Combined Stochastic and Set-membership Information Filtering in Multisensor Systems

Benjamin Noack, Florian Pfaff, and Uwe D. Hanebeck
Intelligent Sensor-Actuator-Systems Laboratory (ISAS),
Institute for Anthropomatics,
Karlsruhe Institute of Technology (KIT), Germany.
Email: benjamin.noack@ieee.org, florian.pfaff@student.kit.edu, uwe.hanebeck@ieee.org

Abstract—In state estimation theory, stochastic and set-membership approaches are generally considered separately from each other. Both concepts have distinct advantages and disadvantages making each one inherently better suited to model different sources of estimation uncertainty. In order to better utilize the potentials of both concepts, the core element of this paper is a Kalman filtering scheme that allows for a simultaneous treatment of stochastic and set-membership uncertainties. An uncertain quantity is herein modeled by a set of Gaussian densities. Since many modern applications operate in networked systems that may consist of a multitude of local processing units and sensor nodes, estimates have to be computed in a distributed manner and measurements may arrive at high frequency. An algebraic reformulation of the Kalman filter, the information filter, significantly eases the implementation of such distributed fusion architectures. This paper explicates how stochastic and set-membership uncertainties can simultaneously be treated within this information form and compared to the Kalman filter, it becomes apparent that the quality of some required approximations is enhanced.
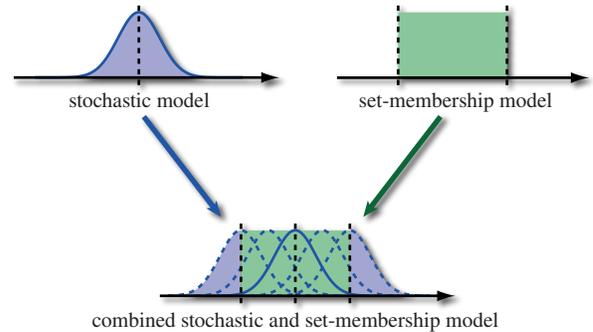
Figure 1. A simultaneous consideration of stochastic and set-membership uncertainties can be achieved by means of sets of probability densities. An additive superposition of a normally distributed and an interval-based error model becomes a set of translated Gaussian densities.

I. INTRODUCTION

Models of uncertainty have become a central component in state estimation frameworks, since neither system dynamics and sensor properties can precisely be identified nor external influences can be taken into account in their entirety. Employing such models can contribute to ensure robustness and reliability in decision and control applications, but for this purpose, it is necessary to propagate and update uncertainty descriptions throughout the entire state estimation process. Essentially, two different directions have been pursued in literature: Bayesian and set-membership state estimation.

In Bayesian state estimators, an uncertain quantity is characterized by a random variable or vector. The generic Bayesian inference scheme then operates on the underlying probability densities and provides an optimal solution to the estimation problem. Unfortunately, this scheme is only of conceptual value, since in general, the absence of finite parameterizations prevents efficient and closed-form calculations of the densities. However, the most notable exception is given by the Kalman filter [1] formulas in the case of linear system dynamics and sensor models perturbed by Gaussian noise. Conditional mean and covariance matrix, which uniquely characterize the posterior Gaussian density, can then be calculated in closed form. Stochastic variables allow for modeling uncertainty over unbounded domains, for example, in order to account for outliers.

Set-membership estimators are most appropriate to cope with unknown but bounded quantities. More precisely, no knowledge is required about the error behavior within the bounds. Very common implementations of set-membership estimation algorithms model bounds by means of ellipsoidal sets [2]–[4], which enable efficient computations of, for instance, Minkowski sums or intersections. Since the latter operation is required for fusing information, the treatment of outliers that may even lead to non-intersecting sets is often difficult.

In the recent past, several approaches towards a simultaneous treatment of both types of uncertainty have been investigated, in order to profit from the individual advantages and to ensure reliable estimation results. For instance, [5] and [6] model uncertainties by means of sum quadratic or relative entropy constraints. Such an approach is useful when a constraint-based description of errors is possible irrespective of their actual nature. Other concepts aspire to explicitly distinguish between stochastic and set-membership uncertainties. [7] considers an extension of the Kalman filter where the gain is computed to minimize a cost function that depends on both the stochastic and the set-membership perturbation term. In this work, we follow the direction of [8], [9], and [10], where a set-bounded uncertainty is incorporated in the prior. This combination of stochastic and set-membership uncertainty then leads to a set of probability densities characterizing the state estimate,

as illustrated in Fig. 1. In [11] and [12], this characterization has been extended to also integrate set-membership errors in the system evolution and sensor observations. In the presence of both types of uncertainty, the proposed Kalman filter based on sets of densities still guarantees that the true mean and error statistics are captured. A deeper look at this concept is provided by Section II. Also, it is strongly related to Kalman filtering approaches based on coherent lower and upper previsions [13], which also correspond to convex sets of probability densities. Alternatively, a combined consideration is also possible by means of random sets [14], [15].

Due to the rapid advances in sensor and communication technologies, the demand for distributed implementations [16] of estimation algorithms is steadily increasing. A distributed formulation of aggregating different estimates and multiple measurements is not trivial in stochastic state estimators, but again for linear systems affected by Gaussian noise, closed-form solutions are derived by means of the information filter [17]–[19], which essentially is an inverse covariance formulation of the Kalman filter. This reformulation encompasses several advantages over the standard Kalman filter formulas with respect to, for instance, the fusion of local estimates or the elimination of double-counted information. In this paper, we show how to also incorporate set-membership uncertainties into the information filter. The according derivations lie in the focus of Sec. III. Due to the required computation of Minkowski sums, approximation errors are generally introduced. However, the discussions in Sec. IV reveal that the information form has advantageous properties in this regard, especially when multiple measurements are received. Also, the distributed evaluation of Minkowski sums is discussed. A performance analysis of the proposed approach and an outlook to prospective work in Sec. V and Sec. VI, respectively, conclude this paper.

II. REVIEW: COMBINED STOCHASTIC AND SET-MEMBERSHIP KALMAN FILTERING

A combined treatment of stochastic and set-membership uncertainties can be achieved by employing a more general concept than classical probability theory, i.e., imprecise probabilities. Instead of characterizing an uncertain quantity by means of a single probability density function, the characterization is expanded to include sets of probability densities, as it is done in [11]–[13]. In this paper, we confine ourselves to additive stochastic and set-membership perturbation terms in order to obtain tractable formulas. Furthermore, the former type of uncertainty is modeled as Gaussian noise, the latter type is modeled by means of ellipsoids, and linear models are considered. Arbitrary models and perturbations have been studied in [11], but they, in general, permit only approximate calculations of posterior estimates.

