Bayesian Estimation with Uncertain Parameters of Probability Density Functions

Vesa Klumpp and **Uwe D. Hanebeck**
Intelligent Sensor-Actuator-Systems Laboratory (ISAS),
Institute for Anthropomatics,
Universität Karlsruhe (TH), Germany.
klumpp@ira.uka.de, uwe.hanebeck@ieee.org

Abstract − *In this paper, we address the problem of processing imprecisely known probability density functions by means of Bayesian estimation. The imprecise knowledge about probability density functions is given as stochastic uncertainty about their parameters. The proposed processing of this special density in a Bayesian estimator is accomplished by reinterpretation of the filter and prediction equations. Here, the parameters are treated as a higher order state, which can be processed by Bayesian estimation techniques. For state estimation, this avoids the need to select specific values for unknown parameters and, thus, allows the processing of all potential parameters at once. The proposed approach further allows the use of imprecisely known model equations for measurement and state prediction by the same principle.*

Keywords: Bayesian state estimation, Hierarchical density, Imprecise probability, Uncertain systems

1   Introduction

In many technical applications, knowledge about the internal, not directly observable, state of a physical system is required. For these problems, the Bayesian estimator is a possible approach. Typical examples for Bayesian state estimation include localization of vehicles, tracking of aircrafts, or speech recognition.

Common approaches to the Bayesian estimation problem are the Kalman Filter and extensions to it, like the Extended Kalman Filter or the Unscented Kalman Filter [1], in which the state estimate is given as a Gaussian density. Another important approach are particle filters [2], where the state estimate is given as a sample set. These sample-based methods shine through their simplicity and the ability to process arbitrary system models. Their substantive drawback is the huge amount of samples needed, in order to obtain a reasonable density representation. Other density representations, like Gaussian mixture densities, which are used in the Hybrid Density Filter [3], or sliced Gaussian mixture densities used in the Sliced Gaussian Mixture Filter [4] exist.

Usually, in continuous state space, the density representation is parametric, i.e., the estimation algorithms work on parameters defining the density function. These can be parameters of Gaussian densities, Gaussian mixture densities, particles, or edgeworth series, for example.

In general, these parameters describe all occuring densities, like noise densities and prior densities. Furthermore, the measurement and prediction model equations can be described by means of parameters. A problem arising in estimation applications is that the true parameters, representing the underlying system and probability density functions, are often hard or even impossible to find. Therefore, arbitrary values are chosen. Due to the sensitivity of the estimation result regarding to these parameters, the quality of the result depends on these assumptions. This means, by choosing model parameters that differ from the truth, the resulting state estimation differs from the true system state to a certain degree.

Different approaches to the problem of (partially) unknown parameters exist. The most common approach is state augmentation [5], where the state of the system is expanded by unknown parameters. Here, the unknown parameters can be estimated, but only under the assumption that a proper model of the (time-variant) parameters is given.

Another solution to this problem is considering every plausible case, which leads to sets of assumptions and therefore to sets of probability density functions. Several works cope with sets of probabilities. Often, convex sets of probability density functions [6] are utilized. An example of Bayesian state estimation with convex sets of probability density functions is [7]. Other ways of defining sets of probability density functions are the famous theory of coherent previsions [8], distribution bands [9], or the class of ϵ-contaminations, just to name

a few.

Instead of allowing a set of densities, whose elements are all equally favorable and possible, a distribution over these densities can express preferences about specific densities. This is often utilized in robust Bayesian analysis [10]. In that field, hierarchical Bayesian approaches [11] are used to process densities, whose parameters are described by random variables. Another way to define uncertainty about densities in a Bayesian framework is by means of Dirichlet processes. An application of Dirichlet processes, which is related to the work presented here, is state estimation of linear systems under consideration of unknown system or measurement noise [12].

In this paper, we consider a special density representation, so-called type-2 densities, which are named after type-2 fuzzy sets [13]. Here, the parameters of the densities are described by densities, too. The difference to usual hierarchical Bayesian approaches lies in the processing within state estimation, as described later.

The paper is structured as follows. In Section 2, a short introduction to Bayesian estimation is given and the problems herein are discussed. Section 3 introduces this special method of processing type-2 density functions in a Bayesian estimator. Section 4 illustrates the formal framework by means of examples. The paper closes with conclusions and an outlook onto future work.

2 Problem Formulation

In this section, Bayesian state estimation is explained and the special problem of imprecisely known parameters within the estimation procedure is addressed.

2.1 Bayesian State Estimation

In Bayesian estimation theory, the measurement model describes the relation between the not directly observable internal state $\underline{\boldsymbol{x}}_k$ and a measurement $\underline{\hat{y}}_k$

$$\underline{\hat{y}}_k = \underline{h}_k\left(\underline{\boldsymbol{x}}_k\right) + \underline{\boldsymbol{v}}_k \tag{1}$$

at discrete time steps k. The measurement model $\underline{h}_k$ is time-variant and describes the relation between the internal state and the measurement. Here, we consider the special case of an additive stochastic noise term $\underline{\boldsymbol{v}}_k$. The system equation describes the evolution of the system state over one time step k to $k+1$ according to

$$\underline{\boldsymbol{x}}_{k+1} = \underline{a}_k\left(\underline{\boldsymbol{x}}_k\right) + \underline{\boldsymbol{w}}_k \quad, \tag{2}$$

with the additional noise term $\underline{\boldsymbol{w}}_k$. Note, that throughout this paper, random variables are bold face ($\boldsymbol{x}$), vectors are underlined ($\underline{x}$), and the state space is denoted by Ω. Random variables are described by probability density functions ($\boldsymbol{x} \sim f(x)$).

