

Nonlinear Estimation With State-Dependent Gaussian Observation Noise

— [Source link](#) 

Daive Spinello, Daniel J. Stilwell

Institutions: University of Ottawa

Published on: 02 Feb 2010 - IEEE Transactions on Automatic Control (IEEE)

Topics: Extended Kalman filter, Invariant extended Kalman filter, Gaussian noise, Nonlinear filter and Ensemble Kalman filter

Related papers:

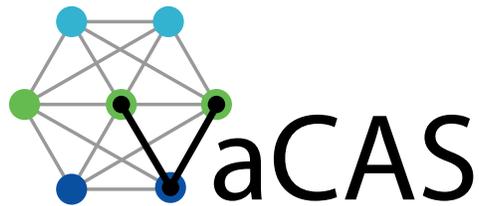
- [A New Approach to Linear Filtering and Prediction Problems](#)
- [The iterated Kalman filter update as a Gauss-Newton method](#)
- [Kalman filtering with intermittent observations](#)
- [Generalized Multiple-Model Adaptive Estimation Using an Autocorrelation Approach](#)
- [An information theoretic approach to observer path design for bearings-only tracking](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/nonlinear-estimation-with-state-dependent-gaussian-2gsf36at02>

NONLINEAR ESTIMATION WITH STATE-DEPENDENT GAUSSIAN OBSERVATION NOISE

D. Spinello & D. J. Stilwell



Virginia Center for Autonomous Systems
Virginia Polytechnic Institute & State University
Blacksburg, VA 24060
www.unmanned.vt.edu

March 8, 2010

Technical Report No. VaCAS-2008-02
Copyright © 2008

Summary

We consider the problem of estimating the state of a system when measurement noise is a function of the system's state. We propose generalizations of the iterated extended Kalman filter and of the extended Kalman filter that can be utilized when the state estimate distribution is approximately Gaussian. The state estimate is computed by an iterative root-searching method that maximize a maximum likelihood function. For sensor network applications, we also address distributed implementations involving multiple sensors.

Contents

Nomenclature	1
1 Introduction	2
2 Problem description	3
2.1 State transition	3
2.2 Observation model	3
2.3 Example system	4
2.3.1 State transition	4
2.3.2 Bearing-only sensors	5
3 Iterated extended Kalman filter with state-dependent observation noise	6
3.1 Bayesian paradigm	6
3.2 State prediction	7
3.3 State update	8
4 Summary of the iterated Kalman filter for state-dependent observation noise	11
4.1 General case	11
4.2 Sensors with scalar output	12
5 Estimation in a decentralized communication network	14
5.1 Sensors network	14
5.2 Bayesian implementation	14
5.3 State prediction	15
5.4 State update	15
5.4.1 Communication of raw sensor data	16
5.4.2 Communication of likelihoods	17
6 Conclusions	19
References	21

Nomenclature

The quantities labeled with a subscript i are referred to the i^{th} sensor. In Section 3, whenever the subscript is dropped, the corresponding function have the same meaning as below without being referred to a specific sensor.

k	integer labeling a time instant
\mathbb{R}	set of real numbers
\mathbf{x}	process state vector
$\mathbf{f}(\cdot)$	process state transition function
\mathbf{Q}	process noise covariance matrix
n	process state dimension
N	number of sensors
\mathbf{z}_i	observation vector
p	observation space dimension
$\mathbf{h}_i(\cdot)$	measurement function
$h_i(\cdot)$	measurement function for $p = 1$
$\Sigma_i(\cdot)$	measurement noise covariance matrix
$\sigma_i(\cdot)$	measurement noise variance for $p = 1$
$p(\cdot, \cdot)$	joint probability density function
$p(\cdot \cdot)$	conditional probability density function
$\mathcal{N}(\cdot, \cdot)$	normal multivariate probability density function
$\mathbf{E}[\cdot]$	expectation operator
$\ell_i(\cdot)$	negative log-likelihood function
$\hat{\mathbf{x}}_i[k l]$	process state estimator at time k given observations up to time $l \leq k$
$\tilde{\mathbf{x}}_i[k l] := \mathbf{x}[k] - \hat{\mathbf{x}}_i[k l]$	estimator error at time k given observations up to time $l \leq k$
$\mathbf{P}_i[k l]$	estimator error covariance at time k given observations up to time $l \leq k$
$\mathcal{I}_i[k] \subseteq \{1, 2, \dots, N\}$	set of integers labeling the sensors communicating with sensor i at time k
tr	trace of a square matrix
\det	determinant of a square matrix
$\nabla_{\mathbf{x}}(\cdot)$	gradient with respect to \mathbf{x}
$(\cdot)^{\top}$	transposition
sym^+	space of symmetric positive-definite matrices

1 Introduction

Advances in embedded computing and wireless communication are facilitating the use of wireless sensor networks for a variety of new applications, including target localization and tracking with bearing-only sensors [12, 25, 39] and with range-only sensors [44]; optimal sensor placement and motion control strategies for target tracking [7, 8, 28, 41]; formation and coverage control [3, 9, 13, 14]; and environmental tracking and monitoring [33, 37, 38].

In this report, we focus on the problem of distributed estimation of the state $\mathbf{x}[k] \in \mathbb{R}^n$ of a dynamical process through a set of noisy measurements $\mathbf{z}_i[k] \in \mathbb{R}^p$, $i = 1, \dots, N$, taken by N sensors at discrete time instants labeled by the integer k . We assume $p \leq n$. The system state evolution and the measurements are modeled as stochastic processes with additive Gaussian noise terms. The measurement noise covariance is a given function of the process state \mathbf{x} .

In works dedicated to sensors motion control strategies for target localization and tracking, the estimation problem is commonly solved by assuming the observation noise to be independent on the process state, see for example [7, 12, 28, 41]. However, as pointed out in [25], this assumption is not realistic for some applications such as those in which bearings-only sensors are employed. We are interested in developing estimators that generalize the extended Kalman filter [1] to the case in which the observation noise is state dependent. By using a maximum likelihood approach coupled with the nonlinear root-searching Newton-Raphson algorithm, we derive a recursive algorithm to estimate the state of a stochastic process from measurements with known state dependent observation noise. As in the iterated extended Kalman filter, only the first two moments of the posterior probability distribution are propagated. Our work is motivated from the study of a class of cooperative motion-control problems in which a group of mobile sensors attempt to configure themselves spatially so that the noise associated to their measurements is minimized, and therefore the feature's estimate is the best achievable given the sensor model [7, 28]. In particular, for a uniform linear array of bearing only sensors, the measurement noise depends on the relative distance to the target and on the orientation of the array with respect to the incident source signal [16].

For the case in which the observation noise is state-independent, different filters have been proposed to address the nonlinear estimation problem in the Bayesian framework: the extended Kalman filter [1, 27, 29]; the extended information filter [30]; the iterated extended Kalman filter [1, 29]; the linear regression Kalman filter [22], which comprises the central difference filter [34], the first-order divided difference filter [31], and the unscented Kalman filter [20]; the sigma point Kalman filter [40], the iterated sigma point Kalman filter [36], and the iterated unscented Kalman filter [43]. For a comparison between these different nonlinear filtering techniques see [23]. A class of nonlinear filters based on Gaussian distributions with multiple moments propagation is proposed in [19].

In [2] it is shown that the iterated extended Kalman filter update is an application of the Gauss-Newton method [5] for approximating a maximum likelihood estimate. The Gauss-Newton method is an approximation of the Newton-Raphson method [5] that applies to

minimum least square problems by discarding second derivatives in the iterates. The update method of the iterated extended Kalman filter reduces to that of the extended Kalman filter for a single iteration, and both reduce to the ordinary Kalman filter when the observation equation is affine in the state to be estimated. Therefore, the extended Kalman filter is inherently suboptimal in the fact that it propagates only the first two moments of the estimator probability density function, and in the fact that the root-searching method used to find the maximum likelihood estimate is not iterated to convergence. For bearings-only tracking, an estimation criterion based on gradient-search iterative methods has been proposed in [18]. The convergence of iterative methods for bearing-only tracking has been addressed in [21].

The rest of the report is organized as follows. In Section 2 we briefly describe the models for the process evolution and for the observations. In Section 3 we derive the iterated extended Kalman filter and the extended Kalman filter updates for a single sensor with state dependent observation noise, and specialize the equations for the case in which the sensor's output is a scalar measurement. In Section 4 we introduce a cooperative scenario in which multiple sensors share information, and we show how the individual estimates are accordingly modified. Section 6 is left for conclusions.

2 Problem description

2.1 State transition

The evolution of the state is described by a nonlinear discrete-time stochastic state transition equation

$$\mathbf{x}[k+1] = \mathbf{f}(\mathbf{x}[k]) + \boldsymbol{\nu}[k] \quad (1)$$

where $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is the possibly nonlinear state transition function, and $\boldsymbol{\nu}[k] \in \mathbb{R}^n$ is an additive process noise. We assume that $\boldsymbol{\nu}$ is a sequence of independent random variables with normal probability distribution $\mathcal{N}(\mathbf{0}, \mathbf{Q})$, therefore satisfying the relations

$$\mathbb{E}[\boldsymbol{\nu}[k]] = \mathbf{0} \quad \forall k \quad (2a)$$

$$\mathbb{E}[\boldsymbol{\nu}[k]\boldsymbol{\nu}^\top[l]] = \mathbf{Q}[k]\delta_{kl} \quad \forall k, l \quad (2b)$$

where $\mathbb{E}[\cdot]$ is the expectation operator, \mathbf{Q} is the target's noise covariance matrix, δ_{kl} is the Kronecker delta, and \top denotes transposition.

2.2 Observation model

We assume that there are N sensors that each generate measurements of the system (1) at discrete time instants. The observation made by the i^{th} sensor at time k is

$$\mathbf{z}_i[k] = \mathbf{h}_i(\mathbf{x}[k]) + \mathbf{v}_i(\mathbf{x}[k]) \quad (3)$$

where $\mathbf{z}_i[k] \in \mathbb{R}^p$, with $p \leq n$, $\mathbf{h}_i : \mathbb{R}^n \rightarrow \mathbb{R}^p$ is the observation function of the model, and the noise $\mathbf{v}_i : \mathbb{R}^n \rightarrow \mathbb{R}^p$ is a sequence of independent random variables with normal probability

distribution $\mathcal{N}(\mathbf{0}, \Sigma_i(\mathbf{x}))$, where $\Sigma_i \in \text{sym}^+ \mathbb{R}^{p \times p}$ in the observation noise covariance, and sym^+ denotes the space of symmetric positive-definite matrices. We emphasize that Σ_i depends on the state \mathbf{x} . It follows that the measurements $\mathbf{z}_i[k]$ can be treated as realizations of multivariate normal distributions described by the conditional moments

$$\mathbb{E}[\mathbf{z}_i[k] \mid \mathbf{x}[k]] = \mathbf{h}_i(\mathbf{x}[k]) \quad (4a)$$

$$\mathbb{E}\left[(\mathbf{z}_i[k] - \mathbf{h}_i(\mathbf{x}[k]))(\mathbf{z}_i[k] - \mathbf{h}_i(\mathbf{x}[k]))^\top \mid \mathbf{x}[k]\right] = \Sigma_i(\mathbf{x}[k]) \quad (4b)$$

We assume that noise terms associated to the same sensor are time uncorrelated, and that noise terms associated to different sensors are mutually independent, that is

$$\mathbb{E}[\mathbf{v}_i(\mathbf{x}[k]) \mathbf{v}_j^\top(\mathbf{x}[l])] = \Sigma_i(\mathbf{x}[k]) \delta_{kl} \delta_{ij} \quad \forall k, l, \quad \forall i, j \quad (5)$$

Additionally, we assume the following cross-correlation independence condition [1]

$$\mathbb{E}[\boldsymbol{\nu}[k] \mathbf{v}_i^\top(\mathbf{x}[l])] = \mathbf{0} \quad \forall i, k, l \quad (6)$$

2.3 Example system

For the purpose of illustration, we briefly describe an example system for which measurement noise is dependent on the state of the system.