An uncertain system input or uncertain sensor observation can now be characterized according to

$$\underline{\boldsymbol{u}} = \hat{\underline{u}} + \underline{\boldsymbol{w}} + \underline{d},$$

where $\hat{\underline{u}}$ is the known quantity affected by a zero-mean Gaussian noise $\underline{\boldsymbol{w}} \sim \mathcal{N}(\underline{0}, \mathbf{C})$ with covariance matrix $\mathbf{C}$ and an unknown but bounded perturbation $\underline{d} \in \mathcal{D} \subset \mathbb{R}^n$. For an arbitrary but fixed $\underline{d}$, the quantity $\underline{\boldsymbol{u}}$ can be regarded as a random variable with mean $\hat{\underline{u}} + \underline{d}$. More precisely, $\underline{\boldsymbol{u}}$ is normally distributed according to $\mathcal{N}(\hat{\underline{u}} + \underline{d}, \mathbf{C})$. Since $\underline{d}$ is only specified through its membership to $\mathcal{D}$, we cannot assign a single distribution to $\underline{\boldsymbol{u}}$, but a set

$$\{\mathcal{N}(\hat{\underline{u}} + \underline{d}, \mathbf{C}) \,|\, \underline{d} \in \mathcal{D}\}$$

of distributions, which contains translated versions of a normal distribution. In order to simplify the parameterization and the calculations, which are needed in the following, we confine the set-membership to ellipsoids

$$\mathcal{X} = \mathcal{E}(\hat{\underline{c}}, \mathbf{X}) = \left\{\underline{x} \in \mathbb{R}^n \,|\, (\underline{x} - \hat{\underline{c}})^{\mathrm{T}} \mathbf{X}^{-1} (\underline{x} - \hat{\underline{c}}) \leq 1\right\} .$$

Thus, a set of Gaussian distributions is characterized by an ellipsoidal set $\mathcal{X}$ of means with center $\hat{\underline{c}}$ and nonnegative definite shape matrix $\mathbf{X}$ and a covariance matrix $\mathbf{C}$. The following two subsections will explicate how such a combined stochastic and set-membership state estimate can be updated and be predicted, respectively.

As indicated above, vectorial quantities are underlined and they are marked in boldface type, if random. Uppercase bold letters denote matrices and the calligraphic type is used for sets.

A. Filtering

Let $\underline{\boldsymbol{x}}_k^{\mathrm{p}}$ be a prior or predicted state estimate that has the error characteristics $\mathcal{E}(\hat{\underline{x}}_k^{\mathrm{p}}, \mathbf{X}_k^{\mathrm{p}})$ and $\mathbf{C}_k^{\mathrm{p}}$ for the set-membership uncertainties and the stochastic uncertainties, respectively. Observed sensor data

$$\hat{\underline{\boldsymbol{z}}}_k = \underline{\boldsymbol{z}}_k + \underline{\boldsymbol{v}}_k + \underline{e}_k = \mathbf{H}_k \underline{\boldsymbol{x}}_k + \underline{\boldsymbol{v}}_k + \underline{e}_k$$

is assumed to be potentially affected by both stochastic and set-membership disturbances. The former errors are modeled as a zero-mean white Gaussian noise $\underline{\boldsymbol{v}}_k$ with covariance matrix $\mathbf{C}_k^z$. The latter bounded error term $\underline{e}_k$ is considered to lie within the ellipsoidal set $\mathcal{E}(\underline{0}, \mathbf{X}_k^z)$. In the case of vanishing set-membership errors, i.e., $\mathbf{X}_k^{\mathrm{p}} = \mathbf{X}_k^z = \mathbf{0}$, fusing the measurement information into the current state estimate is conducted by means of the standard Kalman fusion formulas

$$\begin{aligned}
\hat{\underline{x}}_k^{\mathrm{e}} &= \hat{\underline{x}}_k^{\mathrm{p}} + \mathbf{K}_k(\hat{\underline{z}}_k - \mathbf{H}_k \hat{\underline{x}}_k^{\mathrm{p}}) \\
&= (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)\hat{\underline{x}}_k^{\mathrm{p}} + \mathbf{K}_k \hat{\underline{z}}_k
\end{aligned} \tag{1}$$

and

$$\mathbf{C}_k^{\mathrm{e}} = \mathbf{C}_k^{\mathrm{p}} - \mathbf{K}_k \mathbf{H}_k \mathbf{C}_k^{\mathrm{p}} \tag{2}$$

for estimated mean and covariance matrix, respectively. The Kalman gain is given by

$$\mathbf{K}_k = \mathbf{C}_k^{\mathrm{p}} \mathbf{H}_k^{\mathrm{T}} (\mathbf{C}_k^z + \mathbf{H}_k \mathbf{C}_k^{\mathrm{p}} \mathbf{H}_k^{\mathrm{T}})^{-1} .$$

In the presence of set-membership uncertainty, we regard the measurement $\hat{\underline{z}}_k$ combined with the ellipsoidal error bound as an ellipsoidal set

$$\mathcal{Z}_k = \left\{\hat{\underline{z}}_k - \underline{e}_k \,|\, \underline{e}_k \in \mathcal{E}(\underline{0}, \mathbf{X}_k^z)\right\} = \mathcal{E}(\hat{\underline{z}}_k, \mathbf{X}_k^z)$$

of possible measurements. With this set, the set $\mathcal{E}(\hat{\underline{x}}_k^{\mathrm{p}}, \mathbf{X}_k^{\mathrm{p}})$ of predicted or prior means can be updated by applying (1) elementwise, i.e.,

$$
\begin{aligned}
\mathcal{X}_k^{\mathrm{e}} &\stackrel{\text{eq. (1)}}{=} (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)\, \mathcal{E}(\hat{\underline{x}}_k^{\mathrm{p}}, \mathbf{X}_k^{\mathrm{p}}) \oplus \mathbf{K}_k\, \mathcal{E}(\hat{\underline{z}}_k, \mathbf{X}_k^z) \\
&\stackrel{\text{eq. (4)}}{=} \mathcal{E}\big((\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)\, \hat{\underline{x}}_k^{\mathrm{p}}, (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)\mathbf{X}_k^{\mathrm{p}}(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)^{\mathrm{T}}\big) \\
&\qquad \oplus \mathcal{E}(\hat{\underline{z}}_k, \mathbf{K}_k \mathbf{X}_k^z \mathbf{K}_k^{\mathrm{T}}) \\
&\stackrel{\text{eq. (5)}}{\subseteq}\ \mathcal{E}(\hat{\underline{x}}_k^{\mathrm{e}}, \mathbf{X}_k^{\mathrm{e}})\ ,
\end{aligned}
$$
(3)

where the formula

$$
\mathbf{A}\mathcal{E}(\hat{\underline{c}}, \mathbf{X}) + \underline{b} = \mathcal{E}(\mathbf{A}\,\hat{\underline{c}} + \underline{b}, \mathbf{A}\mathbf{X}\mathbf{A}^{\mathrm{T}})
$$
(4)

for affine transformations of ellipsoids and

$$
\mathcal{E}(\hat{\underline{c}}_1, \mathbf{X}_1) \oplus \mathcal{E}(\hat{\underline{c}}_2, \mathbf{X}_2) \subseteq \mathcal{E}(\hat{\underline{c}}_1 + \hat{\underline{c}}_2, \mathbf{X}(p))
$$
(5)

for Minkowski sums of ellipsoids with

$$
\mathbf{X}(p) = (1 + p^{-1})\mathbf{X}_1 + (1 + p)\mathbf{X}_2
$$
(6)

have been applied [4]. The parameter p that optimizes the trace (6), which corresponds to sum of squares of semi-axes, is obtained by

$$
p = \mathrm{tr}(\mathbf{X}_1)^{\frac{1}{2}} \cdot \mathrm{tr}(\mathbf{X}_2)^{-\frac{1}{2}}\ .
$$
(7)