The measurement step, according to the measurement equation (1) is calculated using the well known Bayes law by

$$f^e(\underline{x}_k) = \frac{f^p(\underline{x}_k) \cdot f^L(\underline{x}_k)}{\int_\Omega f^p(\underline{\xi}_k) \cdot f^L(\underline{\xi}_k) \, \mathrm{d}\underline{\xi}_k} \quad. \tag{3}$$

The prior probability density function $f^p(\underline{x}_k)$ is processed with the likelihood

$$f^L(\underline{x}_k) = f^v\left(\underline{\hat{y}}_k - \underline{h}_k(\underline{x}_k)\right)$$

that depends on the measurement noise density $f^v(\underline{x}_k)$ and the measurement equation (1). The resulting estimated density $f^e(\underline{x}_k)$ can be further handled in the next step. Equation (2) is processed using the Chapman-Kolmogorov equation

$$f(\underline{x}_{k+1}) = \int_\Omega f^T(\underline{x}_{k+1}|\underline{x}_k) \cdot f^e(\underline{x}_k) \, \mathrm{d}\underline{x}_k \tag{4}$$

and the transition density

$$f^T(\underline{x}_{k+1}|\underline{x}_k) = f^w\left(\underline{x}_{k+1} - \underline{a}_k(\underline{x}_k)\right)$$

that depends on the system noise $\underline{\boldsymbol{w}}_k$ with probability density function $f^w(\underline{x}_k)$ and the system equation (2). In this paper, we assume that the filter step (3) and the prediction step (4) are processed in an alternating way without loss of generality. This procedure is visualized in Figure 1 (a).

A problem that arises in Bayesian estimation is the proper choice of model parameters, especially the characterization of the noise terms and the prior state estimate. Usually, they are unknown or cannot be determined exactly due to the lack of knowledge about the true system. A related problem is the processing of imprecise system or measurement models. This occurs when dealing with black box systems, whose behaviour can only be partially reconstructed by observations, or only imprecise expert knowledge is available. In this case, parameters in the measurement or system model equations remain unspecified or are only imprecisely known.

2.2 Novelties of the paper

The novelty of this approach is the application of Bayesian state estimation for densities with uncertain parameters, i.e., the parameters are described by random variables. This results in a hierarchical density representation that defines densities over parameters of densities over the state space and allows a new interpretation of the filter and prediction equations of the Bayesian estimator: Now, these equations are regarded as generative forward models of the uncertain parameters of the densities. By this means, every possible and probable parameterization can be processed simultaneously by the processing of these models.
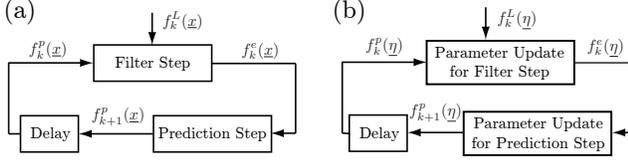
Figure 1: (a) Standard Bayesian estimator, (b) Bayesian estimator working on uncertain density parameters that are processed using a generative forward model according to the standard Bayesian estimator.

3 Bayesian Estimator with Type-2 Densities

This section describes the Bayesian estimator with type-2 densities and derives this basic idea step by step. First, the connection between different types of stochastic uncertainties is explained. Figure 2 depicts the structure of the different types of stochastic uncertainties that are considered in this work. In Figure 2 (a), no uncertainty is considered, i.e., all variables over the state space are deterministic.

Adding stochastic uncertainty about equations and variables in the state space leads to Figure 2 (b). Here, knowledge about the state space is represented by random variables $\underline{x}$, which are distributed according to their densities $f(\underline{x})$. When dealing with continuous density functions, e.g., in a Bayesian estimator, they are described by their parameter vectors $\underline{\eta} \in \Psi^D$, as seen in Figure 2 (c). Similarly to the density parameters, the model equations can be parameterized, too. The parameter space of model parameters is denoted by Ψ^E.

The processing of densities according to a filter step or a prediction step is a mapping of deterministic parameters in the corresponding parameter spaces. The filter step in Bayesian estimation (3) maps the parameters $\underline{\eta}_k^p \in \Psi^D$ from the prior density and the parameters $\underline{\eta}_k^v \in \Psi^D$ of the measurement noise, considering the model parameter vector $\underline{\eta}_k^h \in \Psi^E$ of (1), to the parameters $\underline{\eta}_k^e \in \Psi^D$ of the estimated density. This mapping in parameter space is described by the function $\underline{\mathcal{H}}_k : \Psi^D \times \Psi^D \times \Psi^E \rightarrow \Psi^D$ with

$$\underline{\eta}_k^e = \underline{\mathcal{H}}_k\left(\underline{\eta}_k^p, \underline{\eta}_k^v, \underline{\eta}_k^h\right) \quad . \tag{5}$$

An example of this mapping are the Kalman filter equations of mean and covariance for filtering in linear systems.