2.3.1 State transition

Consider a moving target in plane motion, and let

$$\mathfrak{X} : \mathbb{R}^n \ni \mathbf{x} \mapsto \mathfrak{X}(\mathbf{x}) = \boldsymbol{\varkappa} \in \mathbb{R}^2 \quad (7)$$

be a linear function that maps the target state \mathbf{x} to its position in the plane $\boldsymbol{\varkappa}$, see Figure 1. For nearly constant velocity targets, small fluctuations in the velocity can be modeled as noise, see for example [1]. In this case, the state vector \mathbf{x} includes the position and the velocity, and the trajectory has the following state-space representation

$$\mathbf{x}[k] = \mathbf{f}[k] \mathbf{x}[k-1] + \mathbf{G}[k] \boldsymbol{\nu}[k] \quad (8)$$

where the process noise represents the uncertainty of acceleration. The matrices \mathbf{f} and \mathbf{G} are derived through the kinematics of a point mass with constant acceleration:

$$\mathbf{f} = \begin{pmatrix} 1 & 0 & \gamma & 0 \\ 0 & 1 & 0 & \gamma \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} \gamma^2/2 & 0 \\ 0 & \gamma^2/2 \\ \gamma & 0 \\ 0 & \gamma \end{pmatrix} \quad (9)$$

where γ is the observation time interval. A refined model with constant acceleration included into the state vector has been proposed in [10].

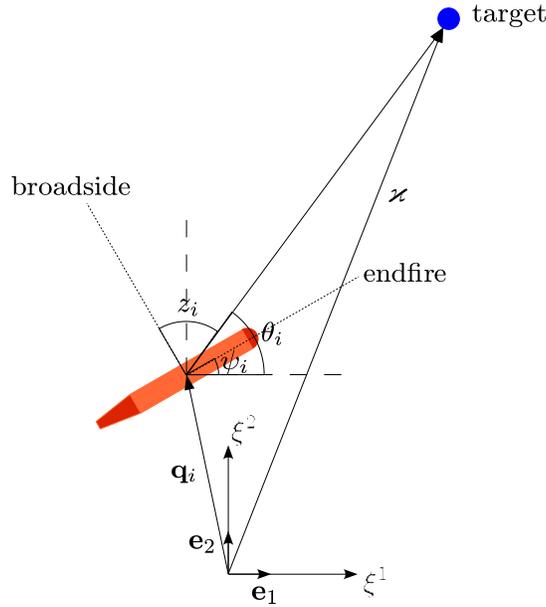


Figure 1: Sketch of the kinematics of a vehicle equipped with a bearing-only sensor.

2.3.2 Bearing-only sensors

We suppose that the state of a moving target is observed by several vehicles (mobile sensors), each equipped with a bearings only sensor. Let $\mathbf{q}_i \in \mathbb{R}^2$ be the position of the center of mass of vehicle i , and $\psi_i \in \mathbb{R}$ the angle formed by the axis of the vehicle with respect to some reference direction, see Figure 1. The vector representing the line of site between the sensor i and the target is $\boldsymbol{\varkappa} - \mathbf{q}_i$. The output of sensor i at time k is the scalar $z_i[k]$ representing the bearing angle with respect to the broadside of the sensor, see Figure 1. In this case, the measurement function h_i , which is the scalar equivalent of \mathbf{h}_i in (3), is given by

$$h_i(\mathbf{q}_i, \psi_i, \mathbf{x}) = \frac{\pi}{2} + \psi_i - \theta_i(\mathbf{q}_i, \mathbf{x}) \quad (10a)$$

$$\theta_i(\mathbf{q}_i, \mathbf{x}) = \arctan\left(\frac{(\boldsymbol{\varkappa} - \mathbf{q}_i)^\top \mathbf{e}_2}{(\boldsymbol{\varkappa} - \mathbf{q}_i)^\top \mathbf{e}_1}\right) \quad (10b)$$

where $\mathbf{e}_1 = (1 \ 0)^\top$ and $\mathbf{e}_2 = (0 \ 1)^\top$ are unit basis vectors for the rectangular Cartesian coordinate system $\{\xi^1, \xi^2\}$, see Figure 1.

The model for the noise variance is given by, see [16, 25, 24]

$$\sigma_i(\mathbf{q}_i, \psi_i, \mathbf{x}) = \alpha \frac{d(\mathbf{q}_i, \psi_i, \mathbf{x})}{\cos^2 h_i(\mathbf{q}_i, \psi_i, \mathbf{x})} \quad (11)$$

where $\sigma_i > 0$ is the scalar equivalent of $\boldsymbol{\Sigma}_i$ in (3), the scalar function d is the inverse of the signal to noise ratio, and α is a constant that depends on parameters of the source signal and the sensor. Expressions for α when the sensor is a uniform linear array are given in

[16]. Note that the variance of the measurement noise in (11) approaches infinity as the bearing angle approaches $\pi/2$, and it is maximum when the bearing angle approaches 0, see Figure 1. For vehicles modeled as point masses, a model for the noise covariance adopted in [7, 8, 41] is $\sigma_i(\mathbf{q}_i, \psi_i, \mathbf{x}) = d(\mathbf{q}_i, \psi_i, \mathbf{x}) = a_2(\|\mathbf{q}_i - \mathbf{x}\| - a_1)^2 + a_0$, in which a_0 , a_1 , and a_2 are constant parameters. This model corresponds to the assumption of the existence of a “sweet spot in sensing”, located at a distance a_1 to the target, where uncertainty in measurement is minimal.

3 Iterated extended Kalman filter with state-dependent observation noise

In this Section we focus on the derivation of the iterated extended Kalman filter and extended Kalman filter prediction and update equations for a single sensor.

3.1 Bayesian paradigm

The approach that is commonly adopted for state estimate of a discrete-time dynamical system can be described as a two-stage recursive process of prediction and update. This is true in particular for the Kalman filter and its extensions to nonlinear systems, the extended Kalman filter and the iterated extended Kalman filter, see for example [29, 30].

Let $\mathbf{z}[k] \in \mathbb{R}^p$ be the observation at time k . Since we restrict our attention to a single sensor, we drop the label i in the observation equation (3). The probabilistic information contained in sensor measurement \mathbf{z} about the unknown state \mathbf{x} is described by the conditional probability distribution $p(\mathbf{z}[k] | \mathbf{x}[k])$, known as the likelihood function.

From Bayes’ rule, see [1, Appendix B.11], we obtain the following representation for the posterior conditional distribution of \mathbf{x} given \mathbf{z}

$$p(\mathbf{x}[k] | \mathbf{z}[k]) = \frac{p(\mathbf{z}[k] | \mathbf{x}[k]) p(\mathbf{x}[k])}{p(\mathbf{z}[k])} \quad (12)$$

where $p(\cdot)$ is the marginal distribution.

In order to reduce the uncertainty, one can consider several measurements taken over time to construct the posterior. We define the collection of observations up to time k

$$\mathbf{Z}[k] := \{\mathbf{z}[k], \mathbf{z}[k-1], \dots, \mathbf{z}[0]\} \quad (13a)$$

In this case, the likelihood function is $p(\mathbf{Z}[k] | \mathbf{x}[k])$, and the posterior is given by

$$p(\mathbf{x}[k] | \mathbf{Z}[k]) = \frac{p(\mathbf{Z}[k] | \mathbf{x}[k]) p(\mathbf{x}[k])}{p(\mathbf{Z}[k])} \quad (14)$$

The posterior can also be computed recursively after each observation $\mathbf{z}[k]$. Let $p(\cdot, \cdot)$ be the joint probability distribution. By applying Bayes rule we obtain

$$\begin{aligned} p(\mathbf{Z}[k], \mathbf{x}[k]) &= p(\mathbf{z}[k], \mathbf{Z}[k-1], \mathbf{x}[k]) \\ &= p(\mathbf{z}[k] | \mathbf{Z}[k-1], \mathbf{x}[k]) p(\mathbf{Z}[k-1], \mathbf{x}[k]) \\ &= p(\mathbf{z}[k] | \mathbf{x}[k]) p(\mathbf{x}[k] | \mathbf{Z}[k-1]) p(\mathbf{Z}[k-1]) \end{aligned} \quad (15)$$

where we used the assumption, intrinsic in the sensor model explained in Section 2.2, that the probability distribution of $\mathbf{z}[k]$ is independent on $\mathbf{Z}[k-1]$ whenever $\mathbf{x}[k]$ is given. From Bayes rule we obtain also

$$p(\mathbf{Z}[k], \mathbf{x}[k]) = p(\mathbf{x}[k] | \mathbf{Z}[k]) p(\mathbf{z}[k] | \mathbf{Z}[k-1]) p(\mathbf{Z}[k-1]) \quad (16)$$

By combining (15) and (16) we obtain the following recursive form of the posterior distribution

$$p(\mathbf{x}[k] | \mathbf{Z}[k]) = \beta[k] p(\mathbf{x}[k] | \mathbf{Z}[k-1]) p(\mathbf{z}[k] | \mathbf{x}[k]) \quad (17)$$

where the proportionality factor is $\beta^{-1}[k] = p(\mathbf{z}[k] | \mathbf{Z}[k-1])$. The probability density function $p(\mathbf{x}[k] | \mathbf{Z}[k-1])$ is associated to a prior estimate of the state $\mathbf{x}[k]$ based on observations up to time $k-1$. The density $p(\mathbf{z}[k] | \mathbf{x}[k])$ can be interpreted as a correction to the prior data based on current observation. The recursive implementation in (17) is often used in estimation theory since it requires limited data storage [27]. The updating scheme in (17) can be interpreted as using new information to correct a previous prediction.

