Optimal Distributed Combined Stochastic and Set-Membership State Estimation

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

***Abstract*—For distributed estimation, algorithms have to be specifically crafted to minimize communication between the sensor nodes. As an adjusted version of the regular Kalman filter, the distributed Kalman filter (DKF) allows for deriving optimal results while not requiring regular communication. To achieve this, the DKF requires that each node has full knowledge about the system model and measurement models of all nodes. However, the DKF is not sufficient if the characteristics of the errors in the system and measurement models are not purely stochastic. In this paper, we present a distributed version of a combined stochastic and set-membership Kalman filter. The proposed filter optimizes the approximations of the set-membership uncertainties and can even yield better results than the regular centralized filter.**

***Index Terms*—Set-membership uncertainty, distributed Kalman filter, recursive Bayesian estimation**

I. Introduction

Frequently used Bayesian estimation approaches rely on precise knowledge about the system and measurement models and all uncertainties involved in each time step. While the required information may be available in simulations, stochastic models used in real-world scenarios are generally only approximations. These approximate models can lead to overconfidence in one's estimation results, an outcome that has to be avoided at all costs for security-relevant infrastructures.

An alternative to a purely stochastic approach is to use set-membership uncertainties [1], [2]. In particular, ellipsoids [3], [4], [5] are frequently used in set-membership state estimation. While set-membership state estimation is a long known and well-studied approach, it does not allow for proper modeling of stochastic variability and outliers. First, treating all uncertainties as set-membership uncertainties is prone to cause overly conservative estimates. Second, if a stochastic uncertainty with an infinite support is approximated using a bounded set, the filter may yield an invalid estimate in the form of an empty set.

To alleviate these issues, filters respecting both stochastic and set-membership uncertainties have been proposed. Approaches dealing with sets of densities include a Kalman filter based on credal sets [6] and a combined stochastic and set-membership Kalman filter based on ellipsoidal sets [7]. In the latter, stochastic uncertainties are described by the mean and covariance, just as in the regular Kalman filter. When using the ellipsoid-based combined stochastic and set-membership Kalman filter of [7] with the standard Kalman gain, the stochastic uncertainty is minimized. In this case, there are formulae for the combined filter that have the same complexity as the regular Kalman filter, just with more matrix operations to calculate in each step. In [8], we proposed another linear filter based on ellipsoids whose gain does not result in a minimization of the stochastic uncertainty but rather in a minimization of a criterion that takes both types of uncertainties into account.

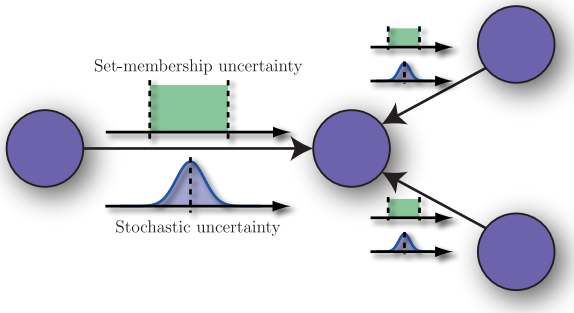

Figure 1. Sensor network with four nodes. Our algorithm minimizes the stochastic uncertainty while keeping the set-membership uncertainty as low as possible.

The combined stochastic and set-membership approaches can be applied to (among others) problems involving discretization, human reports, or negative information [9]. Furthermore, the approaches are well suited to extract information from non-transmitted measurements in event-based state estimation [10].

Due to the set operations involved, an approximation error in the set-membership uncertainty is often inevitable regardless of the gain. First work to minimize approximation errors for multiple sequential operations was done in the context of estimation in sensor networks. In [11], we proposed a variant that allows us to keep the approximation error in a distributed scenario at the same level as when batch processing is performed by a centralized filter. However, it is assumed that there is one central fusing node to which preprocessed data are transferred at each time step. Only the fusing node performs prediction steps. Thus, while multiple filter steps at one time step can be performed optimally in a distributed fashion, regular communication is required and information at the individual nodes cannot be propagated individually through prediction steps.

In [12], the authors describe an approach that only supports stochastic uncertainties that we refer to as the distributed Kalman filter (DKF). The DKF requires that the measurement models of all sensors and the points in time at which measurements are obtained are known by each individual node. Given this information, the DKF can produce the optimal estimate in sensor networks when the fusing node has the current information of all nodes at its disposal at an arbitrary time step. Further, no communication is required prior to the time step at which the information from the multiple sensors is combined.

In this paper, we extend the idea of the DKF to the combined stochastic and set-membership Kalman filter to handle scenarios as the one shown in Fig. 1. We propose a filter that allows us to combine the information of the individual sensors to derive the result with minimal covariance and as little approximation error in the set-membership uncertainty as possible. In case the time step of the fusion is known ahead of time, the approximation of the set-membership uncertainty obtained is the best among a common family of approximations. If the future time step at which the fusion is supposed to occur is not known, the filter still produces valid results that are conservative with regard to the set-membership uncertainty. By conservative, we mean that the uncertainty provided by the filter is never too small. In the context of set-membership uncertainties, the set must at least contain all elements that would be contained if no approximations were performed.

Further related work includes [13], [14], in which the fusion of distributed sensor information featuring merely set-membership uncertainties is regarded. In [15], a filter is presented that is independent of the sequence in which sensor measurements are included in the estimate. However, this filter only yields a suboptimal approximation.

In the next section, we first provide a brief introduction to the combined stochastic and set-membership Kalman filter. Afterward, we give a formulation of the DKF for purely stochastic uncertainties. As the basis for our explanation, we use the IDKF [16], an information filter [17] form of the DKF, which facilitates the formulation of our filter that we present in the fourth section. We evaluate our approach against other centralized and distributed approaches in the fifth section. In the last section, we provide a conclusion and an outlook.

II. Combined Stochastic and Set-Membership Kalman Filter Based on Ellipsoids

In the regular Kalman filter, the prior and posterior densities are described by a mean vector $\hat{\underline{x}}$ and a covariance matrix $\mathbf{C}$. The mean of the state $\hat{\underline{x}}$, the inputs $\hat{\underline{u}}$, and the measurements $\hat{\underline{z}}$ can be written as a single vector. For the combined stochastic and set-membership Kalman filter, there are sets of potential means, inputs, and measurements. The current knowledge about the scenario at hand and all uncertainties are described by a covariance matrix $\mathbf{C}$ and a set of possible means $\mathcal{X}$. Unlike in purely set-membership state estimation, the true state is not guaranteed to lie in $\mathcal{X}$. The vectors in $\mathcal{X}$ rather represent possible means of, e.g., Gaussian distributions describing the distribution of the state.

In the first subsection of this section, we give some basics of ellipsoidal calculus required to understand and implement the set operations involved. In the second and third subsection, we explicate the prediction and filter step of the combined stochastic and set-membership Kalman filter without any extensions as proposed in [7].

A. Ellipsoidal Calculus

In the literature, a variety of conventions regarding the parametrization of an arbitrary-dimensional ellipsoid exist. However, all common conventions describe the ellipsoid using a center vector $\underline{c}$ and a symmetric positive semi-definite shape matrix $\mathbf{Q}$. We use the convention to define the ellipsoidal set $\mathcal{E}(\underline{c}, \mathbf{Q})$ as

$$\mathcal{E}(\underline{c}, \mathbf{Q}) = \{\underline{x} \,|\, (\underline{c} - \underline{x})^T \mathbf{Q}^{-1} (\underline{c} - \underline{x}) \leq 1\} \ .$$

An affine transformation comprising a linear transformation of each individual element of the set $\mathcal{E}(\underline{c}, \mathbf{Q})$ according to a matrix $\mathbf{A}$ and a translation according to a vector $\underline{b}$ can be performed by using simple operations on the parameters of the ellipsoid. For any matrix $\mathbf{A}$ and vector $\underline{b}$,

$$\mathbf{A}\mathcal{E}(\underline{c}, \mathbf{Q}) + \underline{b} = \mathcal{E}(\mathbf{A}\underline{c} + b, \mathbf{A}\mathbf{Q}\mathbf{A}^T) \tag{1}$$

holds.