The mapping of the parameters for the prediction step is defined analogously. Here, the parameters of the estimated density $\underline{\eta}_k^e$, the density function of the system noise $\underline{\eta}_k^w \in \Psi^D$, and the model parameters $\underline{\eta}_k^a \in \Psi^E$ are processed according to the mapping $\underline{\mathcal{A}}_k : \Psi^D \times \Psi^D \times \Psi^E \rightarrow \Psi^D$ with

$$\underline{\eta}_{k+1}^p = \underline{\mathcal{A}}_k\left(\underline{\eta}_k^e, \underline{\eta}_k^w, \underline{\eta}_k^a\right) \quad . \tag{6}$$
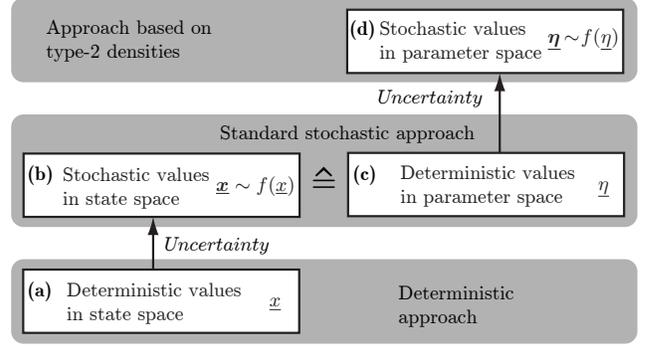

Figure 2: Different levels of uncertainty. In the bottom layer, no stochastic uncertainty is considered at all. In the middle layer, density functions over the state space are considered. The next higher layer deals with type-2 probability density functions.

The result is the parameter vector of the predicted density $\underline{\eta}_{k+1}^p \in \Psi^D$. The deterministic mappings (5) and (6) define the processing of parameters of density functions resulting in a standard Bayesian estimator over densities in state space as visualized in Figure 2 (c).

Now, the uncertainty over the parameters can be added, as depicted in Figure 2 (d). Here, the deterministic parameter vectors $\underline{\eta}$ are extended to random vectors $\underline{\boldsymbol{\eta}}$ in the parameter space. The parameter mappings are now interpreted as forward models of the parameters, which are used for stochastic prediction. The random parameter vector is now regarded as a state estimate in parameter space, which is mapped through the model equations for the filter step and the prediction step. Equations (5) and (6) are applied to random variables over parameters, leading to the stochastic equations

$$\underline{\boldsymbol{\eta}}_k^e = \underline{\mathcal{H}}_k\left(\underline{\boldsymbol{\eta}}_k^p, \underline{\boldsymbol{\eta}}_k^v, \underline{\boldsymbol{\eta}}_k^h\right) \quad , \tag{7}$$

$$\underline{\boldsymbol{\eta}}_{k+1}^p = \underline{\mathcal{A}}_k\left(\underline{\boldsymbol{\eta}}_k^e, \underline{\boldsymbol{\eta}}_k^w, \underline{\boldsymbol{\eta}}_k^a\right) \quad . \tag{8}$$

This converts the given processing rule to a Bayesian estimator working on hierarchical type-2 probability density functions with imprecise input densities and model parameters. The output of such an estimator is a type-2 probability density function, too. Thus, a type-2 density is defined as follows.

Definition: *A type-2 density function is a parameterized density function over state space, whose parameters $\underline{\eta}$ are described by a density function $f(\underline{\eta})$ and are processed by special stochastic forward models $\underline{\mathcal{H}}_k$, $\underline{\mathcal{A}}_k$. These model equations are identical to the equations of the Bayesian state estimator working on the densities over state space for the filter and prediction step.*

For this reason, there is a coupling between the type-2 densities, the parameterization, and the estimator working on state space. It is noteworthy that both equations define stochastic forward models of the pa-

rameter uncertainties. In usual Bayesian state estimation, the measurement equation is processed using Bayes' law. Due to the update of the state estimate according to the measurement, the posterior density $f(\underline{\eta})$ over the parameter space has to be updated, too. This will be explained in the next subsection.

Figure 1 illustrates the conversion to the processing in parameter space. On the left hand side, in Figure 1 (a), the standard Bayesian estimator consisting of filter step and prediction step is visualized. Both processing steps exchange information by means of density functions over the state space $f(\underline{x})$. The conversion of the standard Bayesian estimator is shown in Figure 1 (b). Here, the information is given as probability density functions over the parameter space $f(\underline{\eta})$.

Processing of Type-2 Densities

The processing of these models in parameter space according to the given forward equations are nonlinear stochastic prediction problems. They can be formally stated by the Chapman-Kolmogorov equation

$$f(\underline{\eta}_k^e) = c_k \cdot \int_{\Psi^E} \int_{\Psi^D} \int_{\Psi^D} f_k^T(\underline{\eta}_k^e|\underline{\eta}_k^p,\underline{\eta}_k^v,\underline{\eta}_k^h) \cdot w(\underline{\eta}_k^p,\underline{\eta}_k^v,\underline{\eta}_k^h)$$
$$\cdot \left\{ f(\underline{\eta}_k^p)\, f(\underline{\eta}_k^v)\, f(\underline{\eta}_k^h) \right\}\, \mathrm{d}\underline{\eta}_k^p\, \mathrm{d}\underline{\eta}_k^v\, \mathrm{d}\underline{\eta}_k^h \;, \quad (9)$$

$$f(\underline{\eta}_{k+1}^p) = \int_{\Psi^E} \int_{\Psi^D} \int_{\Psi^D} f_k^T(\underline{\eta}_{k+1}^p|\underline{\eta}_k^e,\underline{\eta}_k^w,\underline{\eta}_k^a)$$
$$\cdot \left\{ f(\underline{\eta}_k^e)\, f(\underline{\eta}_k^w)\, f(\underline{\eta}_k^a) \right\}\, \mathrm{d}\underline{\eta}_k^e\, \mathrm{d}\underline{\eta}_k^w\, \mathrm{d}\underline{\eta}_k^a \;. \quad (10)$$

