Nonlinear Information Filtering for Distributed Multisensor Data Fusion

Benjamin Noack, Daniel Lyons, Matthias Nagel, and Uwe D. Hanebeck

***Abstract*— The information filter has evolved into a key tool for distributed and decentralized multisensor estimation and control. Essentially, it is an algebraical reformulation of the Kalman filter and provides estimates on the information about an uncertain state rather than on a state itself. Whereas many practicable Kalman filtering techniques for nonlinear system and sensor models have been developed, approaches towards nonlinear information filtering are still scarce and limited. In order to deal with nonlinear systems and sensors, this paper derives an approximation technique for arbitrary probability densities that provides the same distributable fusion structure as the linear information filter. The presented approach not only constitutes a nonlinear version of the information filter, but it also points the direction to a Hilbert space structure on probability densities, whose vector space operations correspond to the fusion and weighting of information.**

I. INTRODUCTION

Practical estimation and control applications generally entail the difficulty that the measured data is corrupted by noise. Thus, the derivation of specific estimates from noisy measurements is of little value, if the involved uncertainties are not considered appropriately. A common approach consists in modeling uncertain quantities stochastically by calculating mean and estimate variance. These parameters correspond to a Gaussian density characterizing the uncertainty about the state. For a linear evolution of the state variable and linear observation models, the Kalman filter [1] formulas represent an optimal closed-form solution to the estimation problem. In nonlinear situations, mean and estimate variance often do not suffice to describe the underlying uncertainty and they can even be deceptive, in particular when the true probability density of the state estimate is multi-modal. Since a closed-form computation of the actual density is generally not possible, a lot of effort has been focused on approximative solutions to nonlinear Bayesian state estimation. For this purpose, either the underlying system and measurement mappings or the underlying probability densities are approximated. In the former case, the nonlinearities are generally linearized by first-order Taylor series approximations, which are performed within the extended Kalman filter, or by a linear regression analysis, of which the uncented Kalman filter [2] is a well-known example. Of course, due to the Gaussian assumption, they only provide very limited capabilities for capturing multi-modalities. This can be better achieved through density approximations such as particle filters [3] or finite-dimensional representations via orthonormal bases, e.g., truncated Fourier [4], [5] or wavelet [6] series. All these approaches are intended to provide finite and implementable parameterizations of the state estimates.

In the recent past, the rapid advances in sensor and communication technology have increased the demand for distributed and decentralized estimation and control architectures. For linear distributed sensor systems, again an optimal closed-form solution exists: The information filter [7] has been derived as an inverse covariance formulation of the Kalman filter, with the benefit that the fusion of multiple sensor data can easily be distributed. This reformulation has widely been applied to sensor networks [8], [9]. The information filter herein calculates estimates on the information about the state and not on the state itself, which simplifies the fusion significantly. Also, the elimination of double-counted information between two sensors nodes becomes very simple in the information space.

Unfortunately, distributed fusion is still particularly challenging for nonlinear systems and sensors: The extended information filter [7] comes with the same drawbacks as the extended Kalman filter and developing distributed fusion structures for arbitrary probability densities is elaborate [10]. For example, particles need to be transformed into continuous representations [11]. In order to tackle these issues, the aim of this paper is to lay down the theoretical foundation for tractable nonlinear distributed data fusion. For this purpose, a general formulation of the information filter will be considered, which essentially is a log-likelihood formulation of Bayes' theorem. In order to fuse densities efficiently in this information space representation, the logarithms of the participating densities are approximated by truncated orthonormal series expansions. In terms of the corresponding coefficient vectors, the fusion and removal of information becomes as simple as for the linear information filter.

Moreover, the considered representation is related to a Hilbert space structure on the probability densities themselves. The corresponding vector space operations, addition and scalar multiplication, are herein directly related to the Bayesian fusion and the weighting of information, respectively. Also, the norm in this Hilbert space provides a promising measure of information. In the following section, an overview of linear estimation is provided and the advantages of the information filter against the background of distributed fusion are highlighted, before the remainder of this paper focuses on nonlinear information filtering.

Benjamin Noack, Daniel Lyons, Matthias Nagel, and Uwe D. Hanebeck are with the Intelligent Sensor-Actuator-Systems Laboratory (ISAS), Institute for Anthropomatics, Karlsruhe Institute of Technology (KIT), Germany. Email: {benjamin.noack|daniel.lyons}@ieee.org, matthias.nagel@student.kit.edu, uwe.hanebeck@ieee.org.

II. STATE OF THE ART: LINEAR INFORMATION FILTERING

Practical systems are in general affected by perturbations and inaccuracies, which have to be dealt with. For uncertain linear discrete-time systems and linear observation models

$$\underline{\boldsymbol{x}}_{k+1} = \mathbf{A}_k \underline{\boldsymbol{x}}_k + \underline{\boldsymbol{w}}_k \quad \text{and} \quad \underline{\hat{\boldsymbol{z}}}_k = \mathbf{H}_k \underline{\boldsymbol{x}}_k + \underline{\boldsymbol{v}}_k \ ,$$

the Kalman filter [1] and its derivatives provide estimates on the uncertain state $\underline{\boldsymbol{x}}_k$ by computing mean $\underline{\hat{x}}_k$ and estimate covariance matrix $\mathbf{C}_k$ at each time instant k, where vectors are underlined, matrices are denoted by bold uppercase letters, and random quantities are bold lowercase letters. Uncertainties in the system evolution are characterized by additive zero-mean white Gaussian noise $\underline{\boldsymbol{w}}_k \sim \mathcal{N}(\underline{0}, \mathbf{C}_k^w)$. The Kalman prediction step calculates the conditional mean

$$\underline{\hat{x}}_{k+1}^{\mathrm{p}} = \mathrm{E}\{\underline{x}_{k+1} \,|\, \underline{\hat{z}}_k, \underline{\hat{z}}_{k-1}, \ldots\} = \mathbf{A}_k \, \underline{\hat{x}}_k^{\mathrm{e}} \tag{1}$$