Another operation we require for our filter is the Minkowski sum. This operation yields the set that encompasses all elements that are obtained when adding each element of one set to every element in the other. While affine transformations of an ellipsoid always yield an ellipsoid, the Minkowski sum of two or more ellipsoids is not an ellipsoid in general. However, there are ways to find tight[1] bounding ellipsoids for the Minkowski sum of an arbitrary number of ellipsoids. In the entire paper, we only regard an important family of external approximations of the Minkowski sum described by [18]

$$\mathbf{Q}^{\text{mink}} = \sum_{i=1}^{N} \alpha_i^{-1} \mathbf{Q}_i \text{ with } \alpha_i > 0 \text{ and } \sum_{i=1}^{N} \alpha_i = 0$$

depending on the shape matrices $\mathbf{Q}_1, \ldots, \mathbf{Q}_N$ of the ellipsoids of which we build the Minkowski sum. Among this family, it is easy to obtain the approximation of minimal size in regard to the sum of the squared lengths of the semiaxes, which corresponds to the lowest trace of the shape matrix. Among this family, the shape matrix $\mathbf{Q}^{\text{mink}}$ of the trace-minimal approximation can be calculated according to [3, Ch. 2.7], [18, Theorem 4.4]

$$\mathbf{Q}^{\text{mink}} = \left(\sum_{i=1}^{N} p_i \right) \sum_{i=1}^{N} \frac{1}{p_i} \ \mathbf{Q}_i \qquad p_i = \sqrt{\text{tr}(\mathbf{Q}_i)} \tag{2}$$

[1] An external approximation is called tight when it touches the border of the Minkowski sum and thus cannot be scaled down without losing the property that it is an external approximation.

and the center becomes

$$\underline{c}^{\text{mink}} = \sum_{i=1}^{N} \underline{c}_i \ . \tag{3}$$

As an external approximation,

$$\mathcal{E}(\underline{c}^{\text{mink}}, \mathbf{Q}^{\text{mink}}) \supseteq \bigoplus_{i=1}^{N} \mathcal{E}(\underline{c}_i, \mathbf{Q}_i)$$

holds. For our novel filter, it is important to note that calculating the trace-minimal approximation of the Minkowski sum of a set of N ellipsoids at once yields the same result as calculating the approximation sequentially. In other worlds, if we say that $\text{approx}_{\text{tr}}$ yields the approximation that is minimal in respect to the trace of the shape matrix, then

$$\begin{aligned} &\text{approx}_{\text{tr}}\left(\bigoplus_{i=1}^{N} \mathcal{E}(\underline{c}_i, \mathbf{Q}_i) \right) \\ &= \text{approx}_{\text{tr}}\Big(\ldots \text{approx}_{\text{tr}}\big(\mathcal{E}(\underline{c}_1, \mathbf{Q}_1) \oplus \mathcal{E}(\underline{c}_2, \mathbf{Q}_2)\big) \oplus \ldots \Big) \end{aligned} \tag{4}$$

holds.

B. Prediction Step

In scenarios featuring both stochastic and set-membership uncertainties, our system evolution for the random variable $\underline{\boldsymbol{x}}_k$ describing the state at time step k is given according to

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

with the known system input $\hat{\underline{u}}_k$, the stochastic system noise term $\underline{\boldsymbol{w}}_k$, and the set-membership noise term $\underline{d}_k$.

For the regular Kalman filter, the predicted covariance $\mathbf{C}_{k+1}^{p}$ is calculated from the covariance at time step k respecting all measurements up to time step k that we call $\mathbf{C}_k^e$, the system noise covariance $\mathbf{C}_k^w$ for the evolution from time step k to $k+1$, the system matrix $\mathbf{A}_k$, and the input matrix $\mathbf{B}_k$ according to

$$\mathbf{C}_{k+1}^{p} = \mathbf{A}_k \mathbf{C}_k^e \mathbf{A}_k^T + \mathbf{B}_k \mathbf{C}_k^w \mathbf{B}_k^T \ .$$

The mean $\hat{\underline{x}}_{k+1}^{p}$ of the predicted density is calculated from the current estimated mean $\hat{\underline{x}}_k^e$ according to

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

For the combined stochastic and set-membership Kalman filter, there is a set of possible means $\mathcal{X}_k^e$ instead of the single mean $\hat{\underline{x}}_k^e$. Likewise, only the set $\mathcal{U}_k$ in which the input lies is known. The location of $\mathcal{U}_k$ depends on the input $\hat{\underline{u}}_k$, whereas the extent of $\mathcal{U}_k$ is determined by the values that $\underline{d}_k$ can attain. All predicted means that can result from the prediction step based on any of the combinations of the means in $\mathcal{X}_k^e$ and inputs in $\mathcal{U}_k$ are contained in

$$\tilde{\mathcal{X}}_{k+1}^{p} = \mathbf{A}_k \mathcal{X}_k^e \oplus \mathbf{B}_k \mathcal{U}_k \ ,$$

which is a Minkowski sum of two ellipsoids that is not an ellipsoid in general.

To avoid an increase in the complexity of the algorithm, we perform an external approximation of the new set of prior means $\tilde{\mathcal{X}}_{k+1}^{p}$ using an ellipsoid. This new ellipsoid $\mathcal{X}_{k+1}^{p}$ must fulfill

$$\mathcal{X}_{k+1}^{p} \supseteq \mathbf{A}_k \mathcal{X}_k^e \oplus \mathbf{B}_k \mathcal{U}_k \ .$$

In the combined stochastic and set-membership Kalman filter, we choose to use the trace-minimal external approximation. Without any optimizations to minimize the approximation error over multiple time steps (as are presented in this paper), the center and shape matrix of $\mathcal{X}_{k+1}^{p}$ are calculated according to (2) and (3), which means that $\mathcal{X}_{k+1}^{p} = \text{approx}_{\text{tr}}(\tilde{\mathcal{X}}_{k+1}^{p})$ holds.

In the rest of our paper, we use the model

$$\mathcal{X}_{k+1}^{p} = \mathbf{A}_k \mathcal{X}_k^e \ ,$$

in which no Minkowski sum is involved. The system input (if any) is modeled implicitly as done in many filters such as the regular DKF. While we have also developed a version of our approach with explicit modeling of the system input and additional set-membership input uncertainty, it would require more elaborate explanations in the context of distributed estimation.

C. Filter Step

For measurements with both stochastic and set-membership uncertainties, our measurement model is given according to

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

with the stochastic measurement noise term $\underline{\boldsymbol{v}}_k$ and the set-membership measurement noise term $\underline{e}_k$.

The Kalman filter is a popular filter for filtering problems involving linear models and uncorrelated noise as it is the Linear Minimum Mean Squared Error estimator [19, Ch. 13] for these problems. The formula of the Kalman filter for updating the covariance matrix $\mathbf{C}_k^e$ can be given in Joseph form depending on the measurement matrix $\mathbf{H}_k$ and the measurement covariance matrix $\mathbf{C}_k^v$ according to

$$\mathbf{C}_k^e = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{C}_k^p (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)^T + \mathbf{K}_k \mathbf{C}_k^v \mathbf{K}_k^T$$

with the Kalman gain

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

The estimated mean $\hat{\underline{x}}_k^e$ respecting the new measurement $\hat{\underline{z}}_k$ is calculated from the predicted mean $\hat{\underline{x}}_k^p$ according to

$$\hat{\underline{x}}_k^e = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \hat{\underline{x}}_k^p + \mathbf{K}_k \hat{\underline{z}}_k \ . \tag{5}$$

When dealing with both stochastic and set-membership uncertainties, we have a set of predicted means $\mathcal{X}_k^p$ and only the set $\mathcal{Z}_k$ in which the actual measurement is contained is known. The shape of $\mathcal{Z}_k$ depends on the values that $\underline{e}_k$ can attain at the respective time step. To respect the set-membership uncertainties, we change the formula (5) analogously to the prediction step via the use of the Minkowski sum. Again, we perform an external approximation of

$$\tilde{\mathcal{X}}_k^e = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathcal{X}_k^p \oplus \mathbf{K}_k \mathcal{Z}_k$$

to provide the ellipsoid $\mathcal{X}_k^e$ to be used in the next step. In the combined stochastic and set-membership Kalman filter as introduced in [7], $\mathcal{X}_k^e = \text{approx}_{\text{tr}}(\tilde{\mathcal{X}}_k^e)$ is used.