3.2 State prediction

Following [1], we define the state estimate at time k given the observations up to time $l \leq k$, and the related error covariance matrix as

$$\hat{\mathbf{x}}[k | l] := \mathbb{E}[\mathbf{x}[k] | \mathbf{Z}[l]] \quad (18a)$$

$$\mathbf{P}[k | l] := \mathbb{E}\left[(\mathbf{x}[k] - \hat{\mathbf{x}}[k | l])(\mathbf{x}[k] - \hat{\mathbf{x}}[k | l])^\top | \mathbf{Z}[l]\right] \quad (18b)$$

It is assumed that there exists a state estimate $\hat{\mathbf{x}}[k-1 | k-1]$ at time $k-1$ and associated error covariance $\mathbf{P}[k-1 | k-1]$. The objective is to find a prediction $\hat{\mathbf{x}}[k | k-1]$ of the state at time k based on information available up to time $k-1$. First, (1) is expanded in a Taylor series about $\hat{\mathbf{x}}[k-1 | k-1]$. By truncating terms above the first order and applying definitions (18a) and (18b) with expectations conditioned on $\mathbf{Z}[k-1]$, we obtain the state and the error covariance predictions

$$\begin{aligned} \hat{\mathbf{x}}[k | k-1] &= \mathbb{E}[\mathbf{x}[k] | \mathbf{Z}[k-1]] \\ &= \mathbf{f}(\hat{\mathbf{x}}[k-1 | k-1]) \end{aligned} \quad (19a)$$

$$\begin{aligned} \mathbf{P}[k | k-1] &= \mathbb{E}\left[(\mathbf{x}[k] - \hat{\mathbf{x}}[k | k-1])(\mathbf{x}[k] - \hat{\mathbf{x}}[k | k-1])^\top | \mathbf{Z}[k-1]\right] \\ &= \nabla_{\mathbf{x}} \mathbf{f}(\hat{\mathbf{x}}[k-1 | k-1]) \mathbf{P}[k-1 | k-1] \nabla_{\mathbf{x}} \mathbf{f}^\top(\hat{\mathbf{x}}[k-1 | k-1]) + \mathbf{Q}[k] \end{aligned} \quad (19b)$$

where $\nabla_{\mathbf{x}}$ is the gradient with respect to \mathbf{x} (see [1], for example). The predictions (19) are derived with the aid of (2a), (2b), and (6), and the property of the expectation operator, see [35, Section 1.3.2]

$$\mathbf{E} [\mathbf{A}\mathbf{a}\mathbf{b}^T\mathbf{B}^T] = \mathbf{A}\mathbf{E} [\mathbf{a}\mathbf{b}^T] \mathbf{B}^T \quad (20)$$

where \mathbf{a} and \mathbf{b} are random vectors, and \mathbf{A} and \mathbf{B} are matrices. Higher-order predictions can be obtained by retaining more terms in the Taylor series expansion of the state model [1].

We note that all the information about the target up to time $k - 1$ is included in the estimate $\hat{\mathbf{x}}[k - 1 | k - 1]$, and therefore the computation of the predictions requires only the knowledge of the estimates at time $k - 1$ without storage of past observation data.

3.3 State update

After the derivation of the state and error covariance predictions, we return to the likelihood equation (17) to solve the update problem. We generalize the approach proposed in [2] to find an approximation of a maximum likelihood estimate. This approach allows us to derive the filter update equation for the case in which the covariance associated with the measurement is a function of the state to be estimated, see (5).

The state prediction $\hat{\mathbf{x}}[k | k - 1]$ and the measurement $\mathbf{z}[k]$ are treated as realizations of independent random vectors with multivariate normal distributions (see [2])

$$\hat{\mathbf{x}}[k | k - 1] \sim \mathcal{N}(\mathbf{x}[k], \mathbf{P}[k | k - 1]), \quad \mathbf{z}[k] \sim \mathcal{N}(\mathbf{h}(\mathbf{x}[k]), \Sigma(\mathbf{x}[k])) \quad (21)$$

Although the distribution for $\hat{\mathbf{x}}[k | k - 1]$ is not necessarily Gaussian in practice, it is commonly assumed that a Gaussian approximation is appropriate in many practical applications [12, 41, 25, 28, 32, 39, 44]. Therefore the probability density functions in (17) are given by

$$\begin{aligned} p(\mathbf{x}[k] | \mathbf{Z}[k - 1]) &= \sqrt{\frac{(\det \mathbf{P}[k | k - 1])^{-1}}{(2\pi)^n}} \\ &\times \exp\left(-\frac{1}{2}(\mathbf{x}[k] - \hat{\mathbf{x}}[k | k - 1])^T \mathbf{P}^{-1}[k | k - 1] (\mathbf{x}[k] - \hat{\mathbf{x}}[k | k - 1])\right) \end{aligned} \quad (22a)$$

$$\begin{aligned} p(\mathbf{z}[k] | \mathbf{x}[k]) &= \sqrt{\frac{(\det \Sigma(\mathbf{x}[k]))^{-1}}{(2\pi)^p}} \\ &\times \exp\left(-\frac{1}{2}(\mathbf{z}[k] - \mathbf{h}(\mathbf{x}[k]))^T \Sigma^{-1}(\mathbf{x}[k]) (\mathbf{z}[k] - \mathbf{h}(\mathbf{x}[k]))\right) \end{aligned} \quad (22b)$$

where \det is the determinant.

The maximum likelihood estimator $\hat{\mathbf{x}}[k | k]$ associated to the posterior $p(\mathbf{x}[k] | \mathbf{Z}[k])$ in (17) is the vector $\mathbf{x}[k]$ that maximizes the likelihood $p(\mathbf{x}[k] | \mathbf{Z}[k])$ or, equivalently, the vector $\mathbf{x}[k]$ that minimizes its negative logarithm, see for example [27]. Substituting from (22) into (17),

and taking the negative logarithm we obtain

$$\begin{aligned} \ell(\mathbf{x}[k]) &= \frac{1}{2} \left(\ln \det \boldsymbol{\Sigma}(\mathbf{x}[k]) + (\mathbf{z}[k] - \mathbf{h}(\mathbf{x}[k]))^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}[k]) (\mathbf{z}[k] - \mathbf{h}(\mathbf{x}[k])) \right) \\ &\quad + \frac{1}{2} (\mathbf{x}[k] - \hat{\mathbf{x}}[k | k-1])^\top \mathbf{P}^{-1}[k | k-1] (\mathbf{x}[k] - \hat{\mathbf{x}}[k | k-1]) + c \end{aligned} \quad (23)$$

where c is a constant not dependent on $\mathbf{x}[k]$. The state estimate is given by the solution of the problem

$$\hat{\mathbf{x}}[k | k] = \underset{\mathbf{x}}{\operatorname{argmin}} \ell(\mathbf{x}[k]) \quad (24)$$

which can be equivalently stated as a nonlinear unconstrained optimization problem.

Under the hypothesis that ℓ is twice continuous differentiable, the solution to the optimization problem is found through the Newton-Raphson iterative sequence [5]

$$\hat{\mathbf{x}}^{(\iota+1)}[k | k] = \hat{\mathbf{x}}^{(\iota)}[k | k] - [\nabla_{\mathbf{x}} \nabla_{\mathbf{x}} \ell(\hat{\mathbf{x}}^{(\iota)}[k | k])]^{-1} \nabla_{\mathbf{x}} \ell(\hat{\mathbf{x}}^{(\iota)}[k | k]) \quad (25)$$

with initial guess $\hat{\mathbf{x}}^{(0)}[k | k] = \hat{\mathbf{x}}[k | k-1]$, where (ι) refers to the iteration step. For a single step iteration and state-independent noise covariance, the algorithm (25) defines the extended Kalman filter, while for multiple iterations it defines the iterated extended Kalman filter, see [2]. The extended Kalman filter is suboptimal since the single step convergence of the Newton-Raphson iterates is guaranteed only if the function $\ell(\mathbf{x})$ is quadratic in his argument, see [5], which is true for the case that the observation equation is affine. In [2] the Gauss-Newton algorithm has been used in place of the Newton-Raphson algorithm. The Gauss-Newton algorithm applies to minimum least-square problems, and approximates the Hessian by discarding second derivatives of the residuals. This is consistent with the classical extended Kalman filter derivation, in which the observation function is approximated with a Taylor series truncated at the first order. However, for the case we are studying, the function ℓ cannot be expressed as a quadratic form because of the first term on the right hand side of (23). Therefore we derive the update through (25).

Computation of derivatives of the function ℓ is facilitated by the following relationships, see [4] for example. Let \mathbf{A} and \mathbf{B} be matrices, with \mathbf{A} non singular and dependent on a real parameter τ . Then

$$\frac{\partial}{\partial \tau} \det \mathbf{A} = \det \mathbf{A} \operatorname{tr} \left[\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial \tau} \right], \quad \frac{\partial \mathbf{A}^{-1}}{\partial \tau} = -\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial \tau} \mathbf{A}^{-1} \quad (26a)$$

$$\operatorname{tr}[\mathbf{A}\mathbf{B}] = \operatorname{tr}[\mathbf{B}\mathbf{A}] \quad (26b)$$

where tr is the trace operator. We also introduce the following notation

$$\boldsymbol{\zeta}(\mathbf{x}[k]) := \mathbf{z}[k] - \mathbf{h}(\mathbf{x}[k]) \quad (27)$$

By regarding the quadratic terms in (23) as the trace of a 1×1 matrix, and by using (26b), we rewrite the likelihood function $\ell(\mathbf{x})$ in the more convenient form

$$\begin{aligned} \ell(\mathbf{x}[k]) &= \frac{1}{2} \left(\ln \det \boldsymbol{\Sigma}(\mathbf{x}[k]) + \operatorname{tr} \left[\boldsymbol{\Sigma}^{-1}(\mathbf{x}[k]) \boldsymbol{\zeta}(\mathbf{x}[k]) \boldsymbol{\zeta}^\top(\mathbf{x}[k]) \right] \right) \\ &\quad + \frac{1}{2} \operatorname{tr} \left[\mathbf{P}^{-1}[k | k-1] (\mathbf{x}[k] - \hat{\mathbf{x}}[k | k-1]) (\mathbf{x}[k] - \hat{\mathbf{x}}[k | k-1])^\top \right] + c \end{aligned} \quad (28)$$

By using (26a), the gradient of the log-likelihood function is the n -vector whose l^{th} component given by

$$\begin{aligned} \frac{\partial \ell}{\partial x_l}(\mathbf{x}[k]) &= \text{tr} [\mathbf{P}^{-1}[k | k - 1] (\mathbf{x}[k] - \hat{\mathbf{x}}[k | k - 1]) \mathbf{e}_l^\top] \\ &+ \frac{1}{2} \text{tr} \left[\boldsymbol{\Sigma}^{-1}(\mathbf{x}[k]) \frac{\partial \boldsymbol{\Sigma}(\mathbf{x}[k])}{\partial x_l} (\mathbf{I}_p - \boldsymbol{\Sigma}^{-1}(\mathbf{x}[k]) \boldsymbol{\zeta}(\mathbf{x}[k]) \boldsymbol{\zeta}^\top(\mathbf{x}[k])) \right. \\ &\quad \left. - 2 \frac{\partial \mathbf{h}^\top(\mathbf{x}[k])}{\partial x_l} \boldsymbol{\Sigma}^{-1}(\mathbf{x}[k]) \boldsymbol{\zeta}(\mathbf{x}[k]) \right] \end{aligned} \quad (29)$$

where \mathbf{e}_l is the l^{th} vector of the natural basis in \mathbb{R}^n , that is

$$\mathbf{e}_l = (0 \quad \dots \quad 0 \quad 1 \quad 0 \quad \dots \quad 0)^\top \quad (30)$$

l^{th}

and \mathbf{I}_p is the identity matrix in $\mathbb{R}^{p \times p}$. We note that for the case in which the covariance matrix $\boldsymbol{\Sigma}$ is not state-dependent, (29) reduces to the familiar innovation term for the extended Kalman filter. The Hessian term in (25) is the $n \times n$ -matrix with lm entry given by

$$\begin{aligned} \frac{\partial^2 \ell}{\partial x_l \partial x_m} &= \text{tr} [\mathbf{P}^{-1}[k | k - 1] \mathbf{e}_m \mathbf{e}_l] \\ &+ \frac{1}{2} \text{tr} \left[\left(-\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial x_m} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial x_l} + \boldsymbol{\Sigma}^{-1} \frac{\partial^2 \boldsymbol{\Sigma}}{\partial x_l \partial x_m} \right) (\mathbf{I}_p - \boldsymbol{\Sigma}^{-1} \boldsymbol{\zeta} \boldsymbol{\zeta}^\top) \right. \\ &\quad + \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial x_l} \left(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial x_m} \boldsymbol{\Sigma}^{-1} \boldsymbol{\zeta} \boldsymbol{\zeta}^\top + 2 \frac{\partial \mathbf{h}^\top}{\partial x_m} \boldsymbol{\Sigma}^{-1} \boldsymbol{\zeta} \right) \\ &\quad \left. + 2 \frac{\partial \mathbf{h}^\top}{\partial x_l} \boldsymbol{\Sigma}^{-1} \frac{\partial \mathbf{h}}{\partial x_m} + 2 \frac{\partial \mathbf{h}^\top}{\partial x_l} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial x_m} \boldsymbol{\Sigma}^{-1} \boldsymbol{\zeta} - 2 \frac{\partial^2 \mathbf{h}^\top}{\partial x_l \partial x_m} \boldsymbol{\Sigma}^{-1} \boldsymbol{\zeta} \right] \end{aligned} \quad (31)$$