Hence, the shape matrix of the outer approximation (3) yields

$$
\begin{aligned}
\mathbf{X}_k^{\mathrm{e}} = {} & (1 + p^{-1})(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)\mathbf{X}_k^{\mathrm{p}}(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)^{\mathrm{T}} \\
& + (1 + p)\mathbf{K}_k \mathbf{X}_k^z \mathbf{K}_k^{\mathrm{T}}
\end{aligned}
$$

and the estimated covariance matrix is given by the standard Kalman filter formula (2) since its computation is independent of the mean. Choosing the trace as optimality criterion is reasonable, since selfsame criterion is used to determine the Kalman gain in (2).

B. Prediction

The system's state evolution is characterized by means of a discrete-time linear model

$$
\underline{x}_{k+1} = \mathbf{A}_k\,\underline{x}_k + \mathbf{B}_k\,(\hat{\underline{u}}_k + \underline{w}_k + \underline{d}_k)\ ,
$$
(8)

where $\mathbf{A}_k$ is the system mapping and $\mathbf{B}_k$ is the control-input matrix. The perturbation terms $\underline{w}_k$ and $\underline{d}_k$ are a zero-mean white Gaussian noise with covariance matrix $\mathbf{C}_k^u$ and a set-membership error enclosed by $\mathcal{E}(\underline{0}, \mathbf{X}_k^u)$, respectively. The control input $\hat{\underline{u}}$ combined with the set-membership error can be regarded as a set $\mathcal{E}(\hat{\underline{u}}_k, \mathbf{X}_k^u)$ of possible control inputs. The predicted set of means therefore yields the Minkowski sum

$$
\begin{aligned}
\mathcal{X}_{k+1}^{\mathrm{p}} &\stackrel{\text{eq. (8)}}{=} \mathbf{A}_k\, \mathcal{E}(\hat{\underline{x}}_k^{\mathrm{e}}, \mathbf{X}_k^{\mathrm{e}}) \oplus \mathbf{B}_k \mathcal{E}(\hat{\underline{u}}_k, \mathbf{X}_k^u) \\
&\stackrel{\text{eq. (4)}}{=} \mathcal{E}(\mathbf{A}_k\,\hat{\underline{x}}_k^{\mathrm{e}}, \mathbf{A}_k \mathbf{X}_k^{\mathrm{e}} \mathbf{A}_k^{\mathrm{T}}) \oplus \mathcal{E}(\mathbf{B}_k\,\hat{\underline{u}}_k, \mathbf{B}_k \mathbf{X}_k^u \mathbf{B}_k^{\mathrm{T}}) \\
&\stackrel{\text{eq. (5)}}{\subseteq}\ \mathcal{E}(\hat{\underline{x}}_{k+1}^{\mathrm{p}}, \mathbf{X}_{k+1}^{\mathrm{p}})\ ,
\end{aligned}
$$

where the latter ellipsoid externally approximates the sum with

$$
\hat{\underline{x}}_{k+1}^{\mathrm{p}} = \mathbf{A}_k\,\hat{\underline{x}}_k^{\mathrm{e}} + \mathbf{B}_k\,\hat{\underline{u}}_k
$$

and

$$
\mathbf{X}_{k+1}^{\mathrm{p}} = (1 + p^{-1})\mathbf{A}_k \mathbf{X}_k^{\mathrm{e}} \mathbf{A}_k^{\mathrm{T}} + (1 + p)\mathbf{B}_k \mathbf{X}_k^u \mathbf{B}_k^{\mathrm{T}}\ .
$$

Choosing p according to (7) again minimizes the trace of $\mathbf{X}_{k+1}^{\mathrm{p}}$. The associated covariance matrix is stated by

$$
\mathbf{C}_{k+1}^{\mathrm{p}} = \mathbf{A}_k \mathbf{C}_k^{\mathrm{e}} \mathbf{A}_k^{\mathrm{T}} + \mathbf{B}_k \mathbf{C}_k^u \mathbf{B}_k^{\mathrm{T}}\ ,
$$

which is the same for every possible mean out of $\mathcal{E}(\hat{\underline{x}}_{k+1}^{\mathrm{p}}, \mathbf{X}_{k+1}^{\mathrm{p}})$.

III. Combined Stochastic and Set-Membership Information Filtering

The information filter [17] essentially embodies an algebraic reformulation of the Kalman filter formulas and it provides estimates on the information about an uncertain state rather than on the state itself. More exactly, the information state vector

$$
\hat{\underline{y}}_k := \mathbf{C}_k^{-1} \hat{\underline{x}}_k
$$
(9)

and the information matrix

$$
\mathbf{Y}_k := \mathbf{C}_k^{-1}
$$
(10)

are considered, which are the quantities to be processed and updated in the prediction step and filtering step, respectively. In the presence of set-membership uncertainties, an ellipsoidal set $\mathcal{E}(\hat{\underline{x}}_k, \mathbf{X}_k)$ of means has to be treated in its information form that is obtained by a linear transformation

$$
\begin{aligned}
\mathcal{Y}_k &= \mathbf{C}_k^{-1}\mathcal{E}(\hat{\underline{x}}_k, \mathbf{X}_k) \stackrel{(4)}{=} \mathcal{E}(\mathbf{C}_k^{-1}\hat{\underline{x}}_k, \mathbf{C}_k^{-1}\mathbf{X}_k \mathbf{C}_k^{-\mathrm{T}}) \\
&= \mathcal{E}(\hat{\underline{y}}_k, \mathbf{Q}_k)
\end{aligned}
$$
(11)

with $\mathbf{Q}_k := \mathbf{C}_k^{-1}\mathbf{X}_k \mathbf{C}_k^{-\mathrm{T}}$ and thus again yields an ellipsoid comprising all possible information state vectors. The information form entails an easily distributable formulation of the filtering step, as illustrated in the following subsection.