The weighting factor

$$w(\underline{\eta}_k^p,\underline{\eta}_k^v,\underline{\eta}_k^h) = f(\hat{\underline{y}}_k|\underline{\eta}_k^p,\underline{\eta}_k^v,\underline{\eta}_k^h) =$$
$$\int_{\Psi^E} f(\hat{\underline{y}}_k|\underline{x}_k^p,\underline{\eta}_k^p,\underline{\eta}_k^v,\underline{\eta}_k^h) \cdot f(\underline{x}_k^p|\underline{\eta}_k^p,\underline{\eta}_k^v,\underline{\eta}_k^h)\, \mathrm{d}x_k^p \quad (11)$$

in (9) affects the posterior probability of the parameterization of $\underline{\eta}_k^e$ given the measurement $\hat{\underline{y}}_k$. This is due to the normalization in the denominator of the density over state space in (3), which is equal to $w(\underline{\eta}_k^p,\underline{\eta}_k^v,\underline{\eta}_k^h)$. Here, in this approach, the density over parameter space has to be normalized. This is stated by the normalization constant c_k, which is omitted here, in order to keep the equations more comprehensible.

The transition densities for measurement and prediction

$$f_k^T(\underline{\eta}_k^e|\underline{\eta}_k^p,\underline{\eta}_k^v,\underline{\eta}_k^h) = \delta(\underline{\eta}_k^e - \mathcal{H}_k(\underline{\eta}_k^p,\underline{\eta}_k^v,\underline{\eta}_k^h)) \text{ and}$$
$$f_k^T(\underline{\eta}_{k+1}^p|\underline{\eta}_k^e,\underline{\eta}_k^w,\underline{\eta}_k^a) = \delta(\underline{\eta}_{k+1}^p - \mathcal{A}_k(\underline{\eta}_k^e,\underline{\eta}_k^w,\underline{\eta}_k^a))$$

are defined with the multidimensional Dirac δ distribution. The parameter densities $f(\underline{\eta}_k^p)$, $f(\underline{\eta}_k^v)$, and $f(\underline{\eta}_k^h)$ for the filter step and $f(\underline{\eta}_k^e)$, $f(\underline{\eta}_k^w)$, and $f(\underline{\eta}_k^a)$ for the
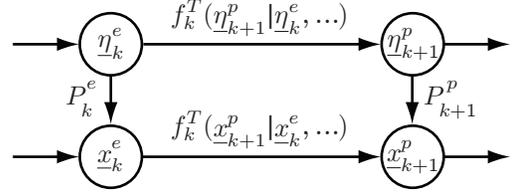

Figure 4: Hierarchical model of the parameters $\underline{\eta}$ and the state estimate $\underline{x}$ for the prediction step.

prediction step over the different input vectors are independent. This allows simpler density representation and processing of the equations above. These equations can be solved by nonlinear estimators, like the one proposed in [4].

Compatibility to Bayesian Approach

The processing of type-2 densities can be regarded as hierarchical Bayesian approach to state estimation with uncertain parameters, as depicted in Figure 4. In the prediction step, the estimated state variable $\underline{x}_k^e$ is processed according to the transition density $f_k^T(\underline{x}_{k+1}^p|\underline{x}_k^e,\ldots)$ to the predicted state $\underline{x}_{k+1}^p$. The densities $f(\underline{x}_k^e)$ and $f(\underline{x}_{k+1}^p)$ are parameterized according to the parameter vectors $\underline{\eta}_k^e$ and $\underline{\eta}_{k+1}^p$, which are given by $P_k^e = f(\underline{x}_k^e|\underline{\eta}_k^e)$ and $P_{k+1}^p = f(\underline{x}_{k+1}^p|\underline{\eta}_{k+1}^p)$. Note, that the parameters $\underline{\eta}_k^a$ and $\underline{\eta}_k^w$ of the mapping are omitted for simplicity and only the prediction step is shown. The predicted density over state space can be calculated according to

$$f(\underline{x}_{k+1}^p) = \int_{\Psi^D} \int_{\Omega} f(x_{k+1}^p,\underline{x}_k^e,\underline{\eta}_k^e)\, \mathrm{d}\underline{x}_k^e\, \mathrm{d}\underline{\eta}_k^e$$
$$= \int_{\Psi^D} \int_{\Omega} f(\underline{\eta}_k^e) \cdot f(\underline{x}_k^e|\underline{\eta}_k^e) \cdot f_k^T(x_{k+1}^p|x_k^e)\, \mathrm{d}\underline{x}_k^e\, \mathrm{d}\underline{\eta}_k^e \;,$$

which is equal to

$$f(\underline{x}_{k+1}^p) = \int_{\Psi^D} \int_{\Psi^D} f(x_{k+1}^p,\underline{\eta}_{k+1}^p,\underline{\eta}_k^e)\, \mathrm{d}\underline{x}_k^e\, \mathrm{d}\eta_k^e$$
$$= \int_{\Psi^D} \int_{\Psi^D} f(\underline{\eta}_k^e) \cdot f_k^T(\underline{\eta}_{k+1}^p|\underline{\eta}_k^e) \cdot f(\underline{x}_{k+1}^p|\underline{\eta}_{k+1}^p)\, \mathrm{d}\underline{\eta}_k^e\, \mathrm{d}\underline{\eta}_{k+1}^p.$$