and the corresponding covariance matrix

$$\begin{aligned} \mathbf{C}_{k+1}^{\mathrm{p}} &= \mathrm{E}\left\{(\underline{x}_{k+1} - \underline{\hat{x}}_{k+1}^{\mathrm{p}})(\underline{x}_{k+1} - \underline{\hat{x}}_{k+1}^{\mathrm{p}})^{\mathrm{T}} \,\middle|\, \underline{\hat{z}}_k, \underline{\hat{z}}_{k-1}, \ldots\right\} \\ &= \mathbf{A}_k \mathbf{C}_k^{\mathrm{e}} \mathbf{A}_k^{\mathrm{T}} + \mathbf{C}_k^w \end{aligned} \tag{2}$$

of the predicted state estimate. By means of the measurement model, the predicted state estimate can then be fused with an observation $\underline{\hat{z}}_k \in \mathbb{R}^m$ corrupted by zero-mean white Gaussian noise $\underline{\boldsymbol{v}}_k \sim \mathcal{N}(\underline{0}, \mathbf{C}_k^v)$. The Kalman filtering step now provides the conditional mean

$$\underline{\hat{x}}_k^{\mathrm{e}} = \mathrm{E}\{\underline{x}_k \,|\, \underline{\hat{z}}_k, \underline{\hat{z}}_{k-1}, \ldots\} = \underline{\hat{x}}_k^{\mathrm{p}} + \mathbf{K}_k \left(\underline{\hat{z}}_k - \mathbf{H}_k \, \underline{\hat{x}}_k^{\mathrm{p}}\right) \tag{3}$$

and covariance matrix

$$\begin{aligned} \mathbf{C}_k^{\mathrm{e}} &= \mathrm{E}\left\{(\underline{x}_k - \underline{\hat{x}}_k^{\mathrm{e}})(\underline{x}_k - \underline{\hat{x}}_k^{\mathrm{e}})^{\mathrm{T}} \,\middle|\, \underline{\hat{z}}_k, \underline{\hat{z}}_{k-1}, \ldots\right\} \\ &= \mathbf{C}_k^{\mathrm{p}} - \mathbf{K}_k \mathbf{H}_k \mathbf{C}_k^{\mathrm{p}} \end{aligned} \tag{4}$$

of the updated estimate $\underline{\boldsymbol{x}}_k^{\mathrm{e}}$, where

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

is the Kalman gain. These prediction and fusion formulas are designed to minimize the trace of the mean-squared error covariance matrices (2) and (4). The Kalman filter is a Bayesian estimator, i.e., the filtering step corresponds to Bayes' rule

$$f_k^{\mathrm{e}}(\underline{x}_k) = f(\underline{x}_k | \underline{\hat{z}}_k, \underline{\hat{z}}_{k-1}, \ldots) = \frac{f_k^{\mathrm{p}}(\underline{x}_k) \cdot f(\underline{\hat{z}}_k \,|\, \underline{x}_k)}{f(\underline{\hat{z}}_k \,|\, \underline{\hat{z}}_{k-1}, \underline{\hat{z}}_{k-2}, \ldots)} \ . \tag{5}$$

The conditional estimated density $f_k^{\mathrm{e}}(\underline{x}_k)$ is normally distributed with mean (3) and variance (4). The predicted density $f_k^{\mathrm{p}}(\underline{x}_k) = f(\underline{x}_k | \underline{\hat{z}}_{k-1}, \underline{\hat{z}}_{k-2}, \ldots)$ is also a Gaussian density with parameters (1) and (2). This state-space formulation of the estimation problem entails the difficulty that, in general, (3) cannot easily be extended to multiple measurements, as it would be useful in sensor networks and multisensor data fusion. This means that for a set of observations $\underline{\hat{z}}_k^1, \ldots, \underline{\hat{z}}_k^M$ from M sensor nodes, the inequality

$$\underline{\hat{x}}_k^{\mathrm{e}} \neq \underline{\hat{x}}_k^{\mathrm{p}} + \sum_{i=1}^{M} \mathbf{K}_k^i \left(\underline{\hat{z}}_k^i - \mathbf{H}_k^i \, \underline{\hat{x}}_k^{\mathrm{p}}\right) \tag{6}$$

holds, because the individual innovations $\left(\underline{\hat{z}}_k^i - \mathbf{H}_k^i \, \underline{\hat{x}}_k^{\mathrm{p}}\right)$ are correlated due to a common prediction of the state [12]. Instead, one has to consider the combined observation vector $\underline{\hat{z}}_k := [(\underline{\hat{z}}_k^1)^{\mathrm{T}}, \ldots, (\underline{\hat{z}}_k^M)^{\mathrm{T}}]^{\mathrm{T}}$ and measurement mapping $\mathbf{H}_k := [(\mathbf{H}_k^1)^{\mathrm{T}}, \ldots, (\mathbf{H}_k^M)^{\mathrm{T}}]^{\mathrm{T}}$ and has to insert them into (3) and (4). Evidently, the need for calculating the Kalman innovation for all measurements at once prohibits efficient distributed data fusion algorithms.

Simply by rearranging the Kalman prediction and filtering formulas and considering the *information state* vector

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

and the *information matrix*

$$\mathbf{Y}_k := \mathbf{C}_k^{-1} \tag{7}$$

instead of the state itself, distributed estimation can be eased significantly [7], [12]. Through this inverse covariance formulation of the estimation problem, the fusion step turns into the sums

