fused given a known prediction model. In such settings, it is clear that the measurement reliability may vary. For example, if the AMR relies on configuration measurements from a simultaneous localization and mapping (SLAM) system, the quality of a

Algorithm 1 Variable-rate Asynchronous VB-KF (AVB-KF).

Define $\boldsymbol{m}_0, \boldsymbol{P}_0, \{\nu_0^{c,-}, \boldsymbol{V}_0^{c,-}\}_{c=1}^N$, set $\mathcal{C}_0 = \emptyset$
Define $\{\tau_c\}_{c=1}^N$, dimensions $\{n_c\}_{c=1}^N$, and $j_{\max}$
for $k = 1, ..., K$ **do**
  // Receive new measurements
  Receive: $t_k, \boldsymbol{y}_k, \mathcal{C}_k$
  // Time prediction
  $h_k \leftarrow t_k - t_{k-1}$
  Determine $\boldsymbol{m}_k^-, \boldsymbol{P}_k^-$ from (8a)
  Determine $(\nu_k^c)^-, (\boldsymbol{V}_k^c)^-$ from (11)
  // Measurement update for sampled channels
  $\boldsymbol{H}_k \leftarrow [\{\boldsymbol{H}_k^c\}_{c\in\mathcal{C}_k}]$
  $\{\boldsymbol{m}_k^{(0)}, \boldsymbol{P}_k^{(0)}\} \leftarrow \{\boldsymbol{m}_k^-, \boldsymbol{P}_k^-\}$
  $\{\nu_k^c, \boldsymbol{V}_k^{c,(0)}\} \leftarrow \{(\nu_k^c)^- + 1, (\boldsymbol{V}_k^c)^-\}$ for all $c \in \mathcal{C}_k$
  $j \leftarrow 0$
  while not converged and $j < j_{\max}$ **do**
    $\{\bar{\boldsymbol{m}}_k, \bar{\boldsymbol{P}}_k\} \leftarrow \{\boldsymbol{m}_k^-, \boldsymbol{P}_k^-\}$
    for $c \in \mathcal{C}_k$ **do**
      $\hat{\boldsymbol{\Sigma}}_k^{c,(j+1)} \leftarrow \frac{1}{\nu_k^c - n_{y^c} - 1} \boldsymbol{V}_k^{c,(j)}$
      $\boldsymbol{S}_k^{(j+1)} \leftarrow \boldsymbol{H}_k^c \bar{\boldsymbol{P}}_k (\boldsymbol{H}_k^c)^\top + \hat{\boldsymbol{\Sigma}}_k^{c,(j+1)}$
      $\boldsymbol{K}_k^{(j+1)} \leftarrow \bar{\boldsymbol{P}}_k (\boldsymbol{H}_k^c)^\top (\boldsymbol{S}_k^{(j+1)})^{-1}$
      $\bar{\boldsymbol{m}}_k \leftarrow \bar{\boldsymbol{m}}_k + \boldsymbol{K}_k^{(j+1)} (\boldsymbol{y}_k^c - \boldsymbol{H}_k^c \bar{\boldsymbol{m}}_k)$
      $\bar{\boldsymbol{P}}_k \leftarrow \bar{\boldsymbol{P}}_k - \boldsymbol{K}_k^{(j+1)} \boldsymbol{S}_k^{(j+1)} (\boldsymbol{K}_k^{(j+1)})^\top$
      $\boldsymbol{V}_k^{c,(j+1)} \leftarrow (\boldsymbol{V}_k^c)^- + (\boldsymbol{y}_k^c - \boldsymbol{H}_k^c \boldsymbol{m}_k^{(j)})(\star)^\top + \boldsymbol{H}_k^c \boldsymbol{P}_k^{(j)} (\boldsymbol{H}_k^c)^\top$
    end for
    $\{\boldsymbol{m}_k^{(j+1)}, \boldsymbol{P}_k^{(j+1)}\} \leftarrow \{\bar{\boldsymbol{m}}_k, \bar{\boldsymbol{P}}_k\}$
    $j \leftarrow j + 1$
  end while
  // Update sufficient statistics
  $\{\boldsymbol{m}_k, \boldsymbol{P}_k\} \leftarrow \{\boldsymbol{m}_k^{(j)}, \boldsymbol{P}_k^{(j)}\}$
  $\boldsymbol{V}_k^c \leftarrow \boldsymbol{V}_k^{c,(j)}$ for all $c \in \mathcal{C}_k$
end for

measurement will depend on how well a given frame can be mapped onto the current keyframes stored in a dynamically changing map representation. Exactly how the uncertainty of the map representation relates to the noise $\boldsymbol{\Sigma}_k^c$ on this channel is unclear, but it will be time varying.