III. Distributed Kalman Filter

When full rate communication is not available, the optimal fusion result that would be obtained by a centralized Kalman filter can be (given all knowledge about the measurement models of the other nodes) obtained using the distributed Kalman filter (DKF) described in [12]. The gist of the DKF is that if all information required to calculate the covariance matrices (system and measurement matrices; system and measurement covariance matrices) are known, local pseudotracks[2] can be generated in which all information is treated as if it was processed in a central node. When an arbitrary node obtains all of the pseudotracks in one time step, the globally optimal fusion result can be derived.

In the following, we use an information form of the distributed Kalman filter (IDKF) [16] (which is an adjusted version of the globalized likelihood form of the DKF) and assume, for simplicity, that all nodes are time-synchronized. While all operations given in [12] yield state space estimates, the equations of the IDKF always yield results in the information space, unless a transformation to state space is invoked explicitly. We build up on this form as performing update steps in information form facilitates the formulation of our novel filter for the combined stochastic and set-membership case.

In the information form [17] of the Kalman filter, we keep track of the inverse of the covariance matrix $\mathbf{Y}_k = \mathbf{C}_k^{-1}$ instead of the covariance matrix and use an information vector $\underline{\hat{y}}_k = \mathbf{C}_k^{-1}\underline{\hat{x}}_k$ instead of the mean $\underline{\hat{x}}_k$ in state space. The mean and covariance in the state space form can always be retrieved via $\underline{\hat{x}}_k = \mathbf{Y}_k^{-1}\underline{\hat{y}}_k$ and $\mathbf{C}_k = \mathbf{Y}_k^{-1}$.

A. Initialization of the IDKF

For the distributed calculation, we let each node s keep track of the information matrix $\mathbf{Y}_{k,\text{glob}}$ representing the global uncertainty and the local information vector $\underline{\hat{y}}_{k,s}$ that was processed in a globalized manner. Keeping track of the global covariance $\mathbf{Y}_{k,\text{glob}}$ instead of a local covariance allows us to perform all operations as if they were performed by a central node, as will become evident in the next subsections. In the IDKF,

$$\underline{\hat{y}}_{k,\text{glob}} = \sum_{s=1}^{N} \underline{\hat{y}}_{k,s} \tag{6}$$

holds, meaning that summing up all local information $\underline{\hat{y}}_{k,s}$ yields the globally optimal information vector $\underline{\hat{y}}_{k,\text{glob}}$ as would be obtained using a centralized Kalman filter.

For the initialization of $\underline{\hat{y}}_{k,s}$ and $\mathbf{Y}_{k,\text{glob}}$ using prior information, we assume that there is precisely one prior distribution, which is in line with the classical Bayesian approach. The information vector of the optimal centralized Kalman filter at the time of the initialization is

$$\underline{\hat{y}}^p_{0,\text{glob}} = (\mathbf{C}^p_0)^{-1}\underline{\hat{x}}^p_0 \ ,$$

[2]We do not refer to them as tracks to emphasize that they are not unbiased local estimates when regarded in insolation and only contain some information that will allow generating a globally optimal estimate.

when $\underline{\hat{x}}^p_0$ is the mean of the prior distribution and $\mathbf{C}^p_0$ is the covariance of the prior.

There are two intuitive ways to initialize the parameters of the filter that are in line with condition (6). One approach is to let all nodes keep a fraction of the global prior information vector. For example, all information vectors of the sensors could be initialized using $\underline{\hat{y}}^p_{0,s} = \frac{1}{N}\underline{\hat{y}}^p_{0,\text{glob}}$, which ensures that they sum up to $\underline{\hat{y}}^p_{0,\text{glob}}$. As an alternative, we can let one node keep track of the prior information, while the other nodes merely accumulate information from the measurements. In this case, the information vector $\underline{\hat{y}}^p_{0,m}$ at the initial time step 0 of the sensor m keeping track of the prior information is initialized via $\underline{\hat{y}}^p_{0,m} = \underline{\hat{y}}^p_{0,\text{glob}}$, while the information vectors of the other nodes $\underline{\hat{y}}^p_{0,s}$ with $s \neq m$ are initialized using zero vectors. Regardless of the approach, all sensors keep track of the global information matrix $\mathbf{Y}_{k,\text{glob}}$ that is initialized using the inverse of the covariance matrix of the prior $\mathbf{C}^p_0$.

B. Update Step of the IDKF

For a single measurement, the prior information matrix $\mathbf{Y}^p_k$ is updated in the information filter according to

$$\mathbf{Y}^e_k = \mathbf{Y}^p_k + \mathbf{J}_k \ , \qquad \mathbf{J}_k = \mathbf{H}^T_k(\mathbf{C}^v_k)^{-1}\mathbf{H}_k$$

and the information vector $\underline{\hat{y}}^p_k$ is updated using the formulae

$$\underline{\hat{y}}^e_k = \underline{\hat{y}}^p_k + \underline{\hat{j}}_k \ , \qquad \underline{\hat{j}}_k = \mathbf{H}^T_k(\mathbf{C}^v_k)^{-1}\underline{\hat{z}}_k \ .$$

To incorporate N uncorrelated measurements in one time step in central processing, the information filter simply uses a sum with additional terms. If we use an additional index s to indicate which measurement is referred to, the formulae become

$$\mathbf{Y}^e_k = \mathbf{Y}^p_k + \sum_{s=1}^{N} \mathbf{J}_{k,s} \ , \qquad \mathbf{J}_{k,s} = \mathbf{H}^T_{k,s}(\mathbf{C}^v_{k,s})^{-1}\mathbf{H}_{k,s}$$

for the information matrix and

$$\underline{\hat{y}}^e_k = \underline{\hat{y}}^p_k + \sum_{s=1}^{N} \underline{\hat{j}}_{k,s} \ , \qquad \underline{\hat{j}}_{k,s} = \mathbf{H}^T_{k,s}(\mathbf{C}^v_{k,s})^{-1}\underline{\hat{z}}_{k,s}$$

for the update of the information vector that integrates all measurements $\underline{\hat{z}}_{k,s}$.

The update step of the IDKF can be thought of as splitting up the information vector $\underline{\hat{y}}_k$ into multiple information vectors $\underline{\hat{y}}_{k,s}$ and accumulating only information $\underline{\hat{j}}_{k,s}$ that can be calculated from the measurements $\underline{\hat{z}}_{k,s}$ at the respective sensor node. The $\underline{\hat{y}}_{k,s}$ are updated according to

$$\underline{\hat{y}}^e_{k,s} = \underline{\hat{y}}^p_{k,s} + \underline{\hat{j}}_{k,s} \quad \text{and} \quad \underline{\hat{j}}_{k,s} = \mathbf{H}^T_{k,s}(\mathbf{C}^v_{k,s})^{-1}\underline{\hat{z}}_{k,s} \ . \tag{7}$$

As we can see, no measurements obtained by other nodes are required by any node to keep track of its part of the information vector. If we update the information like this, it is ensured that (6) still holds after the filter step. A proof for this is provided in [16].

Using the sum of the individual information vectors $\underline{\hat{y}}^e_{k,s}$, the state-space estimate can be calculated via $(\mathbf{Y}^e_{k,\text{glob}})^{-1}\underline{\hat{y}}^e_{k,\text{glob}}$. Since

$$\hat{\underline{x}}^e_{k,\text{glob}} = (\mathbf{Y}^e_{k,\text{glob}})^{-1} \sum_{s=1}^{N} \hat{\underline{y}}^e_{k,s} = \sum_{s=1}^{N} \underbrace{(\mathbf{Y}^e_{k,\text{glob}})^{-1} \hat{\underline{y}}^e_{k,s}}_{:=\hat{\underline{x}}^e_{k,s}}$$

holds, the $\hat{\underline{x}}^e_{k,s}$ can be transferred instead and the fusion of the $\hat{\underline{x}}^e_{k,s}$ provided by the multiple sensors can be performed (in a bit counterintuitive manner) simply using a summation. As all nodes have full information about the system model and the measurement models of all sensors, they can calculate $\mathbf{Y}^e_{k,\text{glob}}$ from $\mathbf{Y}^p_{k,\text{glob}}$ without requiring information to be transmitted from other sensors. The matrix $\mathbf{Y}^e_{k,\text{glob}}$ is not only required to calculate $\hat{\underline{x}}^e_{k,\text{glob}}$ from $\hat{\underline{y}}^e_{k,\text{glob}}$ but also for the prediction step explained in the next subsection.