We note that also in this case, by evaluating the expression (31) at $\hat{\mathbf{x}}[k | k - 1]$ and discarding second derivatives and terms that depend on the derivatives of the of the matrix $\boldsymbol{\Sigma}$, we obtain the familiar expression for the extended Kalman filter, see for example [30]. Since the Gauss-Newton method approximates the Newton-Raphson method by neglecting second derivatives in the computation of the Hessian, its application to a log-likelihood with state-independent observation covariance gives the extended Kalman filter update [2].

We rewrite the Hessian in (31) as

$$\nabla_{\mathbf{x}} \nabla_{\mathbf{x}} \ell(\mathbf{x}[k]) = \mathbf{P}^{-1}[k | k - 1] + \mathbf{R}[k] \quad (32a)$$

$$\begin{aligned} [\mathbf{R}[k]]_{lm} &= \text{tr} \left[\frac{\partial \mathbf{h}^\top}{\partial x_l} \boldsymbol{\Sigma}^{-1} \frac{\partial \mathbf{h}}{\partial x_m} + \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial x_l} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial x_m} \left(\boldsymbol{\Sigma}^{-1} \boldsymbol{\zeta} \boldsymbol{\zeta}^\top - \frac{1}{2} \mathbf{I}_p \right) \right. \\ &\quad + \left(\frac{\partial \mathbf{h}^\top}{\partial x_l} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial x_m} + \frac{\partial \mathbf{h}^\top}{\partial x_m} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial x_l} \right) \boldsymbol{\Sigma}^{-1} \boldsymbol{\zeta} - \frac{\partial^2 \mathbf{h}^\top}{\partial x_l \partial x_m} \boldsymbol{\Sigma}^{-1} \boldsymbol{\zeta} \\ &\quad \left. + \frac{1}{2} \boldsymbol{\Sigma}^{-1} \frac{\partial^2 \boldsymbol{\Sigma}}{\partial x_l \partial x_m} (\mathbf{I}_p - \boldsymbol{\Sigma}^{-1} \boldsymbol{\zeta} \boldsymbol{\zeta}^\top) \right]_{\mathbf{x}[k]} \end{aligned} \quad (32b)$$

The Fisher information matrix is defined as the expected value of the square of the score, that is

$$\mathbf{F} = \mathbb{E} [\nabla_{\mathbf{x}}^T \ell \nabla_{\mathbf{x}} \ell] \quad (33)$$

However, for normal multivariate probability density functions¹ the following identity holds

$$\mathbb{E} [\nabla_{\mathbf{x}}^T \ell \nabla_{\mathbf{x}} \ell] = \mathbb{E} [\nabla_{\mathbf{x}} \nabla_{\mathbf{x}} \ell] \quad (34)$$

Note that the sign on the right hand side is reversed with respect to the usual definition. This is due to the fact that ℓ is the negative of the log-likelihood function, whereas the Fisher information matrix is defined in terms of the log-likelihood function. We use the fact that $\boldsymbol{\zeta}$ is a multivariate normally distributed random vector with covariance $\boldsymbol{\Sigma}$, and compute the Fisher information matrix from (32):

$$\mathbf{F} = \mathbb{E} [\nabla_{\mathbf{x}} \nabla_{\mathbf{x}} \ell] = \mathbf{P}^{-1}[k | k - 1] + \bar{\mathbf{R}} \quad (35a)$$

$$[\bar{\mathbf{R}}]_{lm} = \mathbb{E} [\mathbf{R}]_{lm} = \frac{\partial \mathbf{h}^T}{\partial x_l} \boldsymbol{\Sigma}^{-1} \frac{\partial \mathbf{h}}{\partial x_m} + \frac{1}{2} \text{tr} \left[\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial x_l} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial x_m} \right] \quad (35b)$$

When the state and the observation equations are affine, the Newton-Raphson algorithm converges in a single step, and the covariance in (35) reduces to the Kalman filter error covariance derived in [42] for affine observations with state dependent noise. From the approximation that $\hat{\mathbf{x}}^{(l)}$ is normally distributed, the posterior error covariance equals the inverse of the Fisher information matrix [30, Section 2.3]. Therefore

$$\begin{aligned} \mathbf{P}^{(l)}[k | k] &= [\mathbf{F}^{(l)}[k]]^{-1} \\ &= [\mathbf{P}^{-1}[k | k - 1] + \bar{\mathbf{R}}^{(l)}[k]]^{-1} \end{aligned} \quad (36)$$

We note that since $\mathbf{P}[k | k - 1]$ and $\bar{\mathbf{R}}^{(l)}[k]$ are symmetric positive definite, the matrix $\mathbf{P}^{(l)}[k | k]$ is also symmetric positive-definite, see [4].

4 Summary of the iterated Kalman filter for state-dependent observation noise

4.1 General case

From the nonlinear state equation (1) we have the first-order state and error covariance predictions:

$$\hat{\mathbf{x}}[k | k - 1] = \mathbf{f}(\hat{\mathbf{x}}[k - 1 | k - 1]) \quad (37a)$$

$$\mathbf{P}[k | k - 1] = \nabla_{\mathbf{x}} \mathbf{f}(\hat{\mathbf{x}}[k - 1 | k - 1]) \mathbf{P}[k - 1 | k - 1] \nabla_{\mathbf{x}} \mathbf{f}^T(\hat{\mathbf{x}}[k - 1 | k - 1]) + \mathbf{Q}[k] \quad (37b)$$

¹More generally, the relation (34) hold for all probability distributions whose density function satisfy a specific regularity condition, see [35, Proposition 3.1].

From the nonlinear measurement equation (3) we have the following first-order iterated extended Kalman filter updates:

$$\hat{\mathbf{x}}^{(\iota+1)}[k|k] = \hat{\mathbf{x}}^{(\iota)}[k|k] - [\mathbf{P}^{-1}[k|k-1] + \mathbf{R}^{(\iota)}[k]]^{-1} \mathbf{s}^{(\iota)}[k] \quad (38a)$$

$$\begin{aligned} [\mathbf{s}^{(\iota)}[k]]_l &= \mathbf{e}_l^\top \mathbf{P}^{-1}[k|k-1] (\hat{\mathbf{x}}^{(\iota)}[k|k] - \hat{\mathbf{x}}[k|k-1]) \\ &+ \left[-\frac{\partial \mathbf{h}^\top}{\partial x_l} \boldsymbol{\Sigma}^{-1} \boldsymbol{\zeta} + \frac{1}{2} \text{tr} \left[\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial x_l} (\mathbf{I}_p - \boldsymbol{\Sigma}^{-1} \boldsymbol{\zeta} \boldsymbol{\zeta}^\top) \right] \right]_{\hat{\mathbf{x}}^{(\iota)}[k|k]} \end{aligned} \quad (38b)$$

$$\begin{aligned} [\mathbf{R}^{(\iota)}[k]]_{lm} &= \text{tr} \left[\frac{\partial \mathbf{h}^\top}{\partial x_l} \boldsymbol{\Sigma}^{-1} \frac{\partial \mathbf{h}}{\partial x_m} + \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial x_l} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial x_m} \left(\boldsymbol{\Sigma}^{-1} \boldsymbol{\zeta} \boldsymbol{\zeta}^\top - \frac{1}{2} \mathbf{I}_p \right) \right. \\ &+ \left(\frac{\partial \mathbf{h}^\top}{\partial x_l} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial x_m} + \frac{\partial \mathbf{h}^\top}{\partial x_m} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial x_l} \right) \boldsymbol{\Sigma}^{-1} \boldsymbol{\zeta} - \frac{\partial^2 \mathbf{h}^\top}{\partial x_l \partial x_m} \boldsymbol{\Sigma}^{-1} \boldsymbol{\zeta} \\ &\left. + \frac{1}{2} \boldsymbol{\Sigma}^{-1} \frac{\partial^2 \boldsymbol{\Sigma}}{\partial x_l \partial x_m} (\mathbf{I}_p - \boldsymbol{\Sigma}^{-1} \boldsymbol{\zeta} \boldsymbol{\zeta}^\top) \right]_{\hat{\mathbf{x}}^{(\iota)}[k|k]} \end{aligned} \quad (38c)$$

$$\mathbf{P}^{(\iota)}[k|k] = [\mathbf{P}^{-1}[k|k-1] + \bar{\mathbf{R}}^{(\iota)}[k]]^{-1} \quad (38d)$$

$$[\bar{\mathbf{R}}^{(\iota)}[k]]_{lm} = \left[\frac{\partial \mathbf{h}^\top}{\partial x_l} \boldsymbol{\Sigma}^{-1} \frac{\partial \mathbf{h}}{\partial x_m} + \frac{1}{2} \text{tr} \left[\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial x_l} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial x_m} \right] \right]_{\hat{\mathbf{x}}^{(\iota)}[k|k]} \quad (38e)$$