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

with

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

Data fusion in the information space is obviously predestined for dealing with multiple sensor data $\mathcal{Z}_k = \{\underline{\hat{z}}_k^1, \ldots, \underline{\hat{z}}_k^M\}$ at each time instant. In contrast to inequality (6), the fusion rule becomes

$$\begin{aligned} \underline{\hat{y}}_k^{\mathrm{e}} &= \underline{\hat{y}}_k^{\mathrm{p}} + \sum_{i=1}^{M} \underline{i}_k^i = \underline{\hat{y}}_k^{\mathrm{p}} + \sum_{i=1}^{M} (\mathbf{H}_k^i)^{\mathrm{T}} (\mathbf{C}_k^{v,i})^{-1} \underline{\hat{z}}_k^i \ , \\ \mathbf{Y}_k^{\mathrm{e}} &= \mathbf{Y}_k^{\mathrm{p}} + \sum_{i=1}^{M} \mathbf{I}_k^i = \mathbf{Y}_k^{\mathrm{p}} + \sum_{i=1}^{M} (\mathbf{H}_k^i)^{\mathrm{T}} (\mathbf{C}_k^{v,i})^{-1} \mathbf{H}_k^i \ , \end{aligned} \tag{8}$$

which is due to the fact that the observations are conditionally independent, i.e.,

$$f(\mathcal{Z}_k \,|\, \underline{x}_k) = f(\underline{\hat{z}}_k^1, \ldots, \underline{\hat{z}}_k^M \,|\, \underline{x}_k) = \prod_{i=1}^{M} f^i(\underline{\hat{z}}_k^i \,|\, \underline{x}_k) \ . \tag{9}$$

The matrices $\mathbf{H}_k^i$ and $\mathbf{C}_k^{v,i}$ denote the individual sensor mappings and noises, respectively. This simple filtering algorithm comes at the expense of a slightly more elaborate prediction step, which consists of the calculations

$$\begin{aligned} \mathbf{Y}_{k+1}^{\mathrm{p}} &= [\mathbf{A}_k (\mathbf{Y}_k^{\mathrm{e}})^{-1} \mathbf{A}_k^{\mathrm{T}} + \mathbf{C}_k^w]^{-1} \ , \\ \mathbf{L}_{k+1} &= \mathbf{Y}_{k+1}^{\mathrm{p}} \mathbf{A}_k (\mathbf{Y}_k^{\mathrm{e}})^{-1} \ , \text{ and} \\ \underline{\hat{y}}_{k+1}^{\mathrm{p}} &= \mathbf{L}_{k+1} \, \underline{\hat{y}}_k^{\mathrm{e}} \ . \end{aligned}$$

As illustrated in Fig. 1, the fusion structure in (8) can easily be distributed. Also, more sophisticated distributed fusion systems, in which, for example, local estimates are generated at each node, can efficiently be implemented in information space [12]. Unfortunately, the fusion architecture (8) is only of limited value for nonlinear systems and sensors. This issue is in the focus of the remainder of this paper.

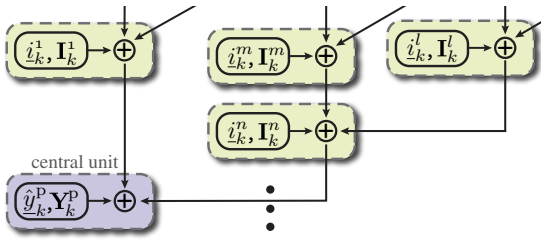

Fig. 1. Every sensor node calculates its contribution and fuses it with data from other nodes. In the data sink, the sensor data is then fused with prior information.

III. THE IDEA BEHIND

It can be shown [7] that the information filter is a log-likelihood representation of Bayesian state estimation and that the inverse covariance matrix (7) is equal to the Fisher information. Thus, minimum mean squared error estimation corresponds to maximizing the Fisher information about the state. In terms of the corresponding densities and by employing the conditional indepence (9), the Bayesian fusion rule

$$f_k^{\mathrm{e}}(\underline{x}_k) = \frac{f_k^{\mathrm{p}}(\underline{x}_k) \cdot \prod_{i=1}^{M} f^i(\hat{\underline{z}}_k^i \,|\, \underline{x}_k)}{f(\mathcal{Z}_k \,|\, \mathcal{Z}_{k-1}, \mathcal{Z}_{k-2}, \ldots)} \;, \tag{10}$$

where $\mathcal{Z}_k = \{\hat{\underline{z}}_k^1, \ldots, \hat{\underline{z}}_k^M\}$ denotes the multiple observations at time instant k, becomes the sum

$$\begin{aligned} \ln f_k^{\mathrm{e}}(\underline{x}_k) = &\ln f_k^{\mathrm{p}}(\underline{x}_k) + \sum_{i=1}^{M} \ln f^i(\hat{\underline{z}}_k^i \,|\, \underline{x}_k) \\ &- \ln f(\mathcal{Z}_k \,|\, \mathcal{Z}_{k-1}, \mathcal{Z}_{k-2}, \ldots) \;. \end{aligned} \tag{11}$$

In the following, we will refer to (11) as the *information space* formulation of the estimation problem, while (10) is referred to as the *state space* formulation. The measurements are conditionally independent given the current state, but the joint probability density $f(\mathcal{Z}_k \,|\, \mathcal{Z}_{k-1}, \mathcal{Z}_{k-2}, \ldots)$ in (10) cannot be split up in the same way as the likelihoods (9). This states the reason why the Kalman filter is not easily distributable, i.e., the reason for inequality (6). In contrast, in the information space, the normalizing constant $\ln f(\mathcal{Z}_k \,|\, \mathcal{Z}_{k-1}, \mathcal{Z}_{k-2}, \ldots)$ is, simply speaking, decoupled and excluded from the fusion process. Instead, it is implicitly calculated whenever the estimate is transformed back into the state space. For Gaussian densities, the sum $\sum_{i=1}^{M} \ln f^i(\hat{\underline{z}}_k^i \,|\, \underline{x}_k)$ in (11) directly corresponds to the sums $\sum_{i=1}^{M} \underline{i}_k^i$ and $\sum_{i=1}^{M} \mathbf{I}_k^i$ in (8) and constant terms in (11) are simply ignored.

In the following, the reformulation (11) of the data fusion problem will provide the basis for information filtering with arbitrary densities and nonlinear models.