C. Prediction Step of the IDKF

In the prediction step, the information of the individual sensors needs to be propagated to the next time step while ensuring that property (6) is maintained. To perform the prediction in a simple way, we temporarily transform the information to state space. We first regard how the information would be processed in a centralized prediction step of a node that has accumulated all information and thus has $\hat{\underline{y}}^e_{k,\text{glob}}$ and $\mathbf{Y}^e_{k,\text{glob}}$ at its disposal. Afterward, we explicate how the calculation can be performed in a distributed fashion.

Let us first assume we are to perform a centralized prediction with $\hat{\underline{y}}^e_{k,\text{glob}}$ and $\mathbf{Y}^e_{k,\text{glob}}$ at our disposal. Then, the prediction step for the information vector and matrix can be performed according to

$$\begin{aligned} \hat{\underline{y}}^p_{k+1,\text{glob}} &= \mathbf{Y}^p_{k+1,\text{glob}} \mathbf{A}_k (\mathbf{Y}^e_{k,\text{glob}})^{-1} \hat{\underline{y}}^e_{k,\text{glob}} \ , \\ \mathbf{Y}^p_{k+1,\text{glob}} &= (\mathbf{A}_k (\mathbf{Y}^e_{k,\text{glob}})^{-1} \mathbf{A}_k^T + \mathbf{C}_k^w)^{-1} \ . \end{aligned} \tag{8}$$

Let us now regard how the formula for $\hat{\underline{y}}^p_{k+1,\text{glob}}$ can be rewritten depending on the individual local information $\hat{\underline{y}}^e_{k,s}$. Using (6), we obtain

$$\hat{\underline{y}}^p_{k+1,\text{glob}} = \mathbf{Y}^p_{k+1,\text{glob}} \mathbf{A}_k (\mathbf{Y}^e_{k,\text{glob}})^{-1} \sum_{s=1}^{N} \hat{\underline{y}}^e_{k,s} \ , \tag{9}$$

which is equivalent to

$$\hat{\underline{y}}^p_{k+1,\text{glob}} = \sum_{s=1}^{N} \underbrace{\mathbf{Y}^p_{k+1,\text{glob}} \mathbf{A}_k (\mathbf{Y}^e_{k,\text{glob}})^{-1} \hat{\underline{y}}^e_{k,s}}_{:=\hat{\underline{y}}^p_{k+1,s}} \ .$$

From the latter form, we can see that we can propagate each summand individually through the prediction step while maintaining the property that the global information vector $\hat{\underline{y}}^p_{k+1,\text{glob}}$ can be calculated from the sum of the individual vectors $\hat{\underline{y}}^p_{k+1,s}$. We can see, however, that the calculation of $\hat{\underline{y}}^p_{k+1,s}$ requires both $\mathbf{Y}^e_{k,\text{glob}}$ and $\mathbf{Y}^p_{k+1,\text{glob}}$. As stated in Sec. III-B, the individual nodes can calculate $\mathbf{Y}^e_{k,\text{glob}}$. As they have all information required for (8), all nodes are also able to calculate $\mathbf{Y}^p_{k+1,\text{glob}}$.

Thus, the nodes can calculate and keep track of only their part of the global information while preserving the property that the fusion can be performed as given in (6). As the filter and prediction steps preserve this property and because all information is multiplied by the same matrices and accumulated (albeit distributedly), the globally optimal estimate can be derived in every time step.

IV. Distributed Combined Stochastic and Set-Membership Kalman Filter

The IDKF serves as the basis for our distributed combined stochastic and set-membership Kalman filter. A naïve adaption of the IDKF to combined stochastic and set-membership state estimation would entail using all formulae as presented in Sec. III and replacing all vectors with the respective sets, using (1) for all linear transformations, and replacing the plus operation with a Minkowski sum with a subsequent approximation. This results in the formulae

$$\mathcal{Y}^e_{k,s} = \operatorname{approx}_{\text{tr}}\big(\mathcal{Y}^p_{k,s} + \mathcal{J}_{k,s}\big) \ , \quad \mathcal{J}_{k,s} = \mathbf{H}^T_{k,s} (\mathbf{C}^v_{k,s})^{-1} \mathcal{Z}_{k,s}$$

for the update step and

$$\mathcal{Y}^p_{k+1,s} = \mathbf{Y}^p_{k+1,\text{glob}} \mathbf{A}_k (\mathbf{Y}^e_{k,\text{glob}})^{-1} \mathcal{Y}^e_{k,s}$$

for the prediction step. While these equations yield valid, conservative results, they lead to suboptimal approximations of the set-membership uncertainty as we will see in this section.

An important aspect of the combined stochastic and set-membership Kalman filter are the approximations of non-ellipsoidal-shaped sets in the filter step. To give an intuition, let us regard multiple filter steps in state space. For the regular Kalman filter, processing multiple measurements at once yields the same results as performing the filter steps sequentially. In the combined stochastic and set-membership Kalman filter, each filter step results in an approximation. As we will see in the following, it is important when the approximations are performed and what processing steps follow the approximation.

It is essential to be aware of that, in general,

$$\mathbf{L} \operatorname{approx}_{\text{tr}}\big(\mathcal{Q}_1 \oplus \mathcal{Q}_2\big) \neq \operatorname{approx}_{\text{tr}}\big(\mathbf{L}\mathcal{Q}_1 \oplus \mathbf{L}\mathcal{Q}_2\big)$$

holds. Performing the transformation before applying the approximation yields superior results as

$$\operatorname{tr}\big(\mathbf{L} \operatorname{approx}_{\text{tr}}\big(\mathcal{Q}_1 \oplus \mathcal{Q}_2\big)\big) \geq \operatorname{tr}\big(\operatorname{approx}_{\text{tr}}\big(\mathbf{L}\mathcal{Q}_1 \oplus \mathbf{L}\mathcal{Q}_2\big)\big)$$

holds. Let us say, for example, $\mathcal{Q}_1$ and $\mathcal{Q}_2$ are two-dimensional, axis-aligned ellipsoids. The shape matrices $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are thus diagonal matrices. As can be seen from (2), the shape matrix of the approximation of the Minkowski sum is a linear combination of the original shape matrices. Thus, the approximation is axis-aligned as well. If we apply, e.g., $\mathbf{L} = \operatorname{diag}([a, b])$ with $a > 1$ and $b < 1$ after the approximation, then the approximation error along one of the axis is amplified, while the error along the other axis is reduced. Thus, it is possible that a different approximation could have been chosen to achieve a lower trace. If the transformation is applied beforehand, the trace-minimal result respecting the transformation can be obtained. The influence of such transformations on the regular combined stochastic and set-membership Kalman filter is illustrated in the following example.