For a single step iteration, (38) evaluated at $\hat{\mathbf{x}}^{(\iota)}[k|k] = \hat{\mathbf{x}}[k|k-1]$ give the extended Kalman filter updates:

$$\hat{\mathbf{x}}[k|k] = \hat{\mathbf{x}}[k|k-1] - [\mathbf{P}^{-1}[k|k-1] + \mathbf{R}^{(0)}[k]]^{-1} \mathbf{s}^{(0)}[k] \quad (39a)$$

$$[\mathbf{s}^{(0)}[k]]_l = \left[-\frac{\partial \mathbf{h}^\top}{\partial x_l} \boldsymbol{\Sigma}^{-1} \boldsymbol{\zeta} + \frac{1}{2} \text{tr} \left[\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial x_l} (\mathbf{I}_p - \boldsymbol{\Sigma}^{-1} \boldsymbol{\zeta} \boldsymbol{\zeta}^\top) \right] \right]_{\hat{\mathbf{x}}[k|k-1]} \quad (39b)$$

$$\begin{aligned} [\mathbf{R}^{(0)}[k]]_{lm} &= \text{tr} \left[\frac{\partial \mathbf{h}^\top}{\partial x_l} \boldsymbol{\Sigma}^{-1} \frac{\partial \mathbf{h}}{\partial x_m} + \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial x_l} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial x_m} \left(\boldsymbol{\Sigma}^{-1} \boldsymbol{\zeta} \boldsymbol{\zeta}^\top - \frac{1}{2} \mathbf{I}_p \right) \right. \\ &+ \left(\frac{\partial \mathbf{h}^\top}{\partial x_l} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial x_m} + \frac{\partial \mathbf{h}^\top}{\partial x_m} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial x_l} \right) \boldsymbol{\Sigma}^{-1} \boldsymbol{\zeta} - \frac{\partial^2 \mathbf{h}^\top}{\partial x_l \partial x_m} \boldsymbol{\Sigma}^{-1} \boldsymbol{\zeta} \\ &\left. + \frac{1}{2} \boldsymbol{\Sigma}^{-1} \frac{\partial^2 \boldsymbol{\Sigma}}{\partial x_l \partial x_m} (\mathbf{I}_p - \boldsymbol{\Sigma}^{-1} \boldsymbol{\zeta} \boldsymbol{\zeta}^\top) \right]_{\hat{\mathbf{x}}[k|k-1]} \end{aligned} \quad (39c)$$

$$\mathbf{P}[k|k] = [\mathbf{P}^{-1}[k|k-1] + \bar{\mathbf{R}}^{(0)}[k]]^{-1} \quad (39d)$$

$$[\bar{\mathbf{R}}^{(0)}[k]]_{lm} = \left[\frac{\partial \mathbf{h}^\top}{\partial x_l} \boldsymbol{\Sigma}^{-1} \frac{\partial \mathbf{h}}{\partial x_m} + \frac{1}{2} \text{tr} \left[\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial x_l} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial x_m} \right] \right]_{\hat{\mathbf{x}}[k|k-1]} \quad (39e)$$

4.2 Sensors with scalar output

We now specialize (38) and (39) for the case in which the output of the sensor is the scalar quantity z , as for bearing-only sensors described in Sec. 2.3.2. Therefore the dimension of

the observation space is $p = 1$. We introduce the scalar-valued functions h and $\sigma > 0$, which represent the observation's expected value and noise variance, respectively. The functions h and σ are therefore the scalar equivalent of \mathbf{h} and $\mathbf{\Sigma}$ in Section 3. Moreover, we introduce the scalar function $\zeta := z - h$.

The iterated extended Kalman filter updates for sensors with scalar output are given by

$$\hat{\mathbf{x}}[k | k] = \hat{\mathbf{x}}[k | k - 1] - [\mathbf{P}^{-1}[k | k - 1] + \mathbf{R}^{(\iota)}[k]]^{-1} \mathbf{s}^{(\iota)}[k] \quad (40a)$$

$$\mathbf{s}^{(\iota)}[k] = \mathbf{P}^{-1}[k | k - 1] (\hat{\mathbf{x}}^{(\iota)}[k | k] - \hat{\mathbf{x}}[k | k - 1]) + \left[-\frac{\zeta}{\sigma} \nabla_{\mathbf{x}}^{\top} h + \frac{1}{2\sigma} \left(1 - \frac{\zeta^2}{\sigma} \right) \nabla_{\mathbf{x}}^{\top} \sigma \right]_{\hat{\mathbf{x}}^{(\iota)}[k | k]} \quad (40b)$$

$$\begin{aligned} \mathbf{R}^{(\iota)}[k] = & \left[\frac{1}{\sigma} \nabla_{\mathbf{x}}^{\top} h \nabla_{\mathbf{x}} h + \frac{\zeta}{\sigma^2} (\nabla_{\mathbf{x}}^{\top} h \nabla_{\mathbf{x}} \sigma + \nabla_{\mathbf{x}}^{\top} \sigma \nabla_{\mathbf{x}} h) + \frac{1}{\sigma^2} \left(\frac{\zeta^2}{\sigma} - \frac{1}{2} \right) \nabla_{\mathbf{x}}^{\top} \sigma \nabla_{\mathbf{x}} \sigma \right. \\ & \left. - \frac{\zeta}{\sigma} \nabla_{\mathbf{x}} \nabla_{\mathbf{x}}^{\top} h + \frac{1}{2\sigma} \left(1 - \frac{\zeta^2}{\sigma} \right) \nabla_{\mathbf{x}} \nabla_{\mathbf{x}}^{\top} \sigma \right]_{\hat{\mathbf{x}}^{(\iota)}[k | k]} \end{aligned} \quad (40c)$$

$$\mathbf{P}^{(\iota)}[k | k] = [\mathbf{P}^{-1}[k | k - 1] + \bar{\mathbf{R}}^{(\iota)}[k]]^{-1} \quad (40d)$$

$$\bar{\mathbf{R}}^{(\iota)}[k] = \left[\frac{1}{\sigma} \nabla_{\mathbf{x}}^{\top} h \nabla_{\mathbf{x}} h + \frac{1}{2\sigma^2} \nabla_{\mathbf{x}}^{\top} \sigma \nabla_{\mathbf{x}} \sigma \right]_{\hat{\mathbf{x}}^{(\iota)}[k | k]} \quad (40e)$$

For a one-step iteration with $\hat{\mathbf{x}}^{(\iota)}[k | k] = \hat{\mathbf{x}}[k | k - 1]$ we obtain the extended Kalman filter for sensors with scalar output:

$$\hat{\mathbf{x}}[k | k] = \hat{\mathbf{x}}[k | k - 1] - [\mathbf{P}^{-1}[k | k - 1] + \mathbf{R}^{(0)}[k]]^{-1} \mathbf{s}^{(0)}[k] \quad (41a)$$

$$\mathbf{s}^{(0)}[k] = \left[-\frac{\zeta}{\sigma} \nabla_{\mathbf{x}}^{\top} h + \frac{1}{2\sigma} \left(1 - \frac{\zeta^2}{\sigma} \right) \nabla_{\mathbf{x}}^{\top} \sigma \right]_{\hat{\mathbf{x}}[k | k - 1]} \quad (41b)$$

$$\begin{aligned} \mathbf{R}^{(0)}[k] = & \left[\frac{1}{\sigma} \nabla_{\mathbf{x}}^{\top} h \nabla_{\mathbf{x}} h + \frac{\zeta}{\sigma^2} (\nabla_{\mathbf{x}}^{\top} h \nabla_{\mathbf{x}} \sigma + \nabla_{\mathbf{x}}^{\top} \sigma \nabla_{\mathbf{x}} h) + \frac{1}{\sigma^2} \left(\frac{\zeta^2}{\sigma} - \frac{1}{2} \right) \nabla_{\mathbf{x}}^{\top} \sigma \nabla_{\mathbf{x}} \sigma \right. \\ & \left. - \frac{\zeta}{\sigma} \nabla_{\mathbf{x}} \nabla_{\mathbf{x}}^{\top} h + \frac{1}{2\sigma} \left(1 - \frac{\zeta^2}{\sigma} \right) \nabla_{\mathbf{x}} \nabla_{\mathbf{x}}^{\top} \sigma \right]_{\hat{\mathbf{x}}[k | k - 1]} \end{aligned} \quad (41c)$$

$$\mathbf{P}[k | k] = [\mathbf{P}^{-1}[k | k - 1] + \bar{\mathbf{R}}^{(0)}[k]]^{-1} \quad (41d)$$

$$\bar{\mathbf{R}}^{(0)}[k] = \left[\frac{1}{\sigma} \nabla_{\mathbf{x}}^{\top} h \nabla_{\mathbf{x}} h + \frac{1}{2\sigma^2} \nabla_{\mathbf{x}}^{\top} \sigma \nabla_{\mathbf{x}} \sigma \right]_{\hat{\mathbf{x}}[k | k - 1]} \quad (41e)$$

5 Estimation in a decentralized communication network

5.1 Sensors network

In Section 4 we derived the update equations of an estimator that generalizes the extended Kalman filter and the iterated extended Kalman filter for a single sensor. The same analysis, with appropriate adaptation of the observation space dimension, directly applies to a centralized system in which a set of sensors communicate only with a central unit that has complete knowledge of the group [17].

In this Section we consider a sensor network comprised of N nodes, in which all sensor nodes acquire measurements, each sensor independently computes a local estimate of the system state \mathbf{x} based on information sensed locally and by information that is communicated to it by other sensor nodes the network. For a fully connected network, in which each sensor receives and sends all the measurements from and to the other sensors in the network at each update instant, individual estimates are the same for all the sensors [27]. Network topologies for which local estimates are identical through information sharing and local data fusion have been studied in [6, 17, 26]. In contrast, we address the case where each local estimate may be different at any point in time. We briefly introduce the Bayesian formalism to describe estimation across a sensor network with a time-varying network topology.

Consider the case in which there is no external unit receiving and sending information, so that the agents exchange information through direct communication. This leads to the definition of a decentralized communication network structure associated to the group of sensors. Individual agents locally acquire measurements $\mathbf{z}_i[k]$ at discrete times k according to the model in (3). In order to model the communication network, we define the time-dependent sets of integers $\mathcal{I}_i[k] \subseteq \{1, 2, \dots, N\}$, for $i = 1, 2, \dots, N$ that represents the set of sensor nodes that communicate with sensor i at time k . Note that $i \in \mathcal{I}_i[k]$ for all k , meaning that each sensor node has access to local information. Therefore, if vehicle i does not receive any data at time k we have $\mathcal{I}_i[k] = \{i\}$. Each local state estimate is denoted by $\hat{\mathbf{x}}_i$. Moreover, we assume that

$$p(\mathbf{z}_i[k] | \mathbf{z}_j[k], \mathbf{x}[k]) = p(\mathbf{z}_i[k] | \mathbf{x}[k]) \quad \forall j \neq i \quad (42)$$

This formalizes the assumption that measurements from different sensors are independent, see [1]. A more refined structure that accounts for correlation between observations is discussed in [15].

5.2 Bayesian implementation

We define the collection of all the available information to sensor i up to time k

$$\mathcal{Z}_i[k] := \bigcup_{l=0}^k \{\mathbf{z}_j[l] : j \in \mathcal{I}_i[l]\} \quad (43)$$

A local estimate at sensor node i is computed from the posterior probability distribution $p(\mathbf{x}[k] | \mathcal{Z}_i[k])$, which accounts for all the data available to sensor i up to time k .