A. Filtering

The standard Kalman filtering step, i.e., equations (1) and (2), cannot easily be extended to a distributable treatment of multiple observations $\mathcal{Z}_k = \{\hat{\underline{z}}_k^1, \ldots, \hat{\underline{z}}_k^M\}$, as it would be useful in multisensor data fusion problems. More precisely, the fusion step cannot simply be expressed in terms of the individual Kalman gains $\mathbf{K}_k^i$ and innovations $(\hat{\underline{z}}_k^i - \mathbf{H}_k^i \hat{\underline{x}}_k^{\mathrm{p}})$, i.e.,

$$
\hat{\underline{x}}_k^{\mathrm{e}} \neq \hat{\underline{x}}_k^{\mathrm{p}} + \sum_{i=1}^{M} \mathbf{K}_k^i \left(\hat{\underline{z}}_k^i - \mathbf{H}_k^i \hat{\underline{x}}_k^{\mathrm{p}}\right)\ ,
$$
(12)

since the individual innovations are affected by a common process noise, which causes them to be correlated with each other. In general, the Kalman filter only allows for a sequential or blockwise processing [20] of the measurement data.

By means of the inverse covariance matrix formulation, i.e., (9) and (10), the filtering step for the mean (1) and covariance matrix (2) becomes the simple sums

$$
\hat{\underline{y}}_k^{\mathrm{e}} = \hat{\underline{y}}_k^{\mathrm{p}} + \underline{i}_k \quad \text{and} \quad \mathbf{Y}_k^{\mathrm{e}} = \mathbf{Y}_k^{\mathrm{p}} + \mathbf{I}_k
$$
(13)

for the information state and matrix, respectively, with

$$\underline{i}_k = \mathbf{H}_k^{\mathrm{T}}(\mathbf{C}_k^z)^{-1}\underline{\hat{z}}_k \quad \text{and} \quad \mathbf{I}_k = \mathbf{H}_k^{\mathrm{T}}(\mathbf{C}_k^z)^{-1}\mathbf{H}_k \ .$$

Unlike (12), multiple measurements can, in this case, be incorporated into the filtered information state by generalizing (13) to

$$\underline{\hat{y}}_k^{\mathrm{e}} = \underline{\hat{y}}_k^{\mathrm{p}} + \sum_{i=1}^{M} \underline{i}_k^i = \underline{\hat{y}}_k^{\mathrm{p}} + \sum_{i=1}^{M} (\mathbf{H}_k^i)^{\mathrm{T}}(\mathbf{C}_k^{z,i})^{-1}\underline{\hat{z}}_k^i \qquad (14)$$

and

$$\mathbf{Y}_k^{\mathrm{e}} = \mathbf{Y}_k^{\mathrm{p}} + \sum_{i=1}^{M} \mathbf{I}_k^i = \mathbf{Y}_k^{\mathrm{p}} + \sum_{i=1}^{M} (\mathbf{H}_k^i)^{\mathrm{T}}(\mathbf{C}_k^{z,i})^{-1}\mathbf{H}_k^i \ , \quad (15)$$

where parts of this calculation can now easily be distributed. More precisely, the measurement data along any communication path can be collected and already be condensed into a single information vector, which has the dimension of the system state.

For an ellipsoidal information state (11) and set-membership perturbations of the measurements, i.e.,

$$\begin{aligned}
\mathcal{I}_k^i &= (\mathbf{H}_k^i)^{\mathrm{T}}(\mathbf{C}_k^{z,i})^{-1}\mathcal{E}(\underline{\hat{z}}_k^i, \mathbf{X}_k^{z,i}) \\
&= \mathcal{E}\Big((\mathbf{H}_k^i)^{\mathrm{T}}(\mathbf{C}_k^{z,i})^{-1}\underline{\hat{z}}_k^i, (\mathbf{H}_k^i)^{\mathrm{T}}(\mathbf{C}_k^{z,i})^{-1}\mathbf{X}_k^{z,i}(\mathbf{C}_k^{z,i})^{-\mathrm{T}}\mathbf{H}_k^i\Big)
\end{aligned}$$
$$(16)$$

in information form, equation (14) turns into the Minkowski sum

$$\mathcal{Y}_k^{\mathrm{e}} = \mathcal{Y}_k^{\mathrm{p}} \oplus \mathcal{I}_k^1 \oplus \ldots \oplus \mathcal{I}_k^M \ . \qquad (17)$$

An outer ellipsoidal approximation of this Minkowski sum can be computed for every part of the sum by means of (5) and thus it can also be implemented in distributed fashion, but this procedure may finally result into a very conservative approximation of the total sum. The distributed computation of an ellipsoid that encloses the total sum and is optimal with respect to the trace of the shape matrix is subject of Sec. IV-A.

B. Prediction

The discussed advantages in the filtering step come at the expense of more elaborate formulas for the prediction step of the information filter. As for the covariance matrix, the predicted information matrix is obtained independently from the information vector by means of

$$\mathbf{Y}_{k+1}^{\mathrm{p}} = \big(\mathbf{A}_k(\mathbf{Y}_k^{\mathrm{e}})^{-1}\mathbf{A}_k^{\mathrm{T}} + \mathbf{B}_k\mathbf{C}_k^u\mathbf{B}_k^{\mathrm{T}}\big)^{-1} \ .$$

The predicted information vector can be computed by

$$\underline{\hat{y}}_{k+1}^{\mathrm{p}} = \mathbf{L}_{k+1}\,\underline{\hat{y}}_k^{\mathrm{e}} + \mathbf{Y}_{k+1}^{\mathrm{p}}\mathbf{B}_k\underline{\hat{u}}_k \qquad (18)$$

with

$$\mathbf{L}_{k+1} = \mathbf{Y}_{k+1}^{\mathrm{p}}\mathbf{A}_k(\mathbf{Y}_k^{\mathrm{e}})^{-1} \ .$$

In the case of ellipsoidal sets (11) of information vectors, (18) turns into the Minkowski sum

$$\begin{aligned}
\mathcal{Y}_{k+1}^{\mathrm{p}} &= \mathbf{L}_{k+1}\mathcal{E}(\underline{\hat{y}}_k^{\mathrm{e}}, \mathbf{Q}_k^{\mathrm{e}}) \oplus \mathbf{Y}_{k+1}^{\mathrm{p}}\mathbf{B}_k\mathcal{E}(\underline{\hat{u}}_k, \mathbf{X}_k^u) \\
&\subseteq \mathcal{E}(\underline{\hat{y}}_{k+1}^{\mathrm{p}}, \mathbf{Q}_{k+1}^{\mathrm{p}}) \ ,
\end{aligned}$$

where the outer approximation $\mathcal{E}(\underline{\hat{y}}_{k+1}^{\mathrm{p}}, \mathbf{Q}_{k+1}^{\mathrm{p}})$ is optimal in the sense of the trace, i.e., the sum of the squared semiaxes lengths, if equations (6) and (7) are deployed to compute $\mathbf{Q}_{k+1}^{\mathrm{p}}$.