IV. OUTLINE TO NONLINEAR INFORMATION FILTERING

Of course, extending the linear information filter in the same way the Kalman filter is extended by linearizing nonlinear system and sensor mappings appears to be most apparent, but this also yields the same drawbacks as for the extended Kalman filter: The estimates may severely suffer from linearization errors and the uncertainty is often underestimated by the estimated covariance matrix. Other approaches towards nonlinear distributed estimation [11], [10] directly consider the Bayesian fusion rule (10) and therefore cannot inherit the advantages of the information space formulation. Especially eliminating common information between two local estimates is often a necessity in distributed and decentralized systems in order to avoid double counting of information, i.e.,

$$f(\underline{x}_k \,|\, \mathcal{Z}_k^i \cup \mathcal{Z}_k^j) \propto \frac{f(\underline{x}_k \,|\, \mathcal{Z}_k^i) f(\underline{x}_k \,|\, \mathcal{Z}_k^j)}{f(\underline{x}_k \,|\, \mathcal{Z}_k^i \cap \mathcal{Z}_k^j)} \;, \tag{12}$$

where $f(\underline{x}_k \,|\, \mathcal{Z}_k^{\{i,j\}})$ are the local estimates to be fused [8]. In state space, this requires the division by the density $f(\underline{x}_k \,|\, \mathcal{Z}_k^i \cap \mathcal{Z}_k^j)$ characterizing the common information. In information space, the log of this density has to be subtracted in (11) and, in particular, in the linear case this simply corresponds to the subtraction of the common information vector and matrix in (8).

Thus, in order to keep the benefits of the information space when dealing with arbitrary densities, we have to derive an approximation technique and parameterization of the participating densities that allows for the same simple manipulation of information, i.e., addition and subtraction, as in the linear information filter. For this purpose, we will consider the log-likelihood formulation (11) of data fusion. Of course, only densities that are non-zero almost everwhere can be fused this way, but this issue can easily be worked around by confining oneself to that subset Ω of the probability space, over which all densities to be fused are non-zero. Excluding parts containing "no information" is not a limitation, as the fusion result will also have zero probability mass there. The key idea behind this work is to approximate the log-densities in (11) over Ω by means of orthonormal bases. The choice of appropriate bases and especially the efficient online approximation of the likelihoods are the subject of the following section. The addition and subtraction of log-densities then turn into the simple addition and subtraction of coefficient vectors.

V. APPROXIMATION OF LOG-PROBABILITY DENSITIES

For the purpose of approximating the conditional log-densities in (11) and enabling an efficient fusion methodology, the participating functions will be represented by means of an orthonormal basis $\{\varphi_j\}_{j\geq 0}$ in the Hilbert space of square-integrable functions $L^2(\Omega)$ over the domain Ω. So, by confining oneself to an N-dimensional subspace of $L^2(\Omega)$ spanned by the subset $\{\varphi_j\}_{j=1}^{N}$, an approximation of $\ln f_k^{\mathrm{p}}$ is given by

$$\ln f_k^{\mathrm{p}}(\underline{x}_k) \approx \sum_{j=1}^{N} \alpha_j^{\mathrm{p}} \cdot \varphi_j(\underline{x}_k) \tag{13}$$

with the coefficients

$$\alpha_j^{\mathrm{p}} = \left\langle \ln f_k^{\mathrm{p}}, \varphi_j \right\rangle_{L^2} = \int_{\Omega} \ln f_k^{\mathrm{p}}(\underline{x}_k) \cdot \varphi_j(\underline{x}_k) \, d\underline{x}_k \;,$$

where $\langle \cdot, \cdot \rangle_{L^2}$ denotes the inner product in $L^2(\Omega)$. Analogously, the log-likelihoods $\ln f^i$ of the individual sensor nodes are approximated by

$$\ln f^i(\hat{\underline{z}}_k \,|\, \underline{x}_k) \approx \sum_{j=1}^{N} \gamma_j^i \cdot \varphi_j(\underline{x}_k) \ , \tag{14}$$

where the coefficients

$$\gamma_j^i = \big\langle \ln f^i(\hat{\underline{z}}_k \,|\, \cdot), \varphi_j \big\rangle_{L^2} = \int_\Omega \ln f^i(\hat{\underline{z}}_k \,|\, \underline{x}_k) \varphi_j(\underline{x}_k) \, d\underline{x}_k \tag{15}$$

are calculated by means of the inner product. An essential precondition for representing the densities this way is that their logarithms need to be square-integrable, i.e., $\ln f \in L^2(\Omega)$. In the following, we therefore restrict our discussion to probability densities lying in the set

$$\mathcal{P}(\Omega) := \{ f \in L^1(\Omega) \,|\, \mathrm{supp}(f) = \Omega, \ln f \in L^2(\Omega) \} \ .$$

This idea of representing probability densities in the information space imposes the following condition on the domain Ω: Since a probability density f is a positive and L^1-integrable function, its logarithm can only be square-integrable if Ω is bounded. At first glance, the boundedness of Ω seems to be very restrictive, but the domain can, of course, be chosen large enough, so that all "interesting" parts of the participating densities are captured. For instance, we will particularly apply the presented approach to Gaussian noise terms in the following example and in Section VII. All told, every density function in the following discussions is considered to be non-zero over Ω and the domain Ω to be bounded.

The online applicability of the presented idea strongly depends on a fast evaluation of the sensor log-likelihoods $\ln f^i(\hat{\underline{z}}_k^i \,|\, \cdot)$ for given measurements, i.e., on the calculation of the coefficient vectors $[\gamma_1^i, \ldots, \gamma_M^i]^{\mathrm{T}}$ from the inner product (15). We start with an example calculation for a nonlinear sensor model with a Gaussian pertubation.