As in (12), application of Bayes' rule, see [1, Appendix B.11], yields the posterior conditional distribution of $\mathbf{x}[k]$ given $\mathcal{Z}_i[k]$

$$p(\mathbf{x}[k] | \mathcal{Z}_i[k]) = \frac{p(\mathcal{Z}_i[k] | \mathbf{x}[k]) p(\mathbf{x}[k])}{p(\mathcal{Z}_i[k])} \quad (44)$$

Application of Bayes' rule as in (17) gives the recursive form

$$p(\mathbf{x}[k] | \mathcal{Z}_i[k]) = \beta_i[k] p(\mathbf{x}[k] | \mathcal{Z}_i[k-1]) \prod_{j \in \mathcal{I}_i[k]} p(\mathbf{z}_j[k] | \mathbf{x}[k]) \quad (45)$$

where the proportionality factor is $\beta_i^{-1}[k] = \prod_{j \in \mathcal{I}_i[k]} p(\mathbf{z}_j[k] | \mathcal{Z}_i[k-1])$. The term $p(\mathbf{x}[k] | \mathcal{Z}_i[k-1])$ is the prior distribution, and accounts for sensor's i measurements up to time $k-1$, and for measurements received by sensor i up to time $k-1$.

5.3 State prediction

State prediction in the case of a sensor network is almost identical to that presented in Section 3.2. For each sensor node i , we define the state estimate and the error covariance at time k given the individual and the received observations up to time $l \leq k$ as

$$\hat{\mathbf{x}}_i[k | l] := \mathbb{E}[\mathbf{x}[k] | \mathcal{Z}_i[l]], \quad i = 1, 2, \dots, N \quad (46a)$$

$$\mathbf{P}_i[k | l] := \mathbb{E}\left[(\mathbf{x}_i[k] - \hat{\mathbf{x}}_i[k | l])(\mathbf{x}_i[k] - \hat{\mathbf{x}}_i[k | l])^\top | \mathcal{Z}_i[l]\right] \quad (46b)$$

It is assumed that for each sensor there exist the local state estimate $\hat{\mathbf{x}}_i[k-1 | k-1]$ at time $k-1$, and associated error covariance $\mathbf{P}_i[k-1 | k-1]$. By following exactly the same steps as in Section 3.2, we obtain the following individual state and error covariance predictions:

$$\hat{\mathbf{x}}_i[k | k-1] = \mathbf{f}(\hat{\mathbf{x}}_i[k-1 | k-1]) \quad (47a)$$

$$\mathbf{P}_i[k | k-1] = \nabla_{\mathbf{x}} \mathbf{f}(\hat{\mathbf{x}}_i[k-1 | k-1]) \mathbf{P}_i[k-1 | k-1] \nabla_{\mathbf{x}} \mathbf{f}^\top(\hat{\mathbf{x}}_i[k-1 | k-1]) + \mathbf{Q}[k] \quad (47b)$$

5.4 State update

The Bayesian scheme in (45), which is a direct consequence of assumption (42), is known as independent likelihood pool, see [27, Section 2.2.2]. Its implementation relies on communication of either raw sensor data $\mathbf{z}_j[k]$ or likelihoods $p(\mathbf{z}_j[k] | \mathbf{x}[k])$. According to the type of data communicated, we derive the two corresponding update equations.

The negative log-likelihood function for sensor i with included sensor data received at time

k is given by

$$\begin{aligned} \ell_i(\mathbf{x}[k]) &= \frac{1}{2} \sum_{j \in \mathcal{I}_i[k]} \left(\ln \det \boldsymbol{\Sigma}_j(\mathbf{x}[k]) + (\mathbf{z}_j[k] - \mathbf{h}_j(\mathbf{x}[k]))^\top \boldsymbol{\Sigma}_j^{-1}(\mathbf{x}[k]) (\mathbf{z}_j[k] - \mathbf{h}_j(\mathbf{x}[k])) \right) \\ &\quad + \frac{1}{2} (\mathbf{x}[k] - \hat{\mathbf{x}}_i[k|k-1])^\top \mathbf{P}_i^{-1}[k|k-1] (\mathbf{x}[k] - \hat{\mathbf{x}}_i[k|k-1]) + C \end{aligned} \quad (48)$$

where C does not depend on \mathbf{x} . The local state update is obtained by applying the algorithm explained in Section 3.3, that is, by using Newton-Raphson iterations to solve the following unconstrained minimization problem

$$\hat{\mathbf{x}}_i[k|k] = \underset{\mathbf{x}}{\operatorname{argmin}} \ell_i(\mathbf{x}[k]) \quad (49)$$

5.4.1 Communication of raw sensor data

Sensor i computes a local state estimate update using Newton-Raphson iterations with initial guess $\hat{\mathbf{x}}_i[k|k-1]$ to solve the unconstrained minimization problem (49). This procedure requires only the communication of raw sensor data, since the Newton-Raphson iterations are computed using the local prediction as initial guess. At the ι^{th} iteration we have

$$\hat{\mathbf{x}}_i^{(\iota+1)}[k|k] = \hat{\mathbf{x}}_i^{(\iota)}[k|k] - \left[\mathbf{P}_i^{-1}[k|k-1] + \mathbf{R}_i^{(\iota)}[k] \right]^{-1} \mathbf{s}_i^{(\iota)}[k] \quad (50a)$$

$$\begin{aligned} \left[\mathbf{s}_i^{(\iota)}[k] \right]_l &= \mathbf{e}_l^\top \mathbf{P}_i^{-1}[k|k-1] \left(\hat{\mathbf{x}}_i^{(\iota)}[k|k] - \hat{\mathbf{x}}_i[k|k-1] \right) \\ &\quad + \sum_{j \in \mathcal{I}_i[k]} \left[-\frac{\partial \mathbf{h}_j^\top}{\partial x_l} \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\zeta}_j + \frac{1}{2} \operatorname{tr} \left[\boldsymbol{\Sigma}_j^{-1} \frac{\partial \boldsymbol{\Sigma}_j}{\partial x_l} (\mathbf{I}_p - \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\zeta}_j \boldsymbol{\zeta}_j^\top) \right] \right]_{\hat{\mathbf{x}}_i^{(\iota)}[k|k]} \end{aligned} \quad (50b)$$

$$\begin{aligned} \left[\mathbf{R}_i^{(\iota)}[k] \right]_{lm} &= \operatorname{tr} \sum_{j \in \mathcal{I}_i[k]} \left[\frac{\partial \mathbf{h}_j^\top}{\partial x_l} \boldsymbol{\Sigma}_j^{-1} \frac{\partial \mathbf{h}_j}{\partial x_m} + \boldsymbol{\Sigma}_j^{-1} \frac{\partial \boldsymbol{\Sigma}_j}{\partial x_l} \boldsymbol{\Sigma}_j^{-1} \frac{\partial \boldsymbol{\Sigma}_j}{\partial x_m} \left(\boldsymbol{\Sigma}_j^{-1} \boldsymbol{\zeta}_j \boldsymbol{\zeta}_j^\top - \frac{1}{2} \mathbf{I}_p \right) \right. \\ &\quad + \left. \left(\frac{\partial \mathbf{h}_j^\top}{\partial x_l} \boldsymbol{\Sigma}_j^{-1} \frac{\partial \boldsymbol{\Sigma}_j}{\partial x_m} + \frac{\partial \mathbf{h}_j^\top}{\partial x_m} \boldsymbol{\Sigma}_j^{-1} \frac{\partial \boldsymbol{\Sigma}_j}{\partial x_l} \right) \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\zeta}_j - \frac{\partial^2 \mathbf{h}_j^\top}{\partial x_l \partial x_m} \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\zeta}_j \right. \\ &\quad \left. + \frac{1}{2} \boldsymbol{\Sigma}_j^{-1} \frac{\partial^2 \boldsymbol{\Sigma}_j}{\partial x_l \partial x_m} (\mathbf{I}_p - \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\zeta}_j \boldsymbol{\zeta}_j^\top) \right]_{\hat{\mathbf{x}}_i^{(\iota)}[k|k]} \end{aligned} \quad (50c)$$

$$\mathbf{P}_i^{(\iota)}[k|k] = \left[\mathbf{P}_i^{-1}[k|k-1] + \bar{\mathbf{R}}_i^{(\iota)}[k] \right]^{-1} \quad (50d)$$

$$\left[\bar{\mathbf{R}}_i^{(\iota)}[k] \right]_{lm} = \sum_{j \in \mathcal{I}_i[k]} \left[\frac{\partial \mathbf{h}_j^\top}{\partial x_l} \boldsymbol{\Sigma}_j^{-1} \frac{\partial \mathbf{h}_j}{\partial x_m} + \frac{1}{2} \operatorname{tr} \left[\boldsymbol{\Sigma}_j^{-1} \frac{\partial \boldsymbol{\Sigma}_j}{\partial x_l} \boldsymbol{\Sigma}_j^{-1} \frac{\partial \boldsymbol{\Sigma}_j}{\partial x_m} \right] \right]_{\hat{\mathbf{x}}_i^{(\iota)}[k|k]} \quad (50e)$$

The implementation of the algorithm (50) requires either the sensors to be homogeneous, or the model of sensors $j \in \mathcal{I}_i[k]$ to be known by sensor i . By evaluating (50) at the prediction $\hat{\mathbf{x}}_i[k|k-1]$, we obtain the extended Kalman filter for observation models with state-dependent noise.

For complete communication networks we have $\mathcal{I}_i[k] = \{1, \dots, N\}$ for all i, k . This structure is equivalent to the centralized one, in which there is an external unit receiving and sending data to the vehicles, and prior estimates in (38) and (39) are common to all the vehicles [17].

For sensors with scalar output, as the bearing-only sensors described in Section 2.3.2, the update equations are

$$\hat{\mathbf{x}}_i^{(\iota+1)}[k|k] = \hat{\mathbf{x}}_i^{(\iota)}[k|k] - \left[\mathbf{P}_i^{-1}[k|k-1] + \mathbf{R}_i^{(\iota)}[k] \right]^{-1} \hat{\mathbf{s}}_i^{(\iota)}[k] \quad (51a)$$

$$\begin{aligned} \mathbf{s}_i^{(\iota)}[k] &= \mathbf{P}_i^{-1}[k|k-1] \left(\hat{\mathbf{x}}_i^{(\iota)}[k|k] - \hat{\mathbf{x}}_i[k|k-1] \right) \\ &+ \sum_{j \in \mathcal{I}_i[k]} \left[-\frac{\zeta_j}{\sigma_j} \nabla_{\mathbf{x}}^{\top} h_j + \frac{1}{2\sigma_j} \left(1 - \frac{\zeta_j^2}{\sigma_j} \right) \nabla_{\mathbf{x}}^{\top} \sigma_j \right]_{\hat{\mathbf{x}}_i^{(\iota)}[k|k]} \end{aligned} \quad (51b)$$

$$\begin{aligned} \mathbf{R}_i^{(\iota)}[k] &= \sum_{j \in \mathcal{I}_i[k]} \left[\frac{1}{\sigma_j} \nabla_{\mathbf{x}}^{\top} h_j \nabla_{\mathbf{x}} h_j + \frac{\zeta_j}{\sigma_j} (\nabla_{\mathbf{x}}^{\top} h_j \nabla_{\mathbf{x}} \sigma_j + \nabla_{\mathbf{x}}^{\top} \sigma_j \nabla_{\mathbf{x}} h_j) \right. \\ &+ \left. \frac{1}{\sigma_j^2} \left(\frac{\zeta_j^2}{\sigma_j} - \frac{1}{2} \right) \nabla_{\mathbf{x}}^{\top} \sigma_j \nabla_{\mathbf{x}} \sigma_j - \frac{\zeta_j}{\sigma_j} \nabla_{\mathbf{x}} \nabla_{\mathbf{x}}^{\top} h_j + \frac{1}{2\sigma_j} \left(1 - \frac{\zeta_j^2}{\sigma_j} \right) \nabla_{\mathbf{x}} \nabla_{\mathbf{x}}^{\top} \sigma_j \right]_{\hat{\mathbf{x}}_i^{(\iota)}[k|k]} \end{aligned} \quad (51c)$$

$$\mathbf{P}_i^{(\iota)}[k|k] = \left[\mathbf{P}_i^{-1}[k|k-1] + \bar{\mathbf{R}}_i^{(\iota)}[k] \right]^{-1} \quad (51d)$$

$$\bar{\mathbf{R}}_i^{(\iota)}[k] = \sum_{j \in \mathcal{I}_i[k]} \left[\frac{1}{\sigma_j} \nabla_{\mathbf{x}}^{\top} h_j \nabla_{\mathbf{x}} h_j + \frac{1}{2\sigma_j^2} \nabla_{\mathbf{x}}^{\top} \sigma_j \nabla_{\mathbf{x}} \sigma_j \right]_{\hat{\mathbf{x}}_i^{(\iota)}[k|k]} \quad (51e)$$

5.4.2 Communication of likelihoods

In some cases, it is desirable to share likelihoods $p(\mathbf{z}_j[k] | \mathbf{x}[k])$, instead of sensor measurements as in algorithm (50). Since in our setting the likelihoods are Gaussian, this is equivalent to sharing the estimate $\hat{\mathbf{x}}_j$ and the covariance \mathbf{P}_j .