For known parameterizations, the transition densities describe the same model: $f_k^t(x_{k+1}^p|x_k^e)$ describes the transition of the state estimation, whereas $f_k^t(\eta_{k+1}^p|\underline{\eta}_k^e)$ describes the mapping of the parameters of the state estimation, which gives in the same result in state space by definition of (8). Thus, the principle of type-2 densities is compatible to the hierarchical Bayesian approach.

For the Bayesian filter step, the additional factor w (11) has to be introduced.

Additional Parameter Noise Term

The models (7) and (8) are exact, i.e., there is no additional noise term. However, using an additive stochastic
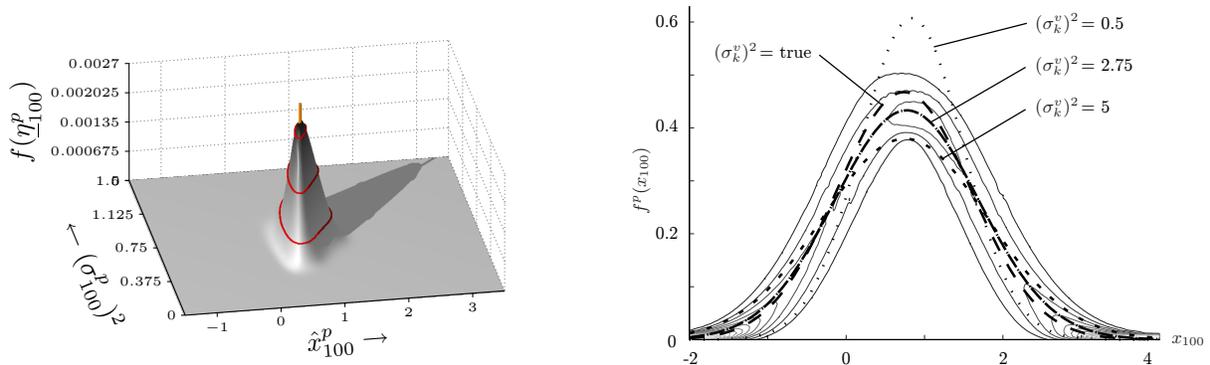
Figure 3: Simulation result at time step 100. Left: Density $f(\underline{\eta}^p_{100})$ over parameter space. Estimation with the true parameter is marked with a bar. Right: Visualization of $f(\underline{\eta}^p_{100})$ in state space as contour plot and estimation results with selected discrete parameter values.

noise term can have several advantages. The first one lies in the fact that by using noise in the parameter model, this model can be regarded as imprecise, too. It is the case when the filter and prediction equations (5) and (6) are not exact, e.g., they implicitly perform an approximation of the resulting density. Especially for nonlinear systems, an approximation of the posterior density function is inevitable in most cases. With this approach, imprecision in the fusion equations can be described by uncertainty in the parameters of the result. Thus, the parameter model becomes a model equation with type-2 stochastic uncertainty. The second advantage is simply to obtain smoother density functions over the parameter space. This allows prediction procedures that make use of system noise, which can reduce the processing effort considerably, e.g., [3]. The model equations (5) and (6) are then given as

$$\underline{\eta}^e_k = \underline{\mathcal{H}}_k\left(\underline{\eta}^p_k, \underline{\eta}^v_k, \underline{\eta}^h_k\right) + \underline{\mathcal{V}}_k \ ,$$

$$\underline{\eta}^p_{k+1} = \underline{\mathcal{A}}_k\left(\underline{\eta}^e_k, \underline{\eta}^w_k, \underline{\eta}^a_k\right) + \underline{\mathcal{W}}_k \ ,$$

with random noise terms $\underline{\mathcal{V}}_k$ and $\underline{\mathcal{W}}_k$.

Note, that these models of the density parameters heavily depend on the parameterization of the probability density functions in the underlying state space. By choosing a different density representation or another implementation of a Bayesian estimator that is incorporated into the parameter mapping, the overall prediction problems (9) and (10) change because their transition densities are a stochastic model of equations (1) and (2).

Different Interpretation of Type-2 Densities
The vector $\underline{\eta}$ sets the parameters for the (conditional) probability density function $f(\underline{x}|\underline{\eta})$. The parameter vector $\underline{\eta}$ can be regarded as constant, so $f(\underline{x}|\underline{\eta})$ is a simple density over the state space Ω. If we build the joint of state space and parameter space $\Omega \times \Psi$, the density over this joint space is $f(\underline{x}|\underline{\eta}) \cdot f(\underline{\eta}) = f(\underline{x}, \underline{\eta})$, where $f(\underline{\eta})$ describes our uncertainty about the param-

eter. Now, the uncertainty about the parameters can be marginalized out according to

$$f(\underline{x}) = \int_\Psi f(\underline{x}|\underline{\eta}) \cdot f(\underline{\eta})\, \mathrm{d}\underline{\eta} \ . \tag{12}$$

This leads to a standard density function $f(\underline{x})$ over state space without parameter uncertainty. This density can be regarded as a mixture of density functions $f(\underline{x}|\underline{\eta})$ with weights $f(\underline{\eta})$ for different parameters $\underline{\eta}$, whereas the number of components depends on the number of elements in Ψ. Typically, Ψ is a continuous vector space with uncountable elements. Thus, the density (12) is an infinite mixture [12], as it is named in the context of Dirichlet processes.