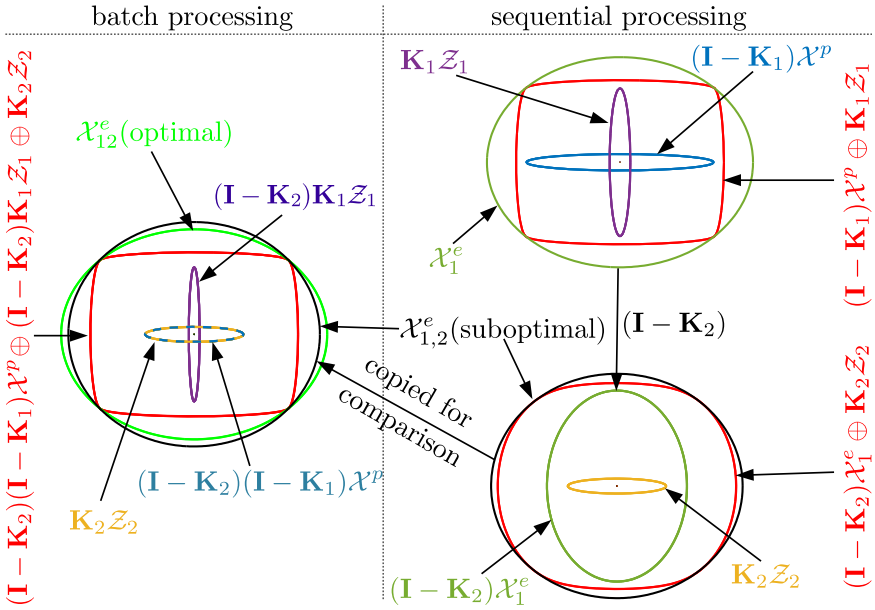

Figure 2. Visualization of the difference between sequential and batch processing in the state space. On the right side, the prior set is combined with the first measurement and the resulting Minkowski sum is approximated. Afterward, the result is fused with the second measurement and another approximation is performed. This is compared with the result of batch processing, in which the Minkowski sum of three ellipsoids is approximated simultaneously. The resulting approximation is superior to that obtained using sequential processing.

Example 1. Assume we have a prior set of means according to $\mathcal{X}^p$ and two sets of possible measurements $\mathcal{Z}_1$ and $\mathcal{Z}_2$ according to $\mathbf{H} = \mathbf{I}$. We call the gain for fusing in the first measurement $\mathbf{K}_1$ and the gain for fusing in the second measurement afterward $\mathbf{K}_2$. Then, using batch processing, we can obtain

$$\mathcal{X}^e_{12} = \text{approx}_{\text{tr}}\big((\mathbf{I}-\mathbf{K}_2)(\mathbf{I}-\mathbf{K}_1)\mathcal{X}^p \oplus (\mathbf{I}-\mathbf{K}_2)\mathbf{K}_1\mathcal{Z}_1 \oplus \mathbf{K}_2\mathcal{Z}_2\big),$$

which is the trace-minimal result that can be obtained for the current time step if no further measurements are obtained. The result and all ellipsoids involved are shown in the left half of Fig. 2.

According to (4), we can first approximate the Minkowski sum of the first two terms and then add the third term and obtain the same result. This, however, is different from performing the filter steps sequentially. If we perform two sequential filter steps, we first approximate the Minkowski sum $(\mathbf{I}-\mathbf{K}_1)\mathcal{X}^p \oplus \mathbf{K}_1\mathcal{Z}_1$ before the result is transformed in the next filter step. The operations performed are

$$\begin{aligned}\mathcal{X}^e_1 &= \text{approx}_{\text{tr}}\big((\mathbf{I}-\mathbf{K}_1)\mathcal{X}^p \oplus \mathbf{K}_1\mathcal{Z}_1\big)\ ,\\ \mathcal{X}^e_{1,2} &= \text{approx}_{\text{tr}}\big((\mathbf{I}-\mathbf{K}_2)\mathcal{X}^e_1 \oplus \mathbf{K}_2\mathcal{Z}_2\big)\ .\end{aligned}$$

In the calculation of $\mathcal{X}^e_{1,2}$, $\mathcal{X}^e_1$ will be transformed according to $\mathbf{L} = (\mathbf{I}-\mathbf{K}_2)$. As we have seen above, this transformation may make the previous approximation suboptimal. As the previous approximation cannot be undone, the suboptimality of the approximation used for $\mathcal{X}^e_1$ may lead to a higher trace in the result of the next filter step than would be achieved if a different approximation was chosen. The sequential processing is illustrated in the right half of Fig. 2.

As shown in the example, performing multiple filter steps using batch processing and approximating only once can yield results superior to those obtained using sequential processing. While the Minkowski sums can be build and approximated sequentially, the linear transformations of the intermediate results lead to conservative but generally suboptimal results. In the first subsection of this section, we begin by laying out how to obtain the trace-minimal approximation using sequential processing when there is only one central node keeping track of the estimate and its uncertainties. In the second subsection, we show how we can obtain the trace-minimal approximation even when the information is split among multiple nodes. The loss of optimality as shown in the example not only applies to transformations caused by sequential filter steps in the state space form but also to transformations caused by prediction steps. Therefore, we lay out how prediction steps can be respected in our novel filter in the third subsection.

A. Central Optimization for the Current Time Step

Before regarding how to perform the approximations when multiple nodes keep track of a part of the state, we discuss how multiple measurements can be combined sequentially in one time step without an additional approximation error. This solution has been presented similarly in [11]. The filter step to update the the set of information vectors in the combined stochastic and set-membership Kalman filter in information filter form can be written as

$$\mathcal{Y}^e_k \supseteq \mathcal{Y}^p_k \oplus \mathcal{J}_{k,1} \oplus \mathcal{J}_{k,2} \oplus \ldots \oplus \mathcal{J}_{k,N}\ ,$$

with sets of information $\mathcal{J}_{k,s}$ derived from the measurements according to

$$\mathcal{J}_{k,s} = \mathbf{H}^T_{k,s}(\mathbf{C}^v_{k,s})^{-1}\mathcal{Z}_{k,s}$$

for $s \in \{1, 2, \ldots, N\}$. If we perform the approximations sequentially, meaning we calculate

$$\mathcal{Y}_k^e = \mathrm{approx}_{\mathrm{tr}}\Big(\ldots \mathrm{approx}_{\mathrm{tr}}\big(\mathrm{approx}_{\mathrm{tr}}\big(\mathcal{Y}_k^p \oplus \mathcal{J}_{k,1}\big) \oplus \mathcal{J}_{k,2}\big)\ldots\Big),$$

then there are no transformations involved after each approximation. Therefore, it is ensured that sequentially approximating the Minkowski sum yields the same result as obtained from batch processing when using the information form. This is a direct result of (4).

However, in information space, it is necessary to be aware of precisely what criterion we aim to optimize when approximating the Minkowski sum using an ellipsoid. Usually, the optimization of the sum of the squared lengths of the semiaxes of the ellipsoid in state space is desired. Thus, the transformation between information and state space has to be taken into account. If we wish to maintain our information form representation, this can be achieved by transferring all ellipsoids to state space, then performing the approximation of the Minkowski sum, and transferring the result back to information space afterward. This means we approximate the Minkowski sums of all transformed ellipsoids on the right-hand side of

$$\mathcal{X}_k^e \supseteq (\mathbf{Y}_k^e)^{-1}\mathcal{Y}_k^p \oplus (\mathbf{Y}_k^e)^{-1}\mathcal{J}_{k,1} \oplus \ldots \oplus (\mathbf{Y}_k^e)^{-1}\mathcal{J}_{k,N} \quad (10)$$

using the formulae (2) and (3) before transferring the result back to the information space.

Instead of transferring all ellipsoids to state space and transferring the result back to information space, we can choose weightings in a way that causes the approximation to be optimal regarding the state space without actually transferring the ellipsoids to state space. If we have multiple ellipsoids $\mathcal{Q}_i = \mathcal{E}(\underline{c}_i, \mathbf{Q}_i)$, we can determine the optimal approximation of the Minkowski sum while taking future affine transformations according to $\mathbf{W}$ into account by calculating the q_i in the formula for the new shape matrix (2) according to

$$q_i = \sqrt{\mathrm{tr}(\mathbf{W}\mathbf{Q}_i\mathbf{W}^T)}\ . \quad (11)$$

This way, it is possible to perform the approximation in information space but still minimize the approximation error in regard to the state space. The weighting that needs to be applied to obtain the optimal approximation regarding the current time step is $\mathbf{W} = (\mathbf{Y}_k^e)^{-1}$.