When communicating likelihoods, the estimate $\hat{\mathbf{x}}_j[k|k-1]$, $j \in \mathcal{I}_i[k]$ is the initial guess for the Newton-Raphson iterations that are used to solve (49). For one step iteration process, we obtain the following extended Kalman filter updates

$$\hat{\mathbf{x}}_i[k|k] = \hat{\mathbf{x}}_i[k|k-1] - \left[\mathbf{P}_i^{-1}[k|k-1] + \mathbf{R}_i^{(0)}[k] \right]^{-1} \mathbf{s}_i^{(0)}[k] \quad (52a)$$

$$\left[\mathbf{s}_i^{(0)}[k] \right]_l = \sum_{j \in \mathcal{I}_i[k]} \left[-\frac{\partial \mathbf{h}_j^{\top}}{\partial x_l} \Sigma_j^{-1} \zeta_j + \frac{1}{2} \text{tr} \left[\Sigma_j^{-1} \frac{\partial \Sigma_j}{\partial x_l} (\mathbf{I}_p - \Sigma_j^{-1} \zeta_j \zeta_j^{\top}) \right] \right]_{\hat{\mathbf{x}}_j[k|k-1]} \quad (52b)$$

$$\begin{aligned} \left[\mathbf{R}_i^{(0)}[k] \right]_{lm} &= \text{tr} \sum_{j \in \mathcal{I}_i[k]} \left[\frac{\partial \mathbf{h}_j^{\top}}{\partial x_l} \Sigma_j^{-1} \frac{\partial \mathbf{h}_j}{\partial x_m} + \Sigma_j^{-1} \frac{\partial \Sigma_j}{\partial x_l} \Sigma_j^{-1} \frac{\partial \Sigma_j}{\partial x_m} \left(\Sigma_j^{-1} \zeta_j \zeta_j^{\top} - \frac{1}{2} \mathbf{I}_p \right) \right. \\ &+ \left. \left(\frac{\partial \mathbf{h}_j^{\top}}{\partial x_l} \Sigma_j^{-1} \frac{\partial \Sigma_j}{\partial x_m} + \frac{\partial \mathbf{h}_j^{\top}}{\partial x_m} \Sigma_j^{-1} \frac{\partial \Sigma_j}{\partial x_l} \right) \Sigma_j^{-1} \zeta_j - \frac{\partial^2 \mathbf{h}_j^{\top}}{\partial x_l \partial x_m} \Sigma_j^{-1} \zeta_j \right] \end{aligned}$$

$$+\frac{1}{2}\Sigma_j^{-1}\frac{\partial^2\Sigma_j}{\partial x_l\partial x_m}(\mathbf{I}_p-\Sigma_j^{-1}\zeta_j\zeta_j^\top)\Big]_{\hat{\mathbf{x}}_j[k|k-1]} \quad (52c)$$

$$\mathbf{P}_i[k|k]=\left[\mathbf{P}_i^{-1}[k|k-1]+\bar{\mathbf{R}}_i^{(0)}[k]\right]^{-1} \quad (52d)$$

$$\left[\bar{\mathbf{R}}_i^{(0)}[k]\right]_{lm}=\sum_{j\in\mathcal{I}_i[k]}\left[\frac{\partial\mathbf{h}_j^\top}{\partial x_l}\Sigma_j^{-1}\frac{\partial\mathbf{h}_j}{\partial x_m}+\frac{1}{2}\text{tr}\left[\Sigma_j^{-1}\frac{\partial\Sigma_j}{\partial x_l}\Sigma_j^{-1}\frac{\partial\Sigma_j}{\partial x_m}\right]\right]_{\hat{\mathbf{x}}_j[k|k-1]} \quad (52e)$$

For any time instant k in which there is no communication event relative to vehicle i , that is, for $\mathcal{I}_i[k]=\{i\}$, we modify the notation in (38) and define the partial estimates based only on the i^{th} sensor's own measurement

$$\tilde{\mathbf{x}}_i[k|k]=\hat{\mathbf{x}}_i[k|k-1]-\tilde{\mathbf{S}}_i^{(0)}[k]\tilde{\mathbf{s}}_i^{(0)}[k] \quad (53a)$$

$$\left[\tilde{\mathbf{S}}_i^{(0)}[k]\right]^{-1}=\mathbf{P}_i^{-1}[k|k-1]+\tilde{\mathbf{R}}_i^0[k] \quad (53b)$$

$$\tilde{\mathbf{P}}_i^{-1}[k|k]=\mathbf{P}_i^{-1}[k|k-1]+\tilde{\mathbf{R}}_i^0[k] \quad (53c)$$

where

$$\left[\tilde{\mathbf{s}}_i^{(0)}[k]\right]_l=\text{tr}\left[-\frac{\partial\mathbf{h}_i^\top}{\partial x_l}\Sigma_i^{-1}\zeta_i+\frac{1}{2}\left(\Sigma_i^{-1}\frac{\partial\Sigma_i}{\partial x_l}(\mathbf{I}_p-\Sigma_i^{-1}\zeta_i\zeta_i^\top)\right)\right]_{\hat{\mathbf{x}}_i[k|k-1]} \quad (54a)$$

$$\begin{aligned} \left[\tilde{\mathbf{R}}_i^0[k]\right]_{lm} &= \text{tr}\left[\frac{\partial\mathbf{h}_i^\top}{\partial x_l}\Sigma_i^{-1}\frac{\partial\mathbf{h}_i}{\partial x_m}+\Sigma_i^{-1}\frac{\partial\Sigma_i}{\partial x_l}\Sigma_i^{-1}\frac{\partial\Sigma_i}{\partial x_m}\left(\Sigma_i^{-1}\zeta_i\zeta_i^\top-\frac{1}{2}\mathbf{I}_p\right)\right. \\ &+ \left.\left(\frac{\partial\mathbf{h}_i^\top}{\partial x_l}\Sigma_i^{-1}\frac{\partial\Sigma_i}{\partial x_m}+\frac{\partial\mathbf{h}_i^\top}{\partial x_m}\Sigma_i^{-1}\frac{\partial\Sigma_i}{\partial x_l}\right)\Sigma_i^{-1}\zeta_i-\frac{\partial^2\mathbf{h}_i^\top}{\partial x_l\partial x_m}\Sigma_i^{-1}\zeta_i\right. \\ &\left.+\frac{1}{2}\Sigma_i^{-1}\frac{\partial^2\Sigma_i}{\partial x_l\partial x_m}(\mathbf{I}_p-\Sigma_i^{-1}\zeta_i\zeta_i^\top)\right]_{\hat{\mathbf{x}}_i[k|k-1]} \end{aligned} \quad (54b)$$

$$\left[\tilde{\mathbf{R}}_i^0[k]\right]_{lm}=\left[\frac{\partial\mathbf{h}_i^\top}{\partial x_l}\Sigma_i^{-1}\frac{\partial\mathbf{h}_i}{\partial x_m}+\frac{1}{2}\text{tr}\left[\Sigma_i^{-1}\frac{\partial\Sigma_i}{\partial x_l}\Sigma_i^{-1}\frac{\partial\Sigma_i}{\partial x_m}\right]\right]_{\hat{\mathbf{x}}_i[k|k-1]} \quad (54c)$$

By combining (52d), (52e), and (54c) we obtain the covariance assimilation equation [11]

$$\mathbf{P}_i^{-1}[k|k]=\mathbf{P}_i^{-1}[k|k-1]+\sum_{j\in\mathcal{I}_i[k]}\left(\tilde{\mathbf{P}}_j^{-1}[k|k]-\mathbf{P}_j^{-1}[k|k-1]\right) \quad (55)$$

which trivially holds for $\mathcal{I}_i[k]=\{i\}$. In order to obtain the state estimate assimilation equation, we define

$$\left[\mathbf{S}_i^{(0)}[k]\right]^{-1}=\mathbf{P}_i^{-1}[k|k-1]+\mathbf{R}_i^{(0)}[k] \quad (56)$$

and, by using definitions (52c) and (54b) we write

$$\left[\mathbf{S}_i^{(0)}[k]\right]^{-1}=\mathbf{P}_i^{-1}[k|k-1]+\sum_{j\in\mathcal{I}_i[k]}\left(\left[\tilde{\mathbf{S}}_j^{(0)}[k]\right]^{-1}-\mathbf{P}_j^{-1}[k|k-1]\right) \quad (57)$$

Moreover, from (53a) and (52b) we have

$$\begin{aligned} \mathbf{s}_i^{(0)}[k] &= \sum_{j \in \mathcal{I}_i[k]} \tilde{\mathbf{s}}_j^{(0)}[k] \\ &= \sum_{j \in \mathcal{I}_i[k]} \tilde{\mathbf{S}}_j^{(0)}[k] (\hat{\mathbf{x}}_j[k | k - 1] - \tilde{\mathbf{x}}_j[k | k]) \end{aligned} \quad (58)$$

Substitution from (57) and (58) into (52a) gives the state estimate assimilation equation

$$\begin{aligned} \hat{\mathbf{x}}_i[k | k] &= \hat{\mathbf{x}}_i[k | k - 1] \\ &+ \left[\mathbf{P}_i^{-1}[k | k - 1] + \sum_{j \in \mathcal{I}_i[k]} \left(\left[\tilde{\mathbf{S}}_j^{(0)}[k] \right]^{-1} - \mathbf{P}_j^{-1}[k | k - 1] \right) \right]^{-1} \\ &\times \sum_{j \in \mathcal{I}_i[k]} \tilde{\mathbf{S}}_j^{(0)}[k] (\tilde{\mathbf{x}}_j[k | k] - \hat{\mathbf{x}}_j[k | k - 1]) \end{aligned} \quad (59)$$

The covariance assimilation and state assimilation equations have been derived in [17, 11] for the extended Kalman filter and for the extended information filter. Whereas the assimilation equation for the covariance has the same structure, the state assimilation equation is different because the extended Kalman filter state update is linear in $\boldsymbol{\zeta}$, while for the case studied here the dependence on $\boldsymbol{\zeta}$ is much more involved, see (52).