Our aim within this approach is not to reduce the parameter uncertainty. The parameter uncertainty expresses our lack of knowledge about the system and marginalizing this uncertainty out would result in a probably wrong choice of parameters that shall be avoided here. This approach can also be regarded as a sensitivity analysis. Here multiple different parameters are propagated through the model equations simultaneously and deviations of the output are observed.

4 Examples

This section illustrates the basic use of the processing of densities over parameters of densities according to Bayesian estimation in two simple examples. To keep the equations as simple as possible, only linear problems are considered. All examples in this paper were performed using the Sliced Gaussian Mixture Filter [4]. Here, linear substructures can be employed in order to reduce the estimation effort considerably.

4.1 Example 1

This first example shows the basic procedure of processing probability density functions with uncertain parameters in a Bayesian estimator. In this example, Gaussian densities $f(x) = \mathcal{N}(x - \hat{x}, \sigma^2)$ with mean $\hat{x}$ and
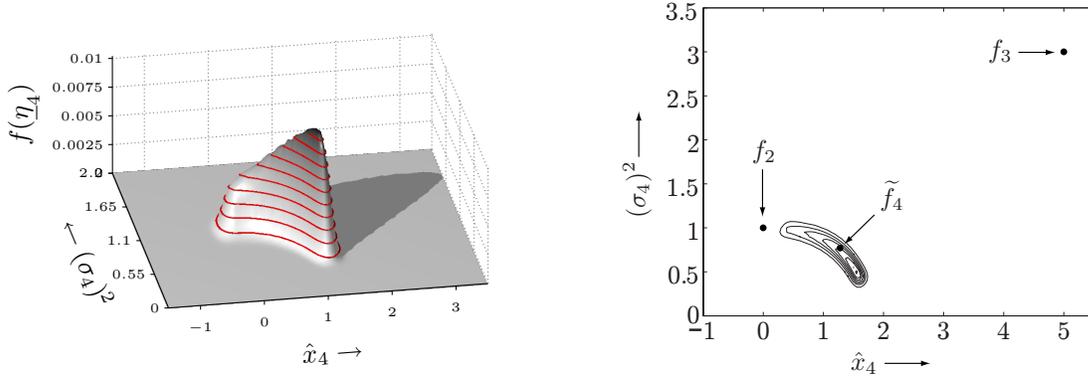
Figure 5: Density $f(\underline{\eta}_4)$ over parameter space for sensor node 4 of Example 2. The correlation coefficient between the two given densities from nodes 2 and 3 is assumed to be uniformly distributed over an interval. On the left hand side, the density is shown. On the right hand side, a contour plot of $f(\underline{\eta}_4)$, the densities $f_2(x_2)$ and $f_3(x_3)$ from sensor nodes 2 and 3, and the true fusion result $\widetilde{f}_4$ are depicted.

variance σ^2 and the parameter vectors

$$\underline{\eta} = \begin{bmatrix} \hat{x} \\ \sigma^2 \end{bmatrix} \in \Psi^D$$

are considered. The measurement model and system model

$$\hat{y}_k = \boldsymbol{x}_k + \boldsymbol{v}_k \quad , \quad \boldsymbol{x}_{k+1} = A \cdot \boldsymbol{x}_k + \boldsymbol{w}_k$$

are both *linear* and depend on *scalar values* only. The noise terms $\boldsymbol{v}_k$ and $\boldsymbol{w}_k$ are zero mean Gaussian distributed random variables. Hence, the parameters of the densities are processed according to the Kalman filter equations.

In the filter step, the prior Gaussian density function $f^p(x_k) = \mathcal{N}\left(x_k - \hat{x}_k^p, (\sigma_k^p)^2\right)$ is defined by the parameter vector $\underline{\eta}_k^p = \left[\hat{x}_k^p, (\sigma_k^p)^2\right]^T$. The measurement $\hat{y}_k$ and the measurement variance $(\sigma_k^v)^2$ define the Gaussian likelihood with the corresponding parameter vector $\underline{\eta}_k^L$. The resulting estimated Gaussian density function is parameterized by $\underline{\eta}_k^e$ with

$$\hat{x}_k^e = \frac{(\sigma_k^v)^2 \cdot (\sigma_k^p)^2}{(\sigma_k^v)^2 + (\sigma_k^p)^2} \left(\frac{\hat{x}_k^p}{(\sigma_k^p)^2} + \frac{\hat{y}_k}{(\sigma_k^v)^2} \right) \quad , \quad (13)$$

$$(\sigma_k^e)^2 = \frac{(\sigma_k^v)^2 \cdot (\sigma_k^p)^2}{(\sigma_k^v)^2 + (\sigma_k^p)^2} \quad , \quad (14)$$

$$w = \mathcal{N}\left(\hat{y}_k - \hat{x}_k^p, (\sigma_k^p)^2 + (\sigma_k^v)^2\right) \quad . \quad (15)$$

Now, consider the mapping $\underline{\mathcal{H}}_k$ of the parameter vectors

$$\underline{\eta}_k^e = \underline{\mathcal{H}}_k(\underline{\eta}_k^p, \underline{\eta}_k^L) + \underline{\mathcal{V}}_k$$

using the equations stated above. $\underline{\mathcal{V}}_k$ is a zero mean Gaussian noise term with small variance. With this mapping, a nonlinear stochastic prediction step in parameter space (9) can be constructed resulting in a posterior density $f(\underline{\eta}_k^e)$ over the parameter space of $\underline{\eta}_k^e$.