Apparently, the prediction step also only requires the approximation of a Minkowski sum of two ellipsoids. The principal difficulty in set-valued information filtering consists of an efficient and distributable approximation of the Minkowski sum (17), which lies in the focus of the following section.

IV. DISTRIBUTED OUTER APPROXIMATION OF SUMS OF INFORMATION ELLIPSOIDS

Aspiring an optimal approximation of the total sum (17) may at first sight conflict with the idea to compute and approximate the sum partially on remote sensor nodes. This section therefore presents an efficient and distributable method for an outer trace-optimal approximation. At first, we can achieve the optimality with respect to the ellipsoid (11) of information vectors. After that, we explain how to ensure that the back-transformed ellipsoid in the state space is trace-optimal.

A. Distributed Optimization

In order to enclose the Minkowski sum (17) by an ellipsoid that is optimal with respect to some criterion, such as the trace, eq. (5) can be generalized into

$$\mathcal{E}(\underline{\hat{c}}_1, \mathbf{X}_1) \oplus \mathcal{E}(\underline{\hat{c}}_2, \mathbf{X}_2) \oplus \ldots \oplus \mathcal{E}(\underline{\hat{c}}_N, \mathbf{X}_N) \subseteq \mathcal{E}(\underline{\hat{c}}^s, \mathbf{X}^s)$$

with

$$\underline{\hat{c}}^s = \sum_{i=1}^{N} \underline{\hat{c}}_i$$

and

$$\mathbf{X}^s = \left(\sum_{i=1}^{N} p_i\right) \cdot \sum_{i=1}^{N} p_i^{-1}\mathbf{X}_i \qquad (19)$$

for some $p_i > 0$ [4], where $\mathcal{E}(\underline{\hat{c}}^s, \mathbf{X}^s)$ is in general not a tight bound for the sum. The set of parameters $\{p_1, \ldots, p_N\}$ has to be determined such that the chosen optimality criterion is fulfilled. Fortunately, considering the trace, as it is done throughout this paper, simplifies matters significantly. In [3], it has been proven that

$$p_i = \sqrt{\mathrm{tr}(\mathbf{X}_i)} \ , \quad \forall i = 1, \ldots, N \qquad (20)$$

minimizes the trace of (19). With (20), the sum (19) furthermore becomes associative in the following sense: For the sum of three ellipsoids with shape matrices $\mathbf{X}_i$, $\mathbf{X}_j$ and $\mathbf{X}_k$, first

$$\mathbf{X}_{j,k} = (p_j + p_k) \cdot \big(p_j^{-1}\mathbf{X}_j + p_k^{-1}\mathbf{X}_k\big)$$

can be computed. The final sum yields

$$\begin{aligned}
\mathbf{X}_{i,j,k} &= (p_i + p_{j,k}) \cdot \big(p_i^{-1}\mathbf{X}_i + p_{j,k}^{-1}\mathbf{X}_{j,k}\big) \\
&= (p_i + p_{j,k}) \cdot \big(p_i^{-1}\mathbf{X}_i + \\
&\qquad p_{j,k}^{-1} \cdot (p_j + p_k) \cdot (p_j^{-1}\mathbf{X}_j + p_k^{-1}\mathbf{X}_k)\big) \\
&= (p_i + p_j + p_k) \cdot \big(p_i^{-1}\mathbf{X}_i + p_j^{-1}\mathbf{X}_j + p_k^{-1}\mathbf{X}_k\big) \ ,
\end{aligned}$$

where the equality

$$p_{j,k} = \sqrt{\mathrm{tr}(\mathbf{X}_{j,k})} = \sqrt{\mathrm{tr}\Big((p_j + p_k) \cdot \big(p_j^{-1}\mathbf{X}_j + p_k^{-1}\mathbf{X}_k\big)\Big)}$$
$$= \sqrt{(p_j + p_k) \cdot \big(p_j^{-1}\mathrm{tr}(\mathbf{X}_j) + p_k^{-1}\mathrm{tr}(\mathbf{X}_k)\big)}$$
$$= \sqrt{(p_j + p_k) \cdot (p_j + p_k)} = p_j + p_k$$

has been employed. Thus, the matrix (19) does not need to be computed at once.

These results directly imply that each sensor node i can autonomically compute the parameter

$$p_i = \sqrt{\mathrm{tr}\Big((\mathbf{H}_k^i)^{\mathrm{T}}(\mathbf{C}_k^{z,i})^{-1}\mathbf{X}_k^{z,i}(\mathbf{C}_k^{z,i})^{-\mathrm{T}}\mathbf{H}_k^i\Big)}$$

corresponding to its contribution (16). In order to simplify the representation of (16), let

$$\underline{i}_k^i := (\mathbf{H}_k^i)^{\mathrm{T}}(\mathbf{C}_k^{z,i})^{-1}\underline{\hat{z}}_k^i ,$$
$$\mathbf{Z}_k^i := (\mathbf{H}_k^i)^{\mathrm{T}}(\mathbf{C}_k^{z,i})^{-1}\mathbf{X}_k^{z,i}(\mathbf{C}_k^{z,i})^{-\mathrm{T}}\mathbf{H}_k^i .$$

A receiving node j on the communication path updates its parameters according to

$$\underline{i}_k^{\{i,j\}} = \underline{i}_k^i + \underline{i}_k^j ,$$
$$\mathbf{I}_k^{\{i,j\}} = \mathbf{I}_k^i + \mathbf{I}_k^j , \qquad (21)$$
$$\mathbf{Z}_k^{\{i,j\}} = (p_i + p_j)\big(p_i^{-1}\mathbf{Z}_k^i + p_j^{-1}\mathbf{Z}_k^j\big) ,$$

before transmitting them to the next node on the path. The first two parameters correspond to the standard information filter, whereas the latter parameter is required for the ellipsoid. In the data sink, these parameters update the predicted information state by means of

$$\underline{\hat{y}}_k^{\mathrm{e}} = \underline{\hat{y}}_k^{\mathrm{p}} + \underline{i}_k^{\{1,\ldots,M\}} ,$$
$$\mathbf{Y}_k^{\mathrm{e}} = \mathbf{Y}_k^{\mathrm{p}} + \mathbf{I}_k^{\{1,\ldots,M\}} , \qquad (22)$$
$$\mathbf{Q}_k^{\mathrm{e}} = \big(p^{\mathrm{p}} + p_{\{1,\ldots,M\}}\big)\big((p^{\mathrm{p}})^{-1}\mathbf{Q}_k^{\mathrm{p}} + p_{\{1,\ldots,M\}}^{-1}\mathbf{Z}_k^{\{1,\ldots,M\}}\big)$$

with $p^{\mathrm{p}} := \sqrt{\mathrm{tr}(\mathbf{Q}_k^{\mathrm{p}})}$. Eventually, the computation of the shape matrix $\mathbf{Q}_k^{\mathrm{e}}$ corresponds to (19) and is optimal with respect to the trace.