A. Modeling

To demonstrate the proposed method, we consider a multi-AMR estimation problem and implement the proposed method in C++, launching each AMR as a node in the robot operating system (ROS) and simulating its dynamics in Gazebo using predefined `TurtleBot3` models with nominal parameters. As we have access to odometry data, we can convert it through the model parameters to a velocity in the body frame, v, and a rotation rate ω. We relate this to the position $\boldsymbol{p}_k = [p_k^X, p_k^Y]$ and orientation θ_k of the AMR,

$$\boldsymbol{x}_{k+1} = \underbrace{\begin{bmatrix} \boldsymbol{p}_k + h_k \boldsymbol{R}(\theta_k)[v_k \ 0]^\top \\ \theta_k + h_k \omega_k \end{bmatrix}}_{\boldsymbol{f}(\boldsymbol{x}_k, \boldsymbol{u}_k, h_k)} + \boldsymbol{w}_k, \boldsymbol{w}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}(h_k)), \tag{17}$$

where $\boldsymbol{Q}(h_k) = h_k \text{blkdiag}(\sigma_v^2, \sigma_v^2, \sigma_w^2)$, $\boldsymbol{x}_k = [\boldsymbol{p}_k, \theta_k]$, and control signals $\boldsymbol{u}_k = [v_k, \omega_k]$. In addition to wheel odometry, we consider $N = 7$ measurement channels:

- $c = 1$: Direct position measurements;
- $c = 2$: Direct configuration measurements;
- $c > 2$: Landmark distance measurements.

To model these measurements in the context of AMRs, let

$$p(\boldsymbol{y}_k^c | \boldsymbol{x}_k) = \mathcal{N}(\boldsymbol{y}_k^c | \boldsymbol{h}^c(\boldsymbol{x}_k), \boldsymbol{\Sigma}_k^c) \qquad \forall c \in \mathbb{N}_{[1,N]}, \tag{18}$$

with

$$\boldsymbol{h}^1(\boldsymbol{x}_k) = \boldsymbol{p}_k \qquad \in \mathbb{R}^2 \quad (n_1 = 2), \tag{19a}$$

$$\boldsymbol{h}^2(\boldsymbol{x}_k) = \begin{bmatrix} \boldsymbol{p}_k \\ \cos(\theta_k/2) \end{bmatrix} \qquad \in \mathbb{R}^3 \quad (n_2 = 3), \tag{19b}$$

$$\boldsymbol{h}^c(\boldsymbol{x}_k) = \|\boldsymbol{p}_k - \boldsymbol{p}^c\|_2 \qquad \in \mathbb{R} \quad (n_c = 1) \ \ \forall c > 2, \tag{19c}$$

and where the true noise covariance matrices $\{\boldsymbol{\Sigma}_k^c\}_{c=1}^N$ are unknown for all $c \in \mathbb{N}_{[1,N]}$ and time-varying. The reason for the $\cos(\cdot)$ in (19b) is that we sample the orientation estimate as a quaternion, and as the AMR only rotates about the vertical axis, the real part of the quaternion is modeled by the last dimension in (19b) (see, e.g., [23]).

B. Results for Ideal Motion Models

We start with an idealized example, where we generate synthetic data by using the estimation model (17),(18). Here, we distinguish between the different measurement channels and denote the time at the kth time step in the cth channel by t_k^c. For the different channels, $t_k^c - t_{k-1}^c \sim \text{Poisson}(\lambda^c)$ and we sample a number of time steps K_c for each channel, such that $t_{K_c}^c - t_0^c > T$ for all $c \in \mathbb{N}_{[1,N]}$. We realize a sequence of time steps $\{t_k^0\}_{k=0}^{K_0}$ with a rate λ_0 at which the inputs are sampled. The time steps of the simulation are $\{t_k\} = \bigcup_{c=0}^N \bigcup_{k=0}^{K_c} t_k^c$, and we let $K = |\{t_k\}|$. This implicitly defines the channel sets $\mathcal{C}_k$, where some are empty.