Example: Gaussian measurement noise
For the sake of simplicity, we consider a one-dimensional state and a scalar-valued measurement function

$$\hat{z}_k = h(\boldsymbol{x}_k) + \boldsymbol{v} \ ,$$

where $\boldsymbol{v}$ is a zero-mean normally distributed random variable with

$$f^v(x) = \tfrac{1}{\sqrt{2\pi}\sigma} \exp\big\{ \tfrac{-(x)^2}{2\sigma^2} \big\} \ .$$

The log-likelihood then simplifies to

$$\ln f(\hat{z}_k \,|\, x_k) = \ln f^v(\hat{z}_k - h(x_k)) = -\tfrac{1}{2\sigma^2} (\hat{z}_k - h(x_k))^2 - C \ ,$$

where C is a constant. The coefficients (15) are now given by

$$\begin{aligned}
\gamma_j &= \langle \ln f^v(\hat{z}_k - h(\cdot)), \varphi_j \rangle_{L^2} \\
&= \int_\Omega \ln f^v(\hat{z}_k - h(x_k)) \varphi_j(x_k) dx_k \\
&= -\tfrac{1}{2\sigma^2} \int_\Omega (\hat{z}_k - h(x_k))^2 \varphi_j(x_k) dx_k - \widehat{C} \\
&= \tfrac{1}{\sigma^2} \hat{z}_k \int_\Omega h(x_k) \varphi_j(x_k) dx_k - \tfrac{1}{2\sigma^2} \int_\Omega h^2(x_k) \varphi_j(x_k) dx_k - \widetilde{C} \ ,
\end{aligned}$$

where all integrals can be computed in advance. All constant summands are subsumed in C. Finally, the coefficients are easily obtained by plugging concrete measurements into the above equation.

In order to deal with other more complex sensor models, the log-likelihoods $\ln f^i(\underline{z}_k \,|\, \underline{x}_k)$ can be interpreted as $\dim(\underline{z}_k) \times \dim(\underline{x}_k)$-dimensional functions in $L^2(\mathcal{Z} \times \Omega)$, where $\mathcal{Z}$ denotes the measurement space. Let $\{\psi_l\}_{l \geq 0}$ be an orthonormal basis in $L^2(\mathcal{Z})$. By virtue of the tensor product basis $\{\psi_l \otimes \varphi_j\}_{(l,j)} = \{\psi_l \cdot \varphi_j\}_{(l,j)}$, the log-likelihood $\ln f^i(\cdot \,|\, \cdot)$ can then be approximated by

$$\ln f^i(\underline{z}_k \,|\, \underline{x}_k) \approx \sum_{l=1}^{L} \sum_{j=1}^{N} \beta_{l,j}^i \cdot \psi_l(\underline{z}_k) \cdot \varphi_j(\underline{x}_k) \tag{16}$$

with

$$\begin{aligned}
\beta_{l,j}^i &= \big\langle \ln f(\cdot \,|\, \cdot), \psi_l \cdot \varphi_j \big\rangle_{L^2} \\
&= \int_{\mathcal{Z}} \int_\Omega \ln f(\underline{z}_k \,|\, \underline{x}_k) \psi_l(\underline{z}_k) \varphi_j(\underline{x}_k) \, d\underline{x}_k d\underline{z}_k \ .
\end{aligned}$$

These integrals can be computed numerically in advance. For an obtained observation $\hat{\underline{z}}_k^i$, the corresponding function values $\psi_l(\hat{\underline{z}}_k^i)$ have to be inserted into (16). Thus, the calculation (14) of the coefficients γ_j^i for $\ln f^i(\hat{\underline{z}}_k^i \,|\, \cdot)$ then becomes the sum

$$\gamma_j^i = \sum_{l=1}^{L} \beta_{l,j}^i \cdot \psi_l(\hat{\underline{z}}_k^i) \ ,$$

which can also be expressed as a matrix-vector multiplication. In Section II, we have highlighted that, for the calculations (11) in information space, the normalizing constants $\ln f(\mathcal{Z}_k \,|\, \mathcal{Z}_{k-1}, \mathcal{Z}_{k-2}, \ldots)$ are decoupled, which provides the foundation for efficient distributed fusion algorithms. More precisely, constants do not affect the parameters (8). Accordingly, we have to demand that $\ln f$ and $\ln f + C$ with constant offset C have the same coefficients (15), which implies

$$\begin{aligned}
\big\langle \ln f, \varphi_j \big\rangle_{L^2} &= \big\langle \ln f + C, \varphi_j \big\rangle_{L^2} \\
&= \big\langle \ln f, \varphi_j \big\rangle_{L^2} + \big\langle C, \varphi_j \big\rangle_{L^2}
\end{aligned}$$

for every $j = 1, \ldots, N$. This means that the subspace spanned by $\{\varphi_j\}_{j=1}^N$ is orthogonal to the set of constant functions, i.e., $\big\langle C, \varphi_j \big\rangle_{L^2} = 0$. In particular, the complete basis $\{\varphi_j\}_{j \geq 0}$ of $L^2(\Omega)$ needs to contain the constant basis element $\varphi_0 = \frac{1}{\sqrt{\mathrm{vol}(\Omega)}}$, so that all other basis elements are orthogonal, i.e., $0 = \langle \varphi_0, \varphi_i \rangle_{L^2}$ for $i \neq 0$. Hence, a wide variety of truncated basis expansions, e.g., Fourier, wavelet, and Legendre basis expansions, can be employed to approximate the log-densities.

With this spadework, the generalized fusion rule (11) for M sensors can now be written in terms of the corresponding coefficient vectors, i.e.,

$$[\alpha_1^{\mathrm{e}}, \ldots, \alpha_N^{\mathrm{e}}]^{\mathrm{T}} = [\alpha_1^{\mathrm{p}}, \ldots, \alpha_N^{\mathrm{p}}]^{\mathrm{T}} + \sum_{i=1}^{M} [\gamma_1^i, \ldots, \gamma_N^i]^{\mathrm{T}} \ ,$$

which yields the parameter vector $\underline{\alpha}_k^{\mathrm{e}} := [\alpha_1^{\mathrm{e}}, \ldots, \alpha_N^{\mathrm{e}}]^{\mathrm{T}}$ of the estimated log-density $\ln f_k^{\mathrm{e}}$. Apparently, this complies with the same simple fusion structure as in the linear case (8).

The two main prerequisites for the presented approach can be summarized as follows:

1) The bounded domain Ω is chosen such that the considered probability densities are non-zero over Ω.
2) The orthonormal basis $\{\varphi_j\}_{j=1}^N$ is orthogonal to constant functions.