B. Optimality in State Space

The procedure from the preceding subsection only provides a trace-optimal ellipsoid in the information space and therefore does not optimize the trace of

$$\mathbf{X}_k^{\mathrm{e}} = (\mathbf{Y}_k^{\mathrm{e}})^{-1}\mathbf{Q}_k^{\mathrm{e}}(\mathbf{Y}_k^{\mathrm{e}})^{-\mathrm{T}}$$

for the ellipsoid

$$\mathcal{E}(\underline{\hat{x}}_k^{\mathrm{e}}, \mathbf{X}_k^{\mathrm{e}}) = (\mathbf{Y}_k^{\mathrm{e}})^{-1}\mathcal{E}(\underline{\hat{y}}_k^{\mathrm{e}}, \mathbf{Q}_k^{\mathrm{e}})$$
$$= \mathcal{E}\big((\mathbf{Y}_k^{\mathrm{e}})^{-1}\underline{\hat{y}}_k^{\mathrm{e}}, (\mathbf{Y}_k^{\mathrm{e}})^{-1}\mathbf{Q}_k^{\mathrm{e}}(\mathbf{Y}_k^{\mathrm{e}})^{-\mathrm{T}}\big)$$

converted back to the state space. This is due to the inequality

$$\mathrm{tr}(\mathbf{X}) \neq \mathrm{tr}(\mathbf{A}\mathbf{X}\mathbf{A}^{\mathrm{T}}) .$$

For the purpose of minimizing the trace of $(\mathbf{Y}_k^{\mathrm{e}})^{-1}\mathbf{Q}_k^{\mathrm{e}}(\mathbf{Y}_k^{\mathrm{e}})^{-\mathrm{T}}$ instead of $\mathbf{Q}_k^{\mathrm{e}}$, we can recognize from (21), (22), and the relationship

$$(\mathbf{Y}_k^{\mathrm{e}})^{-1}\mathbf{Q}_k^{\mathrm{e}}(\mathbf{Y}_k^{\mathrm{e}})^{-\mathrm{T}} = \big(p^{\mathrm{p}} + p_{\{1,\ldots,M\}}\big) \cdot \big((p^{\mathrm{p}})^{-1}\cdot$$
$$(\mathbf{Y}_k^{\mathrm{e}})^{-1}\mathbf{Q}_k^{\mathrm{p}}(\mathbf{Y}_k^{\mathrm{e}})^{-\mathrm{T}} + p_{\{1,\ldots,M\}}^{-1}(\mathbf{Y}_k^{\mathrm{e}})^{-1}\mathbf{Z}_k^{\{1,\ldots,M\}}(\mathbf{Y}_k^{\mathrm{e}})^{-\mathrm{T}}\big)$$

that the parameters $p^{\mathrm{p}}, p_1, \ldots, p_M$ have to be replaced by

$$p^{\mathrm{p}} = \sqrt{\mathrm{tr}\big((\mathbf{Y}_k^{\mathrm{e}})^{-1}\mathbf{Q}_k^{\mathrm{p}}(\mathbf{Y}_k^{\mathrm{e}})^{-\mathrm{T}}\big)} ,$$
$$p^i = \sqrt{\mathrm{tr}\big((\mathbf{Y}_k^{\mathrm{e}})^{-1}\mathbf{Z}_k^i(\mathbf{Y}_k^{\mathrm{e}})^{-\mathrm{T}}\big)} , \quad i \in \{1, \ldots, M\} . \qquad (23)$$

Evidently, each node i must be aware of the estimated information matrix $\mathbf{Y}_k^{\mathrm{e}}$ in order to compute its parameter p_i. This can be achieved by first transmitting the matrices $\mathbf{I}_k^i$ and computing $\mathbf{Y}_k^{\mathrm{e}}$ in the data sink and thereafter broadcasting the information matrix $\mathbf{Y}_k^{\mathrm{e}}$ back to the fusion nodes so that the parameters (23) can be computed locally. With the altered parameters (23), the shape matrices $\mathbf{Z}_k^i$ can then be communicated and fused with $\mathbf{Q}_k^{\mathrm{p}}$ according to (22). In summary, the procedure is

1) first to communicate the information matrices $\mathbf{I}_k^i$ to the data sink, where $\mathbf{Y}_k^{\mathrm{e}}$ is computed.
2) Then, $\mathbf{Y}_k^{\mathrm{e}}$ is send back to each fusing sensor node that needs to know $\mathbf{Y}_k^{\mathrm{e}}$ in order to determine the parameters (23).
3) The parameters p_i can then be locally determined and the matrices $\mathbf{Z}_i$ are transmitted to the data sink according to (21). The data sink computes $\mathbf{Q}_k^{\mathrm{e}}$ by means of (22).
4) In state space, the estimate is given by midpoint $\underline{\hat{x}}_k^{\mathrm{e}} = (\mathbf{Y}_k^{\mathrm{e}})^{-1}\underline{\hat{y}}_k^{\mathrm{e}}$, shape matrix $\mathbf{X}_k^{\mathrm{e}} = (\mathbf{Y}_k^{\mathrm{e}})^{-1}\mathbf{Q}_k^{\mathrm{e}}(\mathbf{Y}_k^{\mathrm{e}})^{-\mathrm{T}}$, and covariance matrix $\mathbf{C}_k^{\mathrm{e}} = (\mathbf{Y}_k^{\mathrm{e}})^{-1}$.

The only additional communication load consists of the transmission of $\mathbf{Y}_k^{\mathrm{e}}$. The communication and computation of $\underline{\hat{y}}_k^{\mathrm{e}}$ can be done either in step 1) or step 3).

In spite of the additional transmission of $\mathbf{Y}_k^{\mathrm{e}}$, this procedure is still more favorable than to directly employ the Kalman filtering formulation from Sec. II. By considering the case of two measurement devices and a sequential processing, an enclosing ellipsoid for

$$(\mathbf{I} - \mathbf{K}_k^2\mathbf{H}_k^2)\Big((\mathbf{I} - \mathbf{K}_k^1\mathbf{H}_k^1)\mathcal{E}(\underline{\hat{x}}_k^{\mathrm{p}}, \mathbf{X}_k^{\mathrm{p}}) \oplus \mathbf{K}_k^1\mathcal{E}(\underline{\hat{z}}_k^1, \mathbf{X}_k^{z,1})\Big)$$
$$\oplus \mathbf{K}_k^2\mathcal{E}(\underline{\hat{z}}_k^2, \mathbf{X}_k^{z,2}) \subseteq \mathcal{E}(\underline{\hat{x}}_k^{\mathrm{e}}, \mathbf{X}_k^{\mathrm{e}})$$

has to be calculated. Approximating first the inner sum optimally and afterwards the outer sum would not provide a trace-optimal result. It becomes apparent that each shape matrix requires an individual linear transformation before a trace-optimal fusion result can be determined. In the above scenario, this is $(\mathbf{I} - \mathbf{K}_k^2\mathbf{H}_k^2)\mathbf{K}_k^1$ for the second ellipsoid and $(\mathbf{I} - \mathbf{K}_k^2\mathbf{H}_k^2)\mathbf{K}_k^2$ for the third ellipsoid. For the information filter, each fusing node only needs to receive the same $\mathbf{Y}_k^{\mathrm{e}}$ to compute the optimal parameter according to (23). These issues are further addressed in the subsequent simulated scenario.