The measurements are realized with time-varying noise covariance matrices $\{\boldsymbol{\Sigma}_{0:k}^{c,\star}\}_{c=1}^N$ to generate a synthetic data set. The measurement data is passed to the ROS node, and the output is analyzed externally in Matlab given the known ground truth, and contains two segments of outliers in the channels $c = 1$ and $c = 2$, respectively, associated with greater measurement noise variance, along with a random noise levels in each of the distance measurements (see Fig. 1). Also, the distance measurements are only available when the landmarks are in close proximity with the AMRs.

Fig. 2 provides a qualitative simulation result using Algorithm 1, indicating that the noise covariance associated with the measurement channels are estimated correctly, even when the measurement data arrives asynchronously. Fig. 3 shows the corresponding results from a Monte-Carlo study with 1000 simulations, with the proposed method and a standard extended KF (EKF) using nominal measurement covariances. The results indicate a significant impact on the estimation performance, measured as root mean-square error (RMSE).

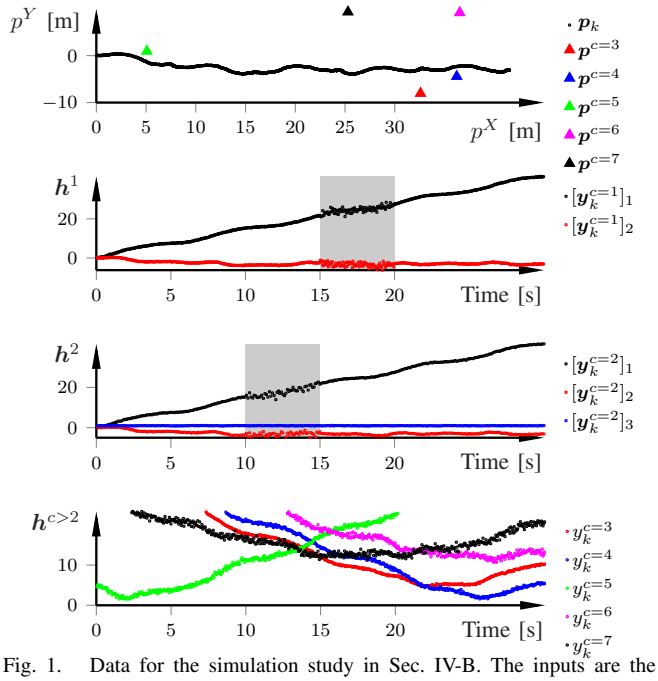

Fig. 1. Data for the simulation study in Sec. IV-B. The inputs are the velocities and rotation rates of the AMR. The measurements arrive at different rates, with two disjoint time intervals of outliers affecting the position (second subplot) and configuration measurements (third subplot), respectively, and different noise levels of the distance measurements (fourth subplot). The shaded areas indicate instances of greater measurement noise, intentionally inflated to study the noise adaptation.

C. Results with Gazebo Models in ROS

In this study, we make the implementation more realistic. In particular, we no longer sample the AMR movement from the prediction model, but use the `Gazebo` simulator with the open-source `TurtleBot3` models. In addition, we now consider a setting with multiple AMRs tasked to navigate in a coordinated manner along a slowly moving reference trajectory, with the super-index $(\cdot)^{(i)}$ indicating the ith AMR, when necessary. The AMRs are controlled using a simple nonlinear proportional controller, with a feedback of the estimates. Furthermore, the relative distance measurements are now evaluated with respect to each AMR; that is, the first AMR samples distance measurements with respect to every other AMR. The related measurement model is identical to (19c), but with the landmark position of channel c in the ith turtlebot defined as the position of the cth turtlebot; that is, $\boldsymbol{p}_k^{c,(i)} = \boldsymbol{p}_k^{(c)}$ when sampling the measurement data, and $\boldsymbol{p}_k^{c,(i)} = \hat{\boldsymbol{p}}_k^{(c)}$ when inferring the state estimate. Here, we ensure that $\mathcal{C}_k^{(i)} \cap i = \emptyset$ for all k, and all agents $i = 1, ..., 4$. This implies that the AMRs communicate their position estimates to one another, and that each AMR has $N = 5$ measurement channels.