Applying this prediction step in parameter space corresponds to a filter step of all prior probability density functions over state space.

The prediction of the linear state space system model can be accomplished analogously. Originating from the linear system model, the Kalman filter equations for the predicted mean and variance are

$$\hat{x}_{k+1}^p = A \cdot \hat{x}_k^e \quad ,$$
$$(\sigma_{k+1}^p)^2 = A^2 \cdot (\sigma_k^e)^2 + (\sigma_k^w)^2 \quad .$$

This leads to the mapping

$$\underline{\eta}_{k+1}^p = \underline{\mathcal{A}}_k(\underline{\eta}_k^e, \underline{\eta}_k^w) + \underline{\mathcal{W}}_k$$

that defines the forward equation of the random vectors $\underline{\eta}_k^e$ and $\underline{\eta}_k^w$ to the predicted random vector $\underline{\eta}_{k+1}^p$. $\underline{\mathcal{W}}_k$ is also a zero mean Gaussian distributed random variable.

In this example, a parameter of the Gaussian distributed random variable $\boldsymbol{v}_k$ is not known exactly. We know that the variance $(\sigma_k^v)^2$ of the Gaussian measurement noise $\boldsymbol{v}_k$ is time-variant with an unknown behaviour over time. We assume that it is uniformly distributed over the complete experiment over the interval $[\frac{1}{2}, 5]$. The variance of the system noise $\boldsymbol{w}_k$ is $\frac{1}{5}$. The type-2 system noise of the measurement step $\underline{\mathcal{V}}_k$ is zero mean Gaussian distributed with variance 0.005. The variance of $\underline{\mathcal{W}}_k$ for the prediction step is 0.005, too. The system is stationary, i.e., $A = 1$ with the real system state 1. The prior density function is assumed to be $f^p(x_0) = \mathcal{N}(x_0 - (-1), 1)$, this means the parameter density is $f(\underline{\eta}_0) = \delta\left([-1, 1]^T\right)$.

Simulation results after 100 consecutive filter and prediction steps are shown in Figure 3. The density $f(\underline{\eta}_{100}^p)$ over the parameter space is shown on the left hand side. The best possible estimation result with exactly known parameter is marked with a bar, which lies near the center of the maximum of the density. On the right hand side, the parameter density is displayed in

state space. Every point $\underline{\eta}$ in parameter space corresponds to a density in state space. This state space density is now drawn for every parameterization $\underline{\eta}$ in relation to its probability $f(\underline{\eta})$ into the same figure. The points, where the density functions are drawn, were weighted according to the probability $f(\underline{\eta})$. A contour plot is used to illustrate the overlappings, and thus, gives an illustrative representation of the type-2 density. For comparison purposes, estimation results for selected parameter values, like the interval borders 0.5 and 5, the median 2.75 and the true, time-variant parameter, are shown. It can be seen, that the type-2 density fits the different parameters very well and gives a good estimate for the true estimation result without any knowledge about the temporal behaviour of this parameter. The case $(\sigma_k^v)^2 = 0.5$ still corresponds to the type-2 density, although with low probability.

The noise terms $\underline{\mathcal{V}}_k$ and $\underline{\mathcal{W}}_k$ in the model are used to obtain more smooth estimation results in parameter space. This supports the use of the Sliced Gaussian Mixture Filter, which exploits overlapping smooth functions. For the assumptions given in the example, the noise terms can be omitted without affecting the conclusion of the results.

4.2 Example 2

The second example relates to information processing in sensor networks. Here, estimation results from individual sensor nodes are transmitted to neighbouring nodes, which fuse this information with own measurements and re-transmit the result to other nodes. Because no history about the fused data is kept, it is not known, which sensor nodes already incorporated their information into the fusion result. Hence, it is possible that some measurements are used several times and thus, correlations between the fusion results occur. These correlations are unknown, but have to be considered when processing these densities.

In this example, estimates from two different information sources have to be processed in order to get an overall fusion result. This is shown in Figure 6. Sensor node 1 transmits its information f_1 to both nodes 2 and 3. They fuse their measurements $\hat{y}_2$ and $\hat{y}_3$ with the received information and transmit the densities f_2 and f_3 to node 4. Node 4 fuses both estimates to the fusion result f_4.