V. SIMULATIONS

To show that adding few additional network transfers enables the information form to create better results not only in constructed cases but also under realistic circumstances a tracking scenario has been simulated. The target object to be tracked runs in a circle with a constant velocity for 19 steps and changes its angle and speed for 16 further time-steps. For the real system we have used

$$\mathbf{A} = \mathbf{I} \ , \qquad \mathbf{B}_k = \begin{bmatrix} \cos(\gamma) & -\sin(\gamma) \\ \sin(\gamma) & \cos(\gamma) \end{bmatrix}$$

with $\underline{u}_1 = \begin{bmatrix} 0.05 & 0 \end{bmatrix}^{\mathrm{T}}$, $\underline{u}_k = \underline{x}_k - \underline{x}_{k-1}$ for $k > 2$, $\gamma = \frac{\pi}{20}$ for $k < 20$ and $\gamma = \frac{3\pi}{40}$ and afterwards. For system model of our estimator we have used $\gamma = \frac{\pi}{10}$ to induce system noise and computed $\hat{\underline{u}}$ using the past estimates via $\hat{\underline{u}}_k = \hat{\underline{x}}_k^{\mathrm{e}} - \hat{\underline{x}}_{k-1}^{\mathrm{e}}$.

The network observing the phenomenon consists of a total of 36 sensors in 9 clusters of 4 sensors each. In each cluster only a single expensive node is capable of communicating to a master node via long distance communication. The cheap nodes have to send measurements as well as information about stochastic and set-membership uncertainty to the expensive nodes and rely on them to fuse or further propagate their information. Whereas only the master node is keeping track of the state the expensive nodes are able to fuse the measurements of the cheap nodes and their own as the measurements are fully dimensional with $\mathbf{H} = \mathbf{I}$.

The need to transfer covariance and shape matrices arises as they are varying depending on the relative position of the node to the sensor. A stochastic noise is assumed and simulated orthogonal to the vector between sensor and target, whereas set-membership uncertainty stemming from a discretization is simulated parallel to the connecting vector. The measurement noises were set as an over-approximation of the applied noise and depend on the distance between $\overline{ts}$ target $\underline{t}$ and sensor $\underline{s}$ and the discretization levels Δd:

$$\mathbf{C}^{\mathrm{v}} = \frac{1}{100} \lceil 10\overline{ts} \rceil \, \mathbf{R} \begin{bmatrix} 1 & 0 \\ 0 & 0.01 \end{bmatrix} \mathbf{R}^{\mathrm{T}} \ ,$$

$$\mathbf{X}^{\mathrm{v}} = \mathbf{R} \begin{bmatrix} 0.01 & 0 \\ 0 & (\frac{\Delta d}{2})^2 \end{bmatrix} \mathbf{R}^{\mathrm{T}}$$

with $\mathbf{R}$ denoting a rotation matrix that causes above described orientation of the uncertainties. The uncertainties are shown on a sample point in Fig. 2.

First the amount of necessary matrix transfers – counting both the transfer of the covariance and shape matrix – need to be considered. The minimal approach that consists of only one transfer of both matrices to the next hop requires 72 matrix transfers. It is the amount of communication required to fuse the measurements of the cheap nodes at the expensive nodes and then transfer the fused result to the master node. This case will be compared to the optimal case in information filter form: this requires 9 additional matrix transfers (a total of 81) as the master node needs to transfer the fused covariance matrix back to the expensive nodes in order to allow them to fuse the ellipsoids optimally. A naive approach to fuse
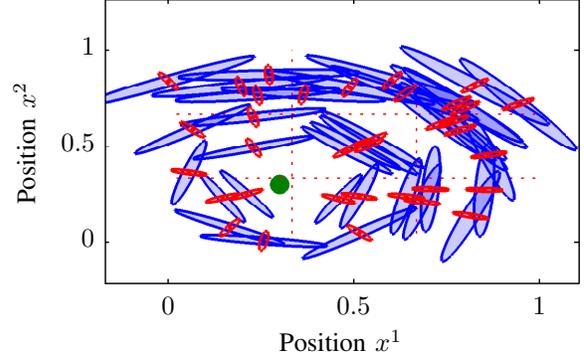

Figure 2.   Uncertainties of each sensor for a sample point. Sample point in green, stochastic uncertainty in blue, and set-membership uncertainty in red.
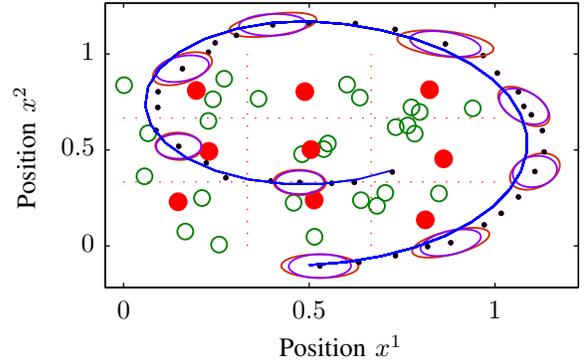

Figure 3.   Network with cheap nodes in green and expensive nodes in red. The real path is drawn in blue, the estimates in black, and the set-membership uncertainty is depicted for every forth step in brown-red for the state-space result and in purple for the optimal information-form result.

information optimally using the regular form from Sec. II by simply sending all covariance and shape matrices to the master node requires a total of 126 matrix transfers.

Secondly, the efficiency of the communication-minimal state-space approach shall be compared to the optimal information-form approach. The additional uncertainty of the state-space formulation is caused by gains of further steps altering the shape of the ellipsoid, retroactively worsening the approximation of the previous step, as discussed in Sec. IV-B. Thus we generally have to expect a worse over-approximation compared to the optimal form in case of higher variety of covariance matrices. At the center the variation of the scale of the covariance matrix is lower than at the side of the environment, causing the gap to generally be smaller when the object is located near the center. The network, the real and estimated trajectory and the set-membership uncertainty are shown in Fig. 3.