Fig. 4 compares the result of using a variable-rate EKF with fixed noise covariances and using Algorithm 1. Table I summarizes the results as time-averaged absolute errors.

V. Conclusion

We developed an asynchronous VB-KF, which can handle measurements from different sources arriving at different rates. When such measurements are independent, our method processes them sequentially, leading to a faster update as

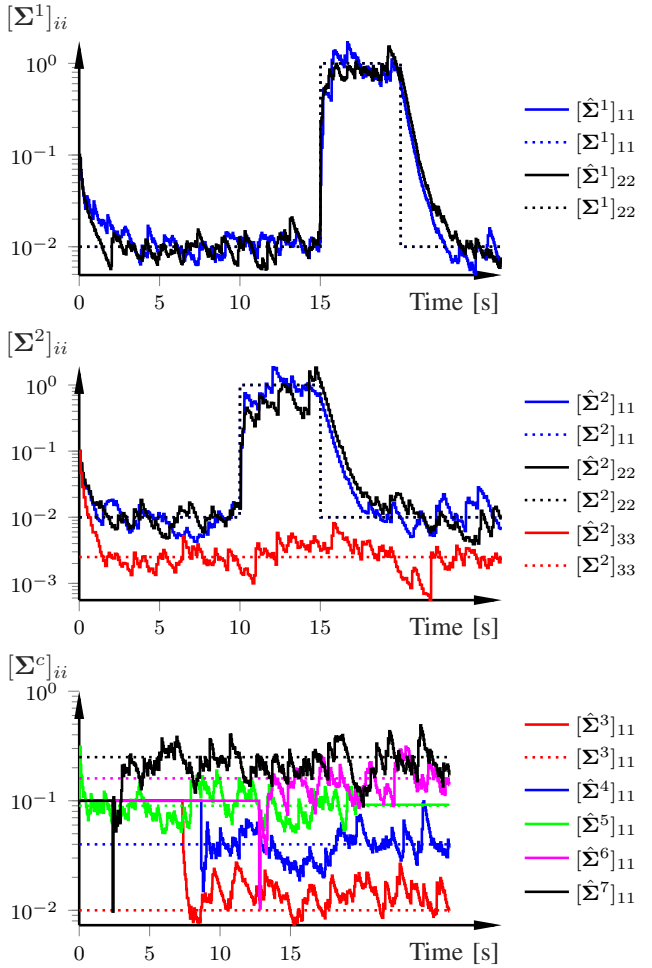

Fig. 2. Diagonal elements of the true noise covariance matrices (dotted) and their respective estimates over time, with the position measurement noise ($c = 1$, top subplot), the configuration measurement noise ($c = 2$, middle subplot), and the distance measurement noise ($c > 2$, bottom subplot).

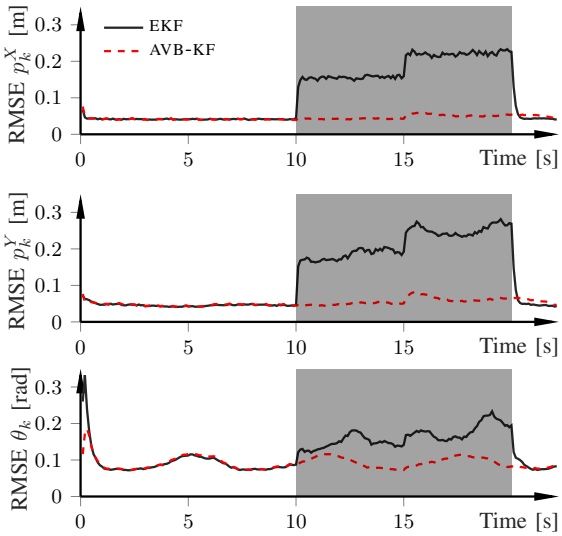

Fig. 3. RMSE from 1000 Monte Carlo simulations of the p^X-position, p^Y-position, and heading-angle estimates. The gray shaded area indicates instances of large-noise measurements, something which the proposed method handles without severely affecting performance.