B. Distributed Optimization for the Current Time Step

For distributed processing, we first have to initialize the pseudotracks in one of the ways described in Sec. III-A. If we choose that the prior is only to be kept track of by one node m, the set $\mathcal{Y}_{0,m}^p$ is initialized using the set of prior means $\mathcal{X}_0^p$ and the covariance of the prior $\mathbf{C}_0^p$ according to

$$\mathcal{Y}_{0,m}^p = \mathcal{Y}_{0,\mathrm{glob}}^p = (\mathbf{C}_0^p)^{-1}\mathcal{X}_0^p$$

and all other sets $\mathcal{Y}_{0,s}^p$ for $s \neq m$ are initialized as sets only containing the zero vector. Analogous to the case without set-membership uncertainty, we can distribute the prior set among the sensor nodes and initialize all as $\frac{1}{N}\mathcal{Y}_0^p$. The result of the combination, which is now performed using the Minkowski sum, still yields $\mathcal{Y}_0^p$ as the Minkowski sum. This directly follows from the fact that the Minkowski sum of multiple identically oriented ellipsoids results in an ellipsoid with no approximation required.

For the distributed filter, we accumulate the knowledge about the set-membership uncertainties in the $\mathcal{Y}_{k,s}^e$ that are kept track of by the individual nodes. Similar to (10), we aim to optimize the required approximation in regard to the state space

$$\mathcal{X}_{k,\mathrm{glob}}^e \supseteq (\mathbf{Y}_{k,\mathrm{glob}}^e)^{-1} \bigoplus_{s=1}^{N} \left(\mathcal{Y}_{k,s}^p \oplus \mathcal{J}_{k,s}^p\right)\ ,$$

for which the transformation $(\mathbf{Y}_{k,\mathrm{glob}}^e)^{-1}$ has to be taken into account. As in Sec. IV-A, we can use weightings instead of transferring the information to state space. In this case, the weighting $\mathbf{W} = (\mathbf{Y}_{k,\mathrm{glob}}^e)^{-1}$ can be used in (11) for all approximations of Minkowski sums occurring to obtain the optimal result. This means if a sensor performs a measurement, it can integrate the new set-membership uncertainty by calculating the approximation required to get the left-hand side of

$$\mathcal{Y}_{k,s}^e \supseteq \mathcal{Y}_{k,s}^p \oplus \mathcal{J}_{k,s}^p \quad (12)$$

using the weighting $\mathbf{W} = (\mathbf{Y}_{k,\mathrm{glob}}^e)^{-1}$. If all $\mathcal{Y}_{k,s}^e$ are fused afterward using a Minkowski sum, the same weighting has to be used to ensure that the approximation is trace-minimal regarding the state space.

C. Distributed Optimization for Future Time Steps

We now regard how the global fusion result after a prediction step of the combined stochastic and set-membership Kalman filter in information form looks like. For each set of information vectors, (9) becomes

$$\mathcal{Y}_{k+1,\mathrm{glob}}^p \supseteq \mathbf{Y}_{k+1,\mathrm{glob}}^p \mathbf{A}_k (\mathbf{Y}_{k,\mathrm{glob}}^e)^{-1} \bigoplus_{s=1}^{N} \mathcal{Y}_{k,s}^e\ , \quad (13)$$

which is equivalent to

$$\mathcal{Y}_{k+1,\mathrm{glob}}^p \supseteq \bigoplus_{s=1}^{N} \underbrace{\mathbf{Y}_{k+1,\mathrm{glob}}^p \mathbf{A}_k (\mathbf{Y}_{k,\mathrm{glob}}^e)^{-1} \mathcal{Y}_{k,s}^e}_{:=\mathcal{Y}_{k+1,s}^p}\ . \quad (14)$$

The optimal approximations of the right-hand sides of (13) and (14) are identical, although different results would be obtained when naïvely approximating (13) as the transformation according to $\mathbf{Y}_{k+1,\mathrm{glob}}^p \mathbf{A}_k (\mathbf{Y}_{k,\mathrm{glob}}^e)^{-1}$ would not be taken into account.

Using weightings, the optimal result can be obtained. As we have shown in Sec. IV-B, $\mathbf{W} = (\mathbf{Y}_{k,\mathrm{glob}}^e)^{-1}$ can be used to fuse the $\mathcal{Y}_{k,s}^e$ in a way that minimizes the approximation error in regard to $\mathcal{X}_{k,\mathrm{glob}}^e$. If we wish to fuse the $\mathcal{Y}_{k+1,s}^p$ in the time step $k+1$ such that $\mathcal{X}_{k+1,\mathrm{glob}}^p = (\mathbf{Y}_{k+1,\mathrm{glob}}^p)^{-1}\mathcal{Y}_{k+1,\mathrm{glob}}^p$ is optimal, then the weighting caused by the different transformation to state space changes to $\mathbf{W} = (\mathbf{Y}_{k+1,\mathrm{glob}}^p)^{-1}$.

Let us now consider how to perform the approximations in the filter steps in the time step k when the trace of the shape

matrix of $\mathcal{X}^p_{k+1,\text{glob}}$ is to be minimized. To be optimal regarding $\mathcal{X}^p_{k+1,\text{glob}}$, we not only have to respect the transformation back to state space but also all linear transformations up to $\mathcal{Y}^p_{k+1,\text{glob}}$. Thus, the weighting for the filter steps in the time step k becomes

$$\mathbf{W} = \underbrace{(\mathbf{Y}^p_{k+1,\text{glob}})^{-1}}_{\text{transformation of } \mathcal{Y}^p_{k+1,s} \text{ to state space}} \overbrace{\mathbf{Y}^p_{k+1,\text{glob}} \mathbf{A}_k (\mathbf{Y}^e_{k,\text{glob}})^{-1}}^{\text{transformation from } \mathcal{Y}^e_{k,s} \text{ to } \mathcal{Y}^p_{k+1,s}} \quad , \quad (15)$$

which can be simplified to $\mathbf{W} = \mathbf{A}_k (\mathbf{Y}^e_{k,\text{glob}})^{-1}$. This weighting, however, only applies to the filter step at time step k. The approximation of the Minkowski sum of the $\mathcal{Y}^p_{k+1,s}$ in the time step $k+1$ to obtain $\mathcal{Y}^p_{k+1,\text{glob}}$ still needs to be performed using the weighting $\mathbf{W} = (\mathbf{Y}^p_{k+1,\text{glob}})^{-1}$.

If filter steps are to be performed at time step $k+1$, meaning, $\mathcal{X}^e_{k+1,\text{glob}}$ is to be optimized, the transformation to state space changes, changing the weighting to be used for the filter steps at time step k. The weighting in the filter steps at time step k then becomes

$$\mathbf{W} = \underbrace{(\mathbf{Y}^e_{k+1,\text{glob}})^{-1}}_{\text{transformation of } \mathcal{Y}^e_{k+1,s} \text{ to state space}} \overbrace{\mathbf{Y}^p_{k+1,\text{glob}} \mathbf{A}_k (\mathbf{Y}^e_{k,\text{glob}})^{-1}}^{\text{transformation from } \mathcal{Y}^e_{k,s} \text{ to } \mathcal{Y}^p_{k+1,s}} \quad . \quad (16)$$

The weighting in the filter steps at time step $k+1$ does not need to respect any transformations due to prediction steps. Thus, the weighting for performing a filter step or fusing the information in the time step $k+1$ is $\mathbf{W} = (\mathbf{Y}^e_{k+1,\text{glob}})^{-1}$.

We can generalize (16) to the optimization of the filter result at a prespecified time step $k+t$ (in other words, optimizing the trace of the shape matrix of $\mathcal{X}^e_{k+t,\text{glob}}$) by integrating multiple prediction steps into the formula. The weighting to be used at time step k becomes

$$\mathbf{W} = \underbrace{(\mathbf{Y}^e_{k+t,\text{glob}})^{-1}}_{\text{transformation of } \mathcal{Y}^e_{k+t,s} \text{ to state space}} \overbrace{\prod_{i=0}^{t-1} \mathbf{Y}^p_{k+i+1,\text{glob}} \mathbf{A}_{k+i} (\mathbf{Y}^e_{k+i,\text{glob}})^{-1}}^{\text{transformation from } \mathcal{Y}^e_{k,s} \text{ to } \mathcal{Y}^p_{k+t,s}} . \quad (17)$$

For the weighting in the time step $k+1$, only the terms from $i=1$ to $i=t-1$ have to be respected in the product. With each additional prediction step performed, the product used to calculate the weighting is reduced by one term until the time step $k+t$ is reached, at which the weighting is $(\mathbf{Y}^e_{k+t,\text{glob}})^{-1}$ as in (12). If the last step is a prediction step instead of a filter step, the leftmost term $(\mathbf{Y}^e_{k+t,\text{glob}})^{-1}$ in (17) needs to be replaced with $(\mathbf{Y}^p_{k+t,\text{glob}})^{-1}$.