The evaluation confirmed the above assumptions and also shows a major point, that is shown clearly in Fig. 4: whereas the information form is only stepwise optimal it performs constantly better over time under reasonable and varying circumstances.
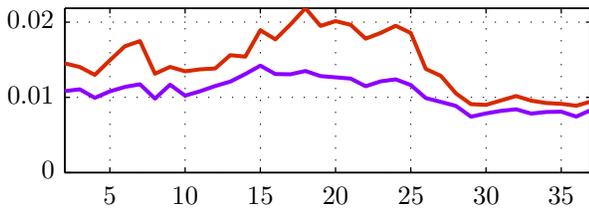
Figure 4. Traces of the ellipsoidmatrices in every time-step. In brown-red the traces for the state-space solution and in purple for the optimal information-form solution.

VI. SUMMARY AND CONCLUSIONS

For the simultaneous treatment of stochastic and unknown but bounded uncertainties, the Kalman filtering scheme has been generalized in [8], [12] to operate on sets of Gaussian densities instead of single densities. Each of these sets has been parameterized by an ellipsoid of means and a covariance matrix. The Kalman filter can then be formulated in terms of the midpoint, the shape matrix, and the covariance matrix. For distributed estimation problems, the single-valued Kalman filter can be transformed into an inverse covariance matrix formulation, i.e., into the information form, which significantly simplifies distributed collection and processing of measurement data. In this paper, an information form for ellipsoidal sets of means has been derived. For this purpose, the ellipsoids of means are mapped to ellipsoids of information vectors. The filtering step of this set-valued information filter constitutes a simple Minkowski sum of the corresponding ellipsoids. Although an outer approximation needs to be determined and optimized for the Minkowski sum, we have shown that a distributable computation is still possible if the trace of the estimated shape matrix is considered as the optimality criterion. Since a trace-optimal enclosing ellipsoid of the possible state estimates and not of the information vectors is desired, we have also proposed a method to achieve optimality in state space. Therefore, the estimated covariance matrix is computed first and broadcasted back to all fusing sensor node. Then, each node is able to compute the optimal weighting parameter for the estimated shape matrix of the external approximation. The proposed method has been evaluated in a simulation and compared with the Kalman filtering scheme.

Prospective work will also include nonlinear versions by following the directions of [9], [12], and [21]. With the information filter, also more sophisticated fusion architectures can be implemented, which avoid the need for a central data sink. For this, we are often required to remove common information between local estimates. Accordingly, it is interesting to formulate the removal of common information in terms of ellipsoidal sets of information vectors.

VII. ACKNOWLEDGMENTS

REFERENCES

[1] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Transactions of the ASME - Journal of Basic Engineering*, pp. 35–45, 1960.

[2] F. C. Schweppe, *Uncertain Dynamic Systems*. Prentice-Hall, 1973.

[3] F. L. Chernousko, *State Estimation for Dynamic Systems*. CRC Press, 1994.

[4] A. Kurzhanski and I. Vályi, *Ellipsoidal Calculus for Estimation and Control*. Birkhäuser, 1997.

[5] A. Kallapur, I. R. Petersen, and S. Anavatti, "A Discrete-time Robust Extended Kalman Filter," in *Proceedings of the 2009 American Control Conference (ACC 2009)*, St. Louis, Missouri, USA, Jun. 2009.

[6] I. R. Petersen, M. R. James, and P. Dupuis, "Minimax Optimal Control of Stochastic Uncertain Systems with Relative Entropy Constraints," *IEEE Transactions on Automatic Control*, vol. 45, pp. 398–412, 2000.

[7] T. Henningsson, "Recursive State Estimation for Linear Systems with Mixed Stochastic and Set-Bounded Disturbances," in *Proceedings of the 47th IEEE Conference on Decision and Control (CDC 2008)*, Cancún, Mexico, Dec. 2008.

[8] D. R. Morrell and W. C. Stirling, "Set-Valued Filtering and Smoothing," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 21, no. 1, pp. 184–193, Jan. 1991.

[9] D. R. Morrell and W. C. Stirling, "An Extended Set-Valued Kalman Filter," in *Proceedings of the Third International Symposium on Imprecise Probabilities and Their Applications (ISIPTA 2003)*, Lugano, Switzerland, Jul. 2003, pp. 395–405.

[10] J. Kenney and W. Stirling, "Nonlinear Filtering of Convex Sets of Probability Distributions," *Journal of Statistical Planning and Inference*, vol. 105, no. 1, pp. 123–147, 2002.

[11] B. Noack, V. Klumpp, D. Brunn, and U. D. Hanebeck, "Nonlinear Bayesian Estimation with Convex Sets of Probability Densities," in *Proceedings of the 11th International Conference on Information Fusion (Fusion 2008)*, Cologne, Germany, Jul. 2008, pp. 1–8.

[12] B. Noack, V. Klumpp, and U. D. Hanebeck, "State Estimation with Sets of Densities considering Stochastic and Systematic Errors," in *Proceedings of the 12th International Conference on Information Fusion (Fusion 2009)*, Seattle, Washington, Jul. 2009.

[13] A. Benavoli, M. Zaffalon, and E. Miranda, "Robust Filtering Through Coherent Lower Previsions," *IEEE Transactions on Automatic Control*, vol. 56, no. 7, pp. 1567–1581, Jul. 2011.

[14] U. D. Hanebeck and J. Horn, "Fusing Information Simultaneously Corrupted by Uncertainties with Known Bounds and Random Noise with Known Distribution," *Information Fusion, Elsevier Science*, vol. 1, no. 1, pp. 55–63, 2000.

[15] A. E. H. Gning, B. Ristic, and L. Mihaylova, "A Box Particle Filter for Stochastic and Set-theoretic Measurements with Association Uncertainty," in *Proceedings of the 14th International Conference on Information Fusion (Fusion 2011)*, Chicago, Illinois, USA, Jul. 2011.

[16] S. Grime and H. Durrant-Whyte, "Data Fusion in Decentralized Sensor Networks," *Control Engineering Practice*, vol. 2, no. 5, pp. 849–863, 1994.

[17] A. G. O. Mutambara, *Decentralized Estimation and Control for Multisensor Systems*. Boca Raton, Florida, USA: CRC Press, Inc., 1998.

[18] H. F. Durrant-Whyte, "A Beginners Guide to Decentralised Data Fusion," Australian Centre Field Robotics, University of Sydney, Sydney, Australia, Tech. Rep., 2006.

[19] B. Noack, D. Lyons, M. Nagel, and U. D. Hanebeck, "Nonlinear Information Filtering for Distributed Multisensor Data Fusion," in *Proceedings of the 2011 American Control Conference (ACC 2011)*, San Francisco, California, USA, Jun. 2011.

[20] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation: Theory, Algorithms, and Software*. John Wiley & Sons, 2001.

[21] B. Noack, V. Klumpp, N. Petkov, and U. D. Hanebeck, "Bounding Linearization Errors with Sets of Densities in Approximate Kalman Filtering," in *Proceedings of the 13th International Conference on Information Fusion (Fusion 2010)*, Edinburgh, United Kingdom, Jul. 2010.