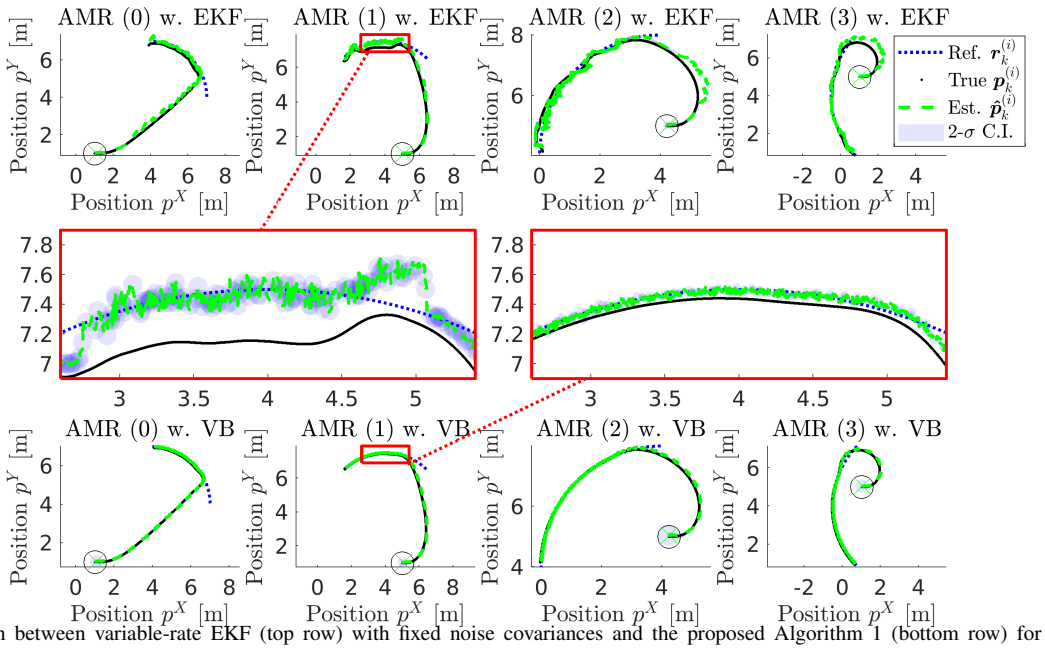

Fig. 4. Comparison between variable-rate EKF (top row) with fixed noise covariances and the proposed Algorithm 1 (bottom row) for a platooning example, involving four AMRs with relative distance measurements similar to (19c). The simulation is done in Gazebo, with the reference trajectory in red dash-dotted, true trajectory in green dashed, estimated trajectory in black dotted, and the blue circles indicate the 2-σ confidence intervals. The middle row shows two zoom-ins of the second AMR for the EKF and Algorithm 1, respectively.

TABLE I

TIME-AVERAGED ABSOLUTE ERROR (TAE) OF THE POSITION ESTIMATES OF EACH AGENT IN THE GAZEBO EXAMPLE IN SEC. IV-C.

Error [m]	Filter	AMR (1)	AMR (2)	AMR (3)	AMR (4)
TAE(p^X)	EKF	0.1275	0.0669	0.1364	0.0486
TAE(p^X)	Alg. 1	0.0573	0.0519	0.0490	0.0477
TAE(p^Y)	EKF	0.1201	0.2008	0.0820	0.1114
TAE(p^Y)	Alg. 1	0.0460	0.0509	0.0511	0.0486

smaller matrix dimensions are handled in serial, as opposed to one large matrix update. We validated the method in using ROS on a Turtlebot mobile robotic setup, also including multiple Turtlebots performing platooning applications. Our results show that the theory of the method is valid also for nonperfect estimation models.

REFERENCES

[1] T. S. Jaakkola, "Variational methods for inference and estimation in graphical models," Ph.D. dissertation, Massachusetts Institute of Technology, 1997.