As for the linear information filter, the benefits concerning distributed data fusion are paid with a more complicated prediction step. For nonlinear systems, the Chapman-Kolmogorov integral for predicting the conditional densities can in general not be solved in closed form and it is a state space formulation. So, either the log-densities are transformed back to state space at each prediction step or the prediction is expressed in information space. For deterministic or static systems, the second solution may be the best choice, but for stochastic sytems, a transformation to state space can be inevitable, i.e., the Chapman-Kolmogorov integral becomes

$$\ln f_{k+1}^{p}(\underline{x}_{k+1}) = \ln \int_{\Omega} f(\underline{x}_{k+1}|\underline{x}_k) \exp\{\ln f_k^{e}(\underline{x}_k)\}\, d\underline{x}_k \quad (17)$$

for the log-densities. For the simulations in this paper, the inverse transformation $\exp$ and the transformation $\ln$ of the integral are approximated. In order to calculate the first operation, the coefficients of $\ln f_k^e$ are approximately mapped to an L^2-basis in state space. In terms of the obtained coefficients, the integral can then be evaluated. Finally, the coefficients are mapped back to the information space basis. Altogether, the implementation of the prediction step depends on the actual system model and a general solution cannot be stated.

While this section focused on the log-densities and their approximations, the following section will consider the implications on the probability densities themselves, i.e., the implications on the state space.

VI. HILBERT SPACE STRUCTURE ON PROBABILITY DENSITIES

For efficient nonlinear state estimation in state space, truncated Fourier [4], [5] or wavelet [6] series expansions have also been successfully applied. In contrast to (13), the densities are then directly approximated by

$$f_k^{\mathrm{p}}(\underline{x}_k) \approx \sum_{j=1}^{N} a_j^{\mathrm{p}} \cdot \phi_j(\underline{x}_k) \quad (18)$$

with $a_j^{\mathrm{p}} = \langle f, \phi_j \rangle_{L^2}$. As expected, we face the same situation as in the linear case: While filtering is easy in information space and prediction is elaborate, the opposite holds for the state space approximation (18). Here, a reapproximation is required for every filtering step, which is a second issue besides the persistent problem of distributed fusion in state space. Another problem of the truncated series (18) is that it in general does not represent a valid probability density, i.e., it possibly does not integrate to one and it can even take negative function values. This is the point where an essential advantage of the information space representation becomes apparent. Transforming (13) back to state space always yields a valid probability density. Even a single basis function φ_j corresponds to a probability density

$$\frac{\exp\{\varphi_j(\underline{x}_k)\}}{\int_{\Omega} \exp\{\varphi_j(\underline{\xi}_k)\}\, d\underline{\xi}_k}$$

in state space and for the series expansion (13), we obtain

$$f(\underline{x}_k) \approx \frac{\prod_{j=1}^{N} \left(\exp\{\varphi_j(\underline{x}_k)\}\right)^{\gamma_j}}{\int_{\Omega} \prod_{j=1}^{N} \left(\exp\{\varphi_j(\underline{\xi}_k)\}\right)^{\gamma_j} d\underline{\xi}_k} \,.$$

That implies that the elementwise sum "$+$" and the scalar multiplication "$\cdot$" in information space correspond to the operations

$$f \oplus g := \frac{f(\cdot) \cdot g(\cdot)}{\int_{\Omega} f(\underline{\xi}) \cdot g(\underline{\xi})\, d\underline{\xi}} \quad (19)$$

and

$$a \odot f := \frac{f^a(\cdot)}{\int_{\Omega} f^a(\underline{\xi})\, d\underline{\xi}} \quad (20)$$

in state space, respectively. As a generalization of the Aitchison geometry [13], a Hilbert space $A(\Omega)$ on probability densities has been developed by means of these operations [14], where

$$\langle f, g \rangle_{A(\Omega)} := \tfrac{1}{2\mathrm{vol}(\Omega)} \int_{\Omega} \int_{\Omega} \ln \frac{f(\underline{x})}{f(\underline{y})} \ln \frac{g(\underline{x})}{g(\underline{y})}\, d\underline{x}\, d\underline{y}$$
$$= \int_{\Omega} \ln f(\underline{x}) \ln g(\underline{x})\, d\underline{x} - \tfrac{1}{\mathrm{vol}(\Omega)} \int_{\Omega} \ln f(\underline{x})\, d\underline{x} \int_{\Omega} \ln g(\underline{x})\, d\underline{x}$$

is the inner product in $A(\Omega)$. This product induces the vector space norm

$$\|f\|_{A(\Omega)} = \left[\int_{\Omega} \left(\ln f(\underline{x})\right)^2 d\underline{x} - \tfrac{1}{\mathrm{vol}(\Omega)} \left(\int_{\Omega} \ln f(\underline{x})\, d\underline{x} \right)^2 \right]^{\frac{1}{2}}$$

on $A(\Omega)$, which has, for example, been applied as an information measure for sensor management in [15]. Simply speaking, the spaces $A(\Omega)$ and $L^2(\Omega)$ can be related by the isometry $\ln : A(\Omega) \to L^2(\Omega)$, with which it can be proven that $A(\Omega)$ is a Hilbert space. For the construction of bases for product spaces $A(\Omega) \times A(\Psi)$, the tensor product used in (16) becomes

$$f \otimes_{A(\Omega)} g = \frac{\exp\{\ln f \cdot \ln g\}}{\int_{\Omega} \int_{\Psi} \exp\{\ln f(\underline{x}) \cdot \ln g(\underline{y})\}\, d\underline{y} d\underline{y}} \,.$$

Of course, likelihoods are not probability densities and therefore no elements of $A(\Omega)$, but this does not pose a problem due to the normalization in (19). In conclusion, the information space considered as the function space L^2 of square-integrable functions is strongly related to a Hilbert space structure in the state space. The vector space addition and multiplication correspond to the Bayesian update (19) and the power transformation (20), which is essentially a weighting of information. All approximation techniques in

this space are compliant to operations on probabilities and therefore yield valid probability densities. In other words, complicated probability densities are approximated by sums of simpler probability densities, which makes the information space representation particularly attractive for stochastic state estimation.