If we have the trace-minimal solution regarding time step $k+t$, any future prediction steps may void the optimality of the approximation. Hence, an arbitrary but prespecified time step $k+t$ needs to be chosen in the approach to attain the optimal fusion result for this time step. For all other time steps, the approximation is still guaranteed to contain all elements that would be contained if the Minkowski sums were never approximated and all sets were propagated to the following time steps. Thus, the approach is conservative for all other time steps. However, the approximation provided is generally suboptimal for all time steps except the chosen one.

Example 2. Let us assume we wish to minimize the sum of the squared lengths of the semiaxes of $\mathcal{X}^e_{k+2,\text{glob}}$. For this, we set $t=2$ in formula (17). Then, we have to use different weightings when approximating the Minkowski sums in different time steps. The different weightings for the filter steps in the time steps k, $k+1$, and $k+2$ as well as for the fusion at time step $k+2$ are given in Table I. No weightings are given for the prediction step as there are no Minkowski sums involved.

Since the stochastic system and measurement models, including the uncertainties, are known by all nodes, each node can calculate the weightings itself. To fuse the information of all sensors, each node only needs to transmit its part of the center and the shape of the set-membership uncertainty. Since the approximation of the Minkowski sum with an identical weighting can be done sequentially, the sets may be fused and approximated along the communication path to reduce the data to be transferred. However, no prediction steps may be performed until all information has been accumulated in the nodes that aim to derive the globally optimal result.

Remark 3. If the set-membership uncertainties of the measurements are also known by all nodes, no communication is required for determining the total set-membership uncertainty. In this case, the nodes are directly initialized using $\mathcal{Y}^p_{0,\text{glob}}$. As an individual node cannot keep track of $\mathcal{Y}_{k,\text{glob}}$ without knowledge about the center of the possible measurements obtained by the other nodes, each node only keeps track of the shape matrix of $\mathcal{Y}_{k,\text{glob}}$. If we use $\bar{\mathcal{J}}_{k,l}$ to denote the set $\mathcal{J}_{k,l}$ with the center set to zero, each node s can calculate the approximation of the Minkowski sum

$$\mathcal{Y}^e_{k,\text{glob}} \supseteq \mathcal{Y}^p_{k,\text{glob}} \oplus \mathcal{J}_{k,s} \bigoplus_{l \in \{1,\dots,n\} \setminus s} \bar{\mathcal{J}}_{k,l}$$

using the weighting described in (17). With all uncertainties correctly modeled, communication is then only required to attain the correct center $\underline{\hat{y}}^p_{k,\text{glob}}$ for the ellipsoid $\mathcal{Y}^p_{k,\text{glob}}$.

Regardless whether the set-membership uncertainties of the measurements are known by all sensors or not, by performing the approximations using weightings, we achieve the optimal approximation. The optimal approximation is generally even better than that obtained by a centralized filter that uses the regular combined stochastic and set-membership Kalman filter that performs the approximations sequentially. As the centralized filter does not use weightings, each intermediate approximation may be suboptimal in respect to the final result. The optimal approximation corresponds to the approximation we would obtain if we propagated all terms of the Minkowski sum to the following time steps and performed the approximation of the Minkowski sum of all ellipsoids involved at the time step $k+t$. This, however, would be much more costly than the approach presented.

Remark 4. Once t has been fixed, the matrix product (17), which is in $O(t)$, only needs to be calculated once. The

Table I
WEIGHTINGS $\mathbf{W}$ TO BE USED IN (11) TO OBTAIN THE OPTIMAL APPROXIMATION IN THE TIME STEP $k+2$.

Time	k	$k+1$	$k+2$	$k+2$
Step	Filter	Filter	Filter	Fusion
Weighting	$(\mathbf{Y}^e_{k+2})^{-1}\mathbf{Y}^p_{k+1}\mathbf{A}_{k+1}(\mathbf{Y}^e_{k+1})^{-1}\mathbf{Y}^p_k\mathbf{A}_k(\mathbf{Y}^e_k)^{-1}$	$(\mathbf{Y}^e_{k+2})^{-1}\mathbf{Y}^p_{k+1}\mathbf{A}_{k+i}(\mathbf{Y}^e_{k+1})^{-1}$	$(\mathbf{Y}^e_{k+2})^{-1}$	$(\mathbf{Y}^e_{k+2})^{-1}$

weighting can then be updated in $O(1)$ by multiplying the previous weighting with $(\mathbf{Y}^p_{k+i+1,\text{glob}}\mathbf{A}_k(\mathbf{Y}^e_{k+i,\text{glob}})^{-1})^{-1}$, which eliminates one term of the product. For software implementations, it can be a challenge that the values of $\mathbf{W}$ in (17) may become very small for large t. However, since the q_i that result in the lowest sum of the squared lengths of the semiaxes also optimize the criterion for a scaled version of the same problem, we can multiply the matrix $\mathbf{W}$ by an arbitrary constant to improve the numerical properties of the algorithm.

V. EVALUATION

In our evaluation, we compare the results of four different filters in a simple scenario involving only two sensors. For one, we use a fully centralized version of the regular combined stochastic and set-membership Kalman filter without any weightings. As briefly mentioned in the previous section, this filter is not optimal if no weightings, as described in this paper, are used. The naïve centralized filter, however, has the highest communication requirement of all filters as all measurements need to be transmitted to the fusing node in each time step. As the second rather communication intensive approach, we use the approach presented in [11], which we shall call Fusion'12. This approach allows us to perform an optimal fusion regarding the current time step but does not support distributed prediction steps. Less communication is required than for the first filter as information can be accumulated along multiple hops.

The third and fourth approach are ones that do not necessitate communication in every time step. In the third approach, each node uses an independent combined stochastic and set-membership Kalman filter. The fusion step is then performed using an adapted version of the covariance intersection (CI) [20] approach that can take set-membership uncertainty into account. As the fourth and last filter, we use our novel approach using weightings in the information space.

As the scenario for our evaluation, we simulate a minimal example involving two sensors communicating over a wireless link that observe a four-dimensional state. The system and measurement model and the stochastic uncertainties of the models are time-invariant, which facilitates the application of our novel approach. Furthermore, the set-membership uncertainties of the measurements are chosen to be time-invariant to allow us to observe clear trends in the evaluation. The set-membership uncertainty is initialized with a large uncertainty of $\mathcal{X}^p_0 = \mathcal{E}(\underline{\hat{x}}^p_0, 150\mathbf{I})$. The system matrix, measurements matrices, and shape matrices of the set-membership uncertainties are diagonal matrices with integer values on the diagonal. The measurement covariance matrices of the two sensors are two dense, independently generated, symmetric positive definite matrices with values between 0.7 and 2.9.

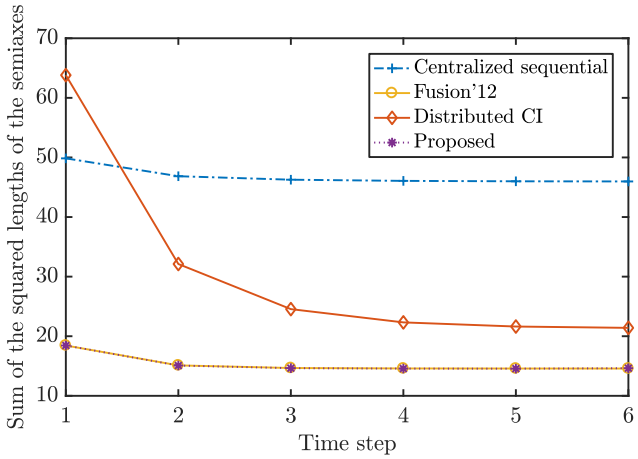

Figure 3. Sum of the squared lengths of the semiaxes of the set-membership uncertainty plotted for different time steps. The approaches in the legend are ordered as in the explanation in Sec. V.