Due to the fact that both estimates f_2 and f_3 contain the same information from node 1, they are dependent. Node 4 doesn't know anything about the specific paths of the densities through the network. Thus, there exists an *unknown correlation* between the two estimates given as scalar random variables $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$, which are distributed according to $\boldsymbol{x}_2 \sim f_2(x)$ and $\boldsymbol{x}_3 \sim f_3(x)$. In this example, the densities are Gaussian $f_2(x) = \mathcal{N}(x - \hat{x}_2, (\sigma_2)^2)$ and $f_3(x) = \mathcal{N}(x - \hat{x}_3, (\sigma_3)^2)$.
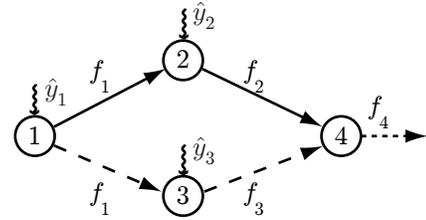

Figure 6: Visualization of the sensor network in Example 2. Node 1 transmits its information given as probability density function f_1 on two paths to nodes 2 and 3. They fuse their local estimate with the received one and re-transmit the result to node 4. Sensor node 4 has to fuse both estimates and outputs the final density f_4.

The resulting density f_4 of sensor node 4 is Gaussian again, with the parameters $\hat{x}_4$ and $(\sigma_4)^2$. The mappings of the parameters for the estimated means and variances are given by

$$\hat{x}_4 = \hat{x}_2 + \frac{\left((\sigma_2)^2 + \sqrt{(\sigma_2)^2 \cdot (\sigma_3)^2} \cdot \rho_4\right) \cdot (\hat{x}_3 - \hat{x}_2)}{(\sigma_2)^2 + (\sigma_3)^2 + 2\sqrt{(\sigma_2)^2 \cdot (\sigma_3)^2} \cdot \rho_4} \, ,$$

$$(\sigma_4)^2 = \frac{(\sigma_2)^2 (\sigma_3)^2 \left(1 - \rho_4^2\right)}{(\sigma_2)^2 + (\sigma_3)^2 + 2\sqrt{(\sigma_2)^2 \cdot (\sigma_3)^2} \cdot \rho_4}$$

and

$$w = \mathcal{N}\left(\hat{x}_3 - \hat{x}_2, (\sigma_2)^2 + (\sigma_3)^2 + 2\sqrt{(\sigma_2)^2 \cdot (\sigma_3)^2} \cdot \rho_4\right)$$

for the fusion step. These equations are processed according to the generative parameter model

$$\underline{\boldsymbol{\eta}}_4 = \underline{\mathcal{H}}_k(\underline{\eta}_2, \underline{\eta}_3, \boldsymbol{\rho}_4) + \underline{\mathcal{V}}_k$$

and equation (9). In contrast to Example 1, the parameter $\rho_4 \in \Psi^E$ belongs to the measurement model. Now, the prior densities are known, whereas the model parameter is unknown to a certain degree.

Sensor node 4 has no information about correlation, so it has to assume correlation to a certain degree. The correlation coefficient ρ_4 is assumed to be uniformly distributed over the interval $[-0.5, 0.5]$.

The measurements and measurement uncertainties of sensors 1,2 and 3 are 1, -0.2, and 9 with variances 6, 1.2, and 6 respectively. Node 4 receives densities from the other nodes with the parameters $\underline{\eta}_2 = [0, 1]^T$ and $\underline{\eta}_3 = [5, 3]^T$. The true fusion result is the Gaussian density $\widetilde{f}_4$ with $\widetilde{\underline{\eta}}_4 \approx [1.29, 0.86]^T$, which is calculated using the information of the three nodes only once. The parameter uncertainty $\underline{\mathcal{V}}_k$ is again zero mean Gaussian distributed with a variance of 0.005.

Figure 5 shows the simulation result for this example. On the left hand side, the parameter density $f(\underline{\eta}_4)$ of the fusion result is shown. On the right hand side, the parameter density, the densities from sensor node 2 and

3, f_2 and f_3, and the true fusion result $\widetilde{f}_4$ is displayed. Here, it can be seen clearly that the resulting type-2 density function is in accordance with the true fusion result.

5 Conclusions and Future Work

In this paper, a method for processing imprecisely known probability density functions and model parameters by means of a Bayesian estimator is stated. The imprecision about the density functions and model parameters is given as probability density function over their parameters. The unique characteristic of this approach is that for state estimation, the Bayesian filter step and prediction step are converted to stochastic forward models, which map the parameters of the state space models and densities onto parameters of the resulting density representation. This allows Bayesian estimation without exact knowledge about certain parameters, which are in several cases hard to find.

The advantage of processing imprecise densities also has a drawback. This approach usually leads to nonlinear prediction problems, which are usually more difficult to solve than the problems in state space. The parameter models are usually nonlinear, even when originating from linear models in state space. Furthermore, the dimensionality depends on the number of parameters needed, which is usually higher than the dimension of the state space.

In comparison to usual hierarchical Bayian modeling, this approach applies a decomposion of the overall problem into a state space part and parameter space part, which allows different esimation techniques and algorithms for both problems, whereas in hierarchcal Bayesian processing, the overall problem has to be solved at once.

Examples show that this approach allows a wide range of applications, like the processing of unknown statistical values, as unknown measurement noise or unknown correlation between density functions, which shall be fused. In this first work, only simple linear systems were considered. When coping with nonlinear systems in state space, the parameter space dimensionality usually increases with every processing step. Although promising preliminary results for nonlinear state space models exist, the reduction of the number of parameters is still an open problem. Instead of only considering type-2 densities with uncertain parameters, the concept can be extended to type-n densities, in which higher order parameters are uncertain, too.

References

[1] S. J. Julier and J. K. Uhlmann, "Unscented Filtering and Nonlinear Estimation," in *Proceedings of the IEEE*, 2004, vol. 92