VII. SIMULATIONS

This section exemplifies the presented approach by means of a cubic sensor that measures a static and a dynamic system. In the first case, the nonlinear information filter even provides optimal estimation results. Also, the communication between two sensor nodes, which compute local estimates, and the removal of common information is discussed. The second case proves the applicability to dynamic state evolutions.

1) Static System and Cubic Sensor: A constant system state, i.e., $x_{k+1} = a(x_k) = 1$, is measured by a cubic sensor

$$\hat{z}_k = h(x_k) = x_k^3 + \mathcal{N}(0, 0.05^2)$$

over a horizon of 50 time steps. The observations are corrupted by zero-mean Gaussian noise with a standard deviation of 0.05. The prior knowledge is modeled by a Gaussian density with mean 1.5 and standard deviation 0.01. In Fig. 2, the estimation results of an optimal Bayesian estimator, the nonlinear information filter, and the extended Kalman filter are compared. The Bayesian estimator serves as ground truth

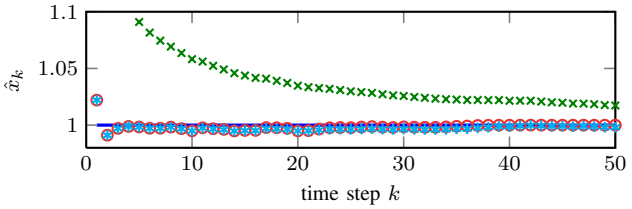

Fig. 2. Simulation run over 50 time steps. Blue: true system state; green crosses: extended Kalman filter; red circles: optimal Bayesian esimator; cyan stars: nonlinear information filter

in this simulation. The estimated density at each time instant is obtained through a numerical computation of the fusion rule (5). For that purpose, the densities have been discretized over the interval $[-2, 3]$. The means of these densities are depicted as circles. Obviously, they capture the true state rather quickly. On the contrary, the estimates provided by the extended Kalman filter converge very slowly to the true state. Due to the bad prior and linearization errors, the extended Kalman filter underestimates the true uncertainty and trusts its own estimates more than it trusts new measurements. The stars in Fig. 2 represents the means of the densities provided by the nonlinear information filter, where Legendre polynomials up to order 6 are employed for the truncated series approximation over the interval $[-2, 3]$. As shown in the example in Sec. V, the log-likelihood to be approximated is given by

$$\ln f(\hat{z}_k \,|\, x_k) = -\tfrac{1}{2\sigma^2}(\hat{z}_k - x_k^3)^2 - C \ ,$$

which is itself a polynomial of degree six and therefore, the first seven Legendre polynomials suffice to represent this log-likelihood even optimally. The Legendre polynomial of order zero, the constant element, is omitted, since it only corresponds to the normalizing constant. Thus, the log-densities are represented by six-dimensional coefficient vectors and the filtering at each time instant simply requires the addition of these vectors. The estimates apparently coincide with the optimal Bayesian results.

2) Two-sensor Data Fusion: For illustrating the benefits of the nonlinear information filter with regard to distributed estimation, we consider two sensor nodes with different priors. Both nodes observe the state independently by means of cubic sensors and calculate local estimates in information space represented by the same Legendre basis as above. Local estimates, where no communication between the nodes takes place, are depicted as dashed lines in Fig. 3. The solid

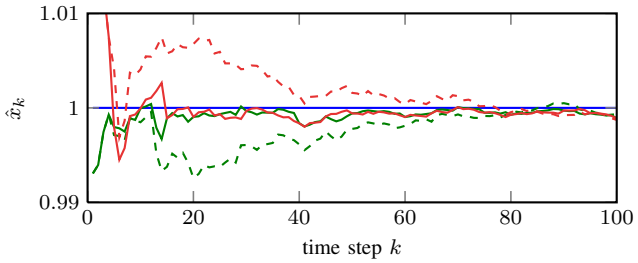

Fig. 3. Simulation with two sensor nodes. Dashed red and green lines denote local estimates without communication. Solid red and green lines shows local estimates that are communicated at every fifth time step. Common information is removed, when the local estimates are fused.

lines correspond to a second simulation run, where the local estimates are exchanged at every fifth time instant. For fusing the local estimates, the information space formulation of (12) has been employed. In order to avoid double counting, the information that both nodes have in common is stored until the next communication takes place, where it can then be subtracted from the fusion result. Hence, the fusion process only comprises vector additions and subtractions.

3) Trigonometric System and Cubic Sensor: As a dynamic example, we consider the dynamic system

$$x_{k+1} = f(x_k) = \cos\left(\frac{\pi}{2} x_k\right) + \mathcal{N}(0, 0.2^2) \ ,$$

which is again observed by means of a cubic sensor with the same parameters as above. Again, the extended Kalman filter and the nonlinear information filter with six Legendre polynomials are compared with the optimal Bayesian estimator. Fig. 4 presents the results of a single simulation

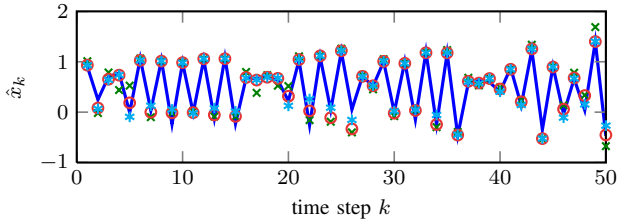

Fig. 4. Blue line: true system state; green crosses: extended Kalman filter; red circles: Bayesian estimator; cyan stars: nonlinear information filter

run. Due to the approximate prediction step, the estimates by the information filter are no longer optimal, but they are still close to the results of the Bayesian estimator. In order to solve the prediction step efficiently, the mappings exp

and $\ln$ in (17) are discretized and precalculated in terms of the coefficients of $\ln f_k^{\mathrm{e}}$. The transition density $f(x_{k+1}|x_k)$ is approximated by an L^2-basis in state space, so that the integral becomes a matrix-vector multiplication. Fig. 5

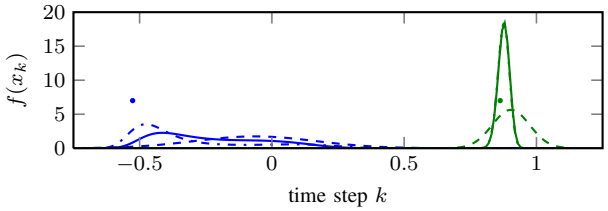

Fig. 5. Densities at time steps $k = 25$ (blue) and $k = 26$ (green). The true states are depicted as dots. The solid lines correspond to the true densities of the state estimate. The dashed lines represents the Gaussian densities, which correspond the Kalman filter estimates. The results from approximating the densities in information space by Legendre polynomials are dash-dotted.

depicts the corresponding densities of the state estimates at two specific time steps. Especially, the green plot shows that the true density is captured well through the approximation in information space, where only six basis elements have been used.

VIII. CONCLUSIONS AND FUTURE WORKS

Instead of approximating probability densities directly, we have shown in this paper that it can be beneficial to approximate their logarithms. The Bayesian fusion formulas then turn into the sums of log-densities, which in particular eases the development of distributed fusion architectures. Furthermore, the normalizing factor in Bayes' rule becomes an additive constant after applying the logarithm. This constant can then be ignored in the information space, since it is implicitly calculated, when the log-densities are transformed back to probability densities, i.e., into state space. As a concrete approximation technique for log-densities, we suggest to employ orthonormal series expansions. Orthonormal bases that contain the constant basis element are especially well-suited, so that the normalizing constant does not affect the coefficients of the non-constant basis elements. For static measurement models, the coefficients of the log-likelihoods can be computed in advance and, for specific observations, they can then easily be evaluated. We pointed out that the orthonormal basis expansions of log-densities impose a different vector space structure on the probability densities themselves. This vector space provides the advantage that addition and scalar multiplication of probability densities again yield valid probability densities.

The weak spot of the information space representation is the difficult prediction step. Not only the Chapman-Kolmogorov integral needs to be computed, but also this integral has to be expressed in terms of the coefficient vectors or, alternatively, complicated transformations between state space and information space are required. Prospective research will particularly focus on providing feasible solutions for the prediction step in terms of the information space. For dynamic systems, also the boundedness of the domain Ω is a problem, if the state moves outside the bounds. A second problem is related to numerical instabilities, which arise when probability densities tend to zero somewhere and their logarithms approaches minus infinity accordingly. A solution to these issues can be an online adaption of the domain Ω, in order to exclude probability masses close to zero. A thorough consideration of the related Hilbert space structure on probability densities and the corresponding norm also appears promising.

IX. ACKNOWLEDGMENTS

This work was partially supported by the German Research Foundation (DFG) within the Research Training Group GRK 1194 "Self-organizing Sensor-Actuator-Networks".

REFERENCES

[1] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Transactions of the ASME - Journal of Basic Engineering*, no. 82 (Series D), pp. 35–45, 1960.

[2] M. E. Liggins, D. L. Hall, and J. Llinas, Eds., *Handbook of Multisensor Data Fusion: Theory and Practice*, 2nd ed., ser. The Electrical Engineering and Applied Signal Processing Series. CRC Press, 2009.

[3] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*, ser. Statistics for Engineering and Information Science. Springer-Verlag, 2001, ch. 1, pp. 6–12.

[4] R. Beard, J. Gunther, J. Lawton, and W. Stirling, "Nonlinear Projection Filter based on Galerkin Approximation," *AIAA Journal of Guidance Control and Dynamics*, vol. 22, no. 2, pp. 258–266, 1999.

[5] D. Brunn, F. Sawo, and U. D. Hanebeck, "Nonlinear Multidimensional Bayesian Estimation with Fourier Densities," in *Proceedings of the 2006 IEEE Conference on Decision and Control (CDC 2006)*, San Diego, California, Dec. 2006, pp. 1303–1308.

[6] A. Hekler, M. Kiefel, and U. D. Hanebeck, "Nonlinear Bayesian Estimation with Compactly Supported Wavelets," in *Proceedings of the 2010 IEEE Conference on Decision and Control (CDC 2010)*, Atlanta, Georgia, Dec. 2010.

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

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

[9] A. Gasparri, F. Pascucci, and G. Ulivi, "A Distributed Extended Information Filter for Self-Localization in Sensor Networks," in *Proceedings of IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2008)*, Sept. 2008, pp. 1–5.

[10] L.-L. Ong, M. Ridley, B. Upcroft, S. Kumar, T. Bailey, S. Sukkarieh, and H. Durrant-Whyte, "A Comparison of Probabilistic Representations for Decentralised Data Fusion," in *Proceedings of the 2005 International Conference on Intelligent Sensors, Sensor Networks and Information Processing Conference (ISSNIP 2005)*, Dec. 2005, pp. 187–192.

[11] L.-L. Ong, B. Upcroft, M. Ridley, T. Bailey, S. Sukkarieh, and H. Durrant-Whyte, "Consistent Methods for Decentralised Data Fusion using Particle Filters," *Proceedings of the 2006 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2006)*, pp. 85–91, 2006.

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

[13] J. Aitchison, *The Statistical Analysis of Compositional Data*, ser. Monographs on Statistics and Applied Probability. Chapman Hall Ltd., London (UK), 1986.

[14] J. J. Egozcue, J. L. Diaz-Barrero, and V. Pawlosky-Glahn, "Hilbert Space of Probability Density Functions Based on Aitchison Geometry," *Acta Mathematica Sinica, English Series*, vol. 22, no. 4, pp. 1175–1182, July 2006.

[15] D. Lyons, B. Noack, and U. D. Hanebeck, "A Log-Ratio Information Measure for Stochastic Sensor Management," in *Proceedings of the IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing (SUTC 2010)*, Newport Beach, California, June 2010.