The results of the evaluation are shown in Fig. 3. For the centralized Kalman filter and the Fusion'12 algorithm, we plot the size of the set-membership uncertainty in the dedicated fusing node at the k-th time. For the distributed filter that merges via CI, the uncertainty obtained when fusing the tracks (only) at the k-th time step is shown. For our novel approach, the results shown were obtained by using the weight that optimizes the results regarding the k-th time step.

In this scenario, it can be seen that the Fusion'12 approach and our novel one perform almost identically. However, our novel approach requires far less communication because the optimal result in the k-th time step can be obtained using only 1 communication event instead of k. Furthermore, our novel approach performs slightly better. The centralized approach with sequential processing in state space performs the worst of the evaluated approaches in regard to the set-membership uncertainty. Due to the multitude of transformations by not only the system matrix but also the different gains, the approximations performed are highly suboptimal. This shows that even for full rate communication (or a centralized phenomenon), it is essential to use weightings when having multiple measurements in one time step to obtain a result that is at least optimal regarding the current time step.

The approach using CI performs better regarding the set-membership uncertainty than the centralized filter that has all measurements at all time steps at its disposal. The approach using CI, however, is the only one that does not yield the optimal covariance and thus results in a higher stochastic uncertainty. We have verbally summarized the performance of all approaches regarding the communication overhead, the

Table II
VERBAL SUMMARY OF THE STRENGTHS AND WEAKNESSES OF THE INDIVIDUAL FILTERS FOR DISTRIBUTED SCENARIOS IN WHICH OUR NOVEL APPROACH IS APPLICABLE. THE COLUMN "COMMUNICATION" INDICATES THE NUMBER OF TRANSMISSIONS REQUIRED TO OBTAIN THE RESULT IN A PRESPECIFIED TIME STEP, "QUALITY" DESCRIBES HOW CLOSE THE APPROXIMATION OF THE SET-MEMBERSHIP UNCERTAINTY IS TO THE OPTIMAL APPROXIMATION, AND THE LAST COLUMN DESCRIBES IF THE COVARIANCE PROVIDED IS OPTIMAL.

Algorithm	Communication	Quality	Optimal covariance
Centralized	High	Low	Yes
FUSION'12	Medium	High	Yes
Distributed CI	Low	Medium	No
Novel approach	Low	Optimal	Yes

quality in terms of the size of the set-membership uncertainty, and the optimality of the covariance in Table II.

VI. CONCLUSION

Our proposed novel filter for combined stochastic and set-membership state estimation provides high quality results and allows performing the prediction and filter steps in a distributed manner. The approach has prerequisites analogous to the IDKF while featuring similar strengths. The IDKF that does not support set-membership uncertainties does not require knowledge about the time step at which the fusion will occur to be optimal regarding the stochastic uncertainty. If the time step at which the fusion will occur is not known, our proposed distributed combined stochastic and set-membership Kalman filter is also optimal regarding the stochastic uncertainty but is generally not optimal regarding the set-membership uncertainty.

To minimize the approximation error in the set-membership uncertainty, we also require knowledge about the time step at which the fusion will occur. Then, however, our proposed approach is as good as a far more expensive filter that requires transmission and propagation of all individual ellipsoids and only approximates the Minkowki sum when a single ellipsoid is to be returned. Thus, our novel approach should even be applied to central processing tasks when multiple measurements are obtained in one time step, especially if the set-membership uncertainty at a future time step is to be minimized. Regardless whether the information is fused at the prespecified time step or at a different time step, the results are always conservative and never feature uncertainties that may be too small.

In future work, we aim to closely regard distributed processing in filters that do not use the regular Kalman gain. By doing so, measures of uncertainty taking both types of uncertainties into account can be improved. Current approaches respecting both types of uncertainties [8] only optimize the result regarding the current filter step. Taking future filter and prediction steps into account may help us obtain smaller uncertainties in the long run.

ACKNOWLEDGMENT

The IGF project 18798 N of the research association Forschungs-Gesellschaft Verfahrens-Technik e.V. (GVT) was supported via the AiF in a program to promote the Industrial Community Research and Development (IGF) by the Federal Ministry for Economic Affairs and Energy on the basis of a resolution of the German Bundestag. This work was also supported by the German Research Foundation (DFG) under grant NO 1133/1-1.

REFERENCES

[1] D. Morrell and W. Stirling, "Set-Valued Filtering and Smoothing," in *Twenty-Second Asilomar Conference on Signals, Systems and Computers*, vol. 1, 1988, pp. 236–240.
[2] F. C. Schweppe, *Uncertain Dynamic Systems*. Prentice Hall, 1973.
[3] A. Kurzhanski and I. Valyi, *Ellipsoidal Calculus for Estimation and Control*. Nelson Thornes, 1997.
[4] F. L. Chernousko, *State Estimation for Dynamic Systems*. CRC Press, 1993.
[5] L. Ros, A. Sabater, and F. Thomas, "An Ellipsoidal Calculus Based on Propagation and Fusion," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 32, no. 4, pp. 430–442, Aug. 2002.
[6] 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.
[7] 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, USA, Jul. 2009.
[8] B. Noack, F. Pfaff, and U. D. Hanebeck, "Optimal Kalman Gains for Combined Stochastic and Set-Membership State Estimation," in *Proceedings of the 51st IEEE Conference on Decision and Control (CDC 2012)*, Maui, Hawaii, USA, Dec. 2012.
[9] B. Noack, F. Pfaff, M. Baum, and U. D. Hanebeck, "State Estimation Considering Negative Information with Switching Kalman and Ellipsoidal Filtering," in *Proceedings of the 19th International Conference on Information Fusion (Fusion 2016)*, Heidelberg, Germany, Jul. 2016.
[10] J. Sijs, B. Noack, and U. D. Hanebeck, "Event-Based State Estimation with Negative Information," in *Proceedings of the 16th International Conference on Information Fusion (Fusion 2013)*, Istanbul, Turkey, Jul. 2013.
[11] B. Noack, F. Pfaff, and U. D. Hanebeck, "Combined Stochastic and Set-Membership Information Filtering in Multisensor Systems," in *Proceedings of the 15th International Conference on Information Fusion (Fusion 2012)*, Singapore, Jul. 2012.
[12] F. Govaers and W. Koch, "An Exact Solution to Track-to-Track-Fusion at Arbitrary Communication Rates," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 3, pp. 2718–2729, 2012.
[13] Y. Becis-Aubry, D. Aubry, and N. Ramdani, "Multisensor Set-Membership State Estimation of Nonlinear Models With Potentially Failing Measurements," in *Proceedings of the 18th IFAC World Congress (IFAC 2011)*, Milan, Italy, Aug. 2011.
[14] Z. Wang, X. Shen, and Y. Zhu, "Set-Membership Information Fusion for Multisensor Nonlinear Dynamic Systems," *arXiv:1702.05214 [cs, math]*, Feb. 2017, arXiv: 1702.05214.
[15] D. Shi, T. Chen, and L. Shi, "On Set-Valued Kalman Filtering and Its Application to Event-Based State Estimation," *IEEE Transactions on Automatic Control*, vol. 60, no. 5, pp. 1275–1290, May 2015.
[16] F. Pfaff, B. Noack, and U. D. Hanebeck, "Information Form Distributed Kalman Filtering (IDKF) with Explicit Inputs," in *Proceedings of the 20th International Conference on Information Fusion (Fusion 2017)*, Xi'an, China, Jul. 2017.
[17] A. G. O. Mutambara, *Decentralized Estimation and Control for Multisensor Systems*. Boca Raton, Florida, USA: CRC Press, Inc., 1998.
[18] C. Durieu, É. Walter, and B. Polyak, "Multi-Input Multi-Output Ellipsoidal State Bounding," *Journal of Optimization Theory and Applications*, vol. 111, no. 2, pp. 273–303, 2001.
[19] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1993.
[20] S. J. Julier and J. K. Uhlmann, "A Non-Divergent Estimation Algorithm in the Presence of Unknown Correlations," in *Proceedings of the 1997 American Control Conference*, vol. 4. IEEE, 1997, pp. 2369–2373.