[2] M. Beal, "Variational algorithms for approximate Bayesian inference," Ph.D. dissertation, University of London, 2003.

[3] 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.

[4] S. Särkkä and J. Hartikainen, "Non-linear noise adaptive Kalman filtering via variational Bayes," in *IEEE Int. Workshop Machine Learning for Signal Processing*, Southampton, UK, Sep. 2013.

[5] F. Gustafsson, *Statistical Sensor Fusion*. Lund, Sweden: Utbildningshuset/Studentlitteratur, 2010.

[6] R. Mehra, "On the identification of variances and adaptive Kalman filtering," *IEEE Trans. Autom. Control*, vol. 15, no. 2, pp. 175–184, 1970.

[7] R. Piche, S. Särkkä, and J. Hartikainen, "Recursive outlier-robust filtering and smoothing for nonlinear systems using the multivariate student-t distribution," in *IEEE Int. Workshop Machine Learning for Signal Processing*, Santander, Spain, Sep. 2012.

[8] G. Agamennoni, J. I. Nieto, and E. M. Nebot, "Approximate inference in state-space models with heavy-tailed noise," *IEEE Trans. Signal Process.*, vol. 60, no. 10, 2012.

[9] G. Storvik, "Particle filters for state-space models with the presence of unknown static parameters," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 281–289, 2002.

[10] P. M. Djurić and J. Miguez, "Sequential particle filtering in the presence of additive Gaussian noise with unknown parameters," in *Int. Conf. Acoustics, Speech, Signal Process.*, Orlando, FL, May 2002.

[11] E. Özkan, V. vSmídl, S. Saha, C. Lundquist, and F. Gustafsson, "Marginalized adaptive particle filtering for nonlinear models with unknown time-varying noise parameters," *Automatica*, vol. 49, no. 6, pp. 1566–1575, 2013.

[12] K. Berntorp and S. Di Cairano, "Tire-stiffness and vehicle-state estimation based on noise-adaptive particle filtering," *IEEE Trans. Control Syst. Technol.*, vol. 27, no. 3, pp. 1100–1114, 2018.

[13] K. Berntorp, "Online Bayesian inference and learning of Gaussian-process state-space models," *Automatica*, vol. 129, p. 109613, 2021.

[14] A. Kullberg, I. Skog, and G. Hendeby, "Online joint state inference and learning of partially unknown state-space models," *IEEE Trans. Signal Process.*, vol. 69, pp. 4149–4161, 2021.

[15] S. Särkkä and J. Sarmavuori, "Gaussian filtering and smoothing for continuous-discrete dynamic systems," *Signal Processing*, vol. 93, no. 2, pp. 500–510, 2013.

[16] S. Särkkä and T. Sottinen, "Application of Girsanov theorem to particle filtering of discretely observed continuous-time non-linear systems," *Bayesian Analysis*, vol. 3, no. 3, pp. 555–584, 2008.

[17] K. Zhang, X. R. Li, and Y. Zhu, "Optimal update with out-of-sequence measurements," *IEEE Trans. Signal Process.*, vol. 53, no. 6, pp. 1992–2004, 2005.

[18] S. Zhang and Y. Bar-Shalom, "Optimal update with multiple out-of-sequence measurements with arbitrary arriving order," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 48, no. 4, pp. 3116–3132, 2012.

[19] U. Orguner and F. Gustafsson, "Storage efficient particle filters for the out of sequence measurement problem," in *Int. Conf. Information Fusion*, Cologne, Germany, June 2008.

[20] K. Berntorp, A. Robertsson, and K.-E. Årzén, "Rao-Blackwellized particle filters with out-of-sequence measurement processing," *IEEE Trans. Signal Process.*, vol. 62, no. 24, pp. 6454–6467, 2014.

[21] K. P. Murphy, "Conjugate Bayesian analysis of the Gaussian distribution," UBC, Tech. Rep., 2007.

[22] K. J. Åström and B. Wittenmark, *Computer-Controlled Systems*. Dover, 2011.

[23] J. Sola, "Quaternion kinematics for the error-state Kalman filter," *arXiv preprint arXiv:1711.02508*, 2017.