The implementation of the state assimilation equations requires the communication of the covariance error info, the Hessian error info, and the state error info, respectively given by

$$\tilde{\mathbf{P}}_j^{-1}[k | k] - \mathbf{P}_j^{-1}[k | k - 1], \quad \left[\tilde{\mathbf{S}}_j^{(0)}[k] \right]^{-1} - \mathbf{P}_j^{-1}[k | k - 1], \quad \tilde{\mathbf{S}}_j^{(0)}[k] (\tilde{\mathbf{x}}_j[k | k] - \hat{\mathbf{x}}_j[k | k - 1]) \quad (60)$$

for a total of $n(2n + 1)$ scalars. The size of the communication package can be reduced to $n(n + 1)$ scalars if sensor i has knowledge of the sensor model of the other sensors, from which $\tilde{\mathbf{P}}_j^{-1}[k | k]$ can be computed as

$$\tilde{\mathbf{P}}_j[k | k] = \mathbb{E} \left[\tilde{\mathbf{S}}_j^{(0)}[k] \right] \quad (61)$$

As in the case studied in Section 5.4.1, for complete communication networks defined by $\mathcal{I}_i[k] = \{1, \dots, N\}$, $\forall i, k$, the decentralized structure is equivalent to the centralized one, and prior estimates are common to all the vehicles [17].

To summarize, each sensor computes a partial estimate using (53) and (54), and assimilates other estimates eventually received through communication by using (59) and (55). If no data is received at time k , equations (59) and (55) reduce to identities.

6 Conclusions

We have derived the update equations for a class of filters that generalize the extended Kalman filter and the iterated extended Kalman filter to the case in which the noise related

to a nonlinear observation model is a known function of the state to be estimated. The proposed filters are suboptimal in the same sense as the extended Kalman filter, since the probability density function associated to the state estimator is approximated as Gaussian, which leads to the propagation of the first two statistical moments only.

We have also considered a communication network structure that consists of a set of sensors that measure the state of a system. From a Bayesian formulation of the network, we have derived the filter update equations that account for information sharing among sensors. Each sensor maintains an individual estimator, and its individual predictions are updated by using individual and received data. This allows for an improvement of local estimates with respect to the case of a single vehicle operating individually, and therefore relying only on its own measurements.

This work can be applied to formation control problems in which a group of mobile sensors attempt to configure themselves spatially in order to minimize the noise associated to their measurements of a mobile target, and therefore obtain the best estimate of the target.

References

- [1] Y. Bar-Shalom and T. E. Fortman. *Tracking and data association*. Academic Press, 1988.
- [2] B. M. Bell and F. W. Cathey. The iterated Kalman filter update as a Gauss-Newton method. *IEEE Transactions on Automatic Control*, 38(2):294–297, 1993.
- [3] C. Belta and V. Kumar. Abstraction and control for groups of robots. *IEEE Transactions on Robotics*, 20(5):865–875, October 2004.
- [4] D. S. Bernstein. *Matrix Mathematics*. Princeton University Press, 2005.
- [5] D. P. Bertsekas and J. N. Tsitsiklis. *Parallel and distributed computation: numerical methods*. Prentice Hall, Englewood Cliffs, New Jersey, 1989.
- [6] F. Bourgault and H. F. Durrant-Whyte. Communication in general decentralized filters and the coordinated search strategy. In *Proceedings of the Seventh International Conference on Information Fusion*, Stockholm, Sweden, 28 June - 1 July 2004.
- [7] T. H. Chung, J. W. Burdick, and R. M. Murray. Decentralized motion control of mobile sensing agents in a network. In *Proceedings of the IEEE International Conference on Robotics and Automation*, Orlando, Florida, May 2006.
- [8] T. H. Chung, V. Gupta, J. W. Burdick, and R. M. Murray. On a decentralized active sensing strategy using mobile sensor platforms in a network. In *Proceedings of the IEEE conference on Decision and Control*, Paradise Island, Bahamas, December 2004.
- [9] J. Cortés, S. Martínez, T. Karatas, and F. Bullo. Coverage control for mobile sensing networks. *IEEE Transactions on Robotics and Automation*, 20(2):243–255, 2004.
- [10] K. Doğançai. On the efficiency of a bearings-only instrumental variable estimator for a target motion analysis. *Signal Processing*, 85:481–490, 2005.
- [11] H. F. Durrant-Whyte, B. Y. S. Rao, and H. Hu. Toward a fully decentralized architecture for multi-sensor data fusion. In *Proceedings of the IEEE International Conference on Robotics and Automation*, volume 2, pages 1331 – 1336, 13-18 May 1990.
- [12] A. Farina. Target tracking with bearings-only measurements. *Signal Processing*, 78:61–78, 1999.
- [13] J. A. Fax and R. M. Murray. Information flow and cooperative control of vehicle formations. *IEEE Transactions on Automatic Control*, 49(9):1465–1476, September 2004.
- [14] R. A. Freeman, P. Yang, and K. M. Lynch. Distributed estimation and control of swarm formation statistics. In *Proceedings of the American Control Conference*, pages 749–755, Minneapolis, Minnesota USA, June 14-16 2006.

- [15] N. Friedman, D. Geiger, and M. Goldszmidt. Bayesian network classifiers. *Machine Learning*, 29(2):131–163, October 1997.
- [16] A. Gadre, M. Roan, and D. J. Stilwell. Sensor model for a uniform linear array. Technical Report 2008-01, VaCAS, 2008.
- [17] S. Grime and H. F. Durrant-Whyte. Data fusion in decentralized sensor networks. *Control Engineering Practice*, 2(5):849–863, 1994.
- [18] R. A. Iltis and K. L. Anderson. A consistent estimation criterion for multisensor bearing-only tracking. *IEEE Transactions on Aerospace and Electronic Systems*, 32(1):108–120, January 1996.
- [19] K. Ito and K. Xiong. Gaussian filters for nonlinear filtering problems. *IEEE Transactions on Automatic Control*, 45(5):910–927, May 2000.
- [20] J. Julier, S. Uhlmann and H. F. Durrant-Whyte. A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Transactions on Automatic Control*, 45(3):477–482, 2000.
- [21] J. P. Le Cadre and C. Jauffret. On the convergence of iterative methods for bearing-only tracking. *IEEE Transactions on Aerospace and Electronic Systems*, 35(3):801–818, July 1999.
- [22] T Lefebvre, H. Bruyninckx, and J. De Schutter. Comment on ‘A new method for the nonlinear transformation of means and covariances in filters and estimators’. *IEEE Transactions on Automatic Control*, 47(8):1406–1408, 2002.
- [23] T. Lefebvre, H. Bruyninckx, and J. De Shutter. Kalman filters for non-linear systems: a comparison of performance. *International Journal of Control*, 77(7):639–653, 2004.
- [24] A. Logothetis, A. Isaksson, and R. J. Evans. Comparison of suboptimal strategies for optimal own-ship maneuvers in bearing-only tracking. In *Proceedings of the American Control Conference*, pages 3334–3338, Philadelphia, Pennsylvania, June 1996.
- [25] A. Logothetis, A. Isaksson, and R. J. Evans. An information theoretic approach to observer path design for bearings-only tracking. In *Proceedings of the 36th Conference on Decision and Control*, pages 3132–3137, San Diego, California, Dec. 1997.
- [26] A. Makarenko and H. F. Durrant-Whyte. Decentralized data fusion and control in active sensor networks. In *Proceedings of the Seventh International Conference on Information Fusion*, Stockholm, Sweden, 28 June - 1 July 2004.
- [27] J. Manyika and H. Durrant-Whyte. *Data fusion and sensor management: a decentralized information-theoretic approach*. Ellis Horwood, London, 1994.
- [28] S. Martínez and F. Bullo. Optimal sensor placement and motion coordination for target tracking. *Automatica*, 42(4):661–668, 2006.

- [29] P. S. Maybeck. *Stochastic models, estimation, and control*, volume 2 of *Mathematics in Science and Engineering*. Academic Press, New York, 1982.
- [30] A. G. O. Mutambara. *Decentralized estimation and control for multisensor systems*. CRC Press LLC, Boca Raton, Florida, 1998.
- [31] M. Nørgaard, N. Poulsen, and O. Ravn. New developments in state estimations for nonlinear systems. *Automatica*, 36(11):1627–1638, 2000.
- [32] Y. Oshman and P. Davidson. Optimization of observer trajectories for bearings-only target localization. *IEEE Transactions on Aerospace and Electronic Systems*, 35(3):892–902, 1999.
- [33] M. Porfiri, D. G. Roberson, and D. J. Stilwell. Tracking and formation control of multiple autonomous agents: A two-level consensus approach. *Automatica*, 43(8):1318–1328, 2007.
- [34] T. Schei. A finite-difference method for linearisation in nonlinear estimation algorithms. *Automatica*, 33(11):2053–2058, 1997.
- [35] J. Shao. *Mathematical Statistics*. Springer texts in statistics. Springer Verlag, New York, 1999.
- [36] G. Sibley, G. Suckhatme, and L. Matthies. The iterated sigma point Kalman filter with application to long range stereo. In *Proceedings of Robotics: Science and Systems*, Philadelphia, PA, August 2006.
- [37] S. Simic and S. Sastry. Distributed environmental monitoring using random sensor networks. In *Proceeding of the 2nd International Workshop on Information Processing in Sensor Networks*, pages 582–592, Palo Alto, CA, 2003.
- [38] S. Susca, S. Martínez, and F. Bullo. Monitoring environmental boundaries with a robotic sensor network. In *Proceedings of the American Control Conference*, pages 2072–2077, 2006.
- [39] S. Thrun, Y. Liu, Koller D., and A. Y. Ng. Simultaneous localization and mapping with sparse extended information filters. *International Journal of Robotics Research*, 23(7-8):693–716, July-August 2004.
- [40] R. van der Merve. *Sigma-point Kalman filters for probabilistic inference in dynamic state-space models*. PhD thesis, Oregon Health & Science University, OGI School of Science & Engineering, 2004.
- [41] P. Yang, R. A. Freeman, and K. M. Lynch. Distributed cooperative active sensing using consensus filters. In *Proceedings of the IEEE International Conference on Robotics and Automation*, Roma, Italy, Feb. 2007.
- [42] B. Zehnwirth. A generalization of the Kalman filter for models with state-dependent observation variance. *Journal of the American Statistical Association*, 83(401):164–167, 1988.

- [43] R. Zhan and J. Wan. Iterated unscented Kalman filter for passive target tracking. *IEEE Transactions on Aerospace and Electronic Systems*, 43(3):1155–1163, July 2007.
- [44] K. X. Zhou and S. I. Roumeliotis. Optimal motion strategies for range-only distributed target tracking. In *Proceedings of the American Control Conference*, pages 5195–5200, Minneapolis, Minnesota USA, June 14-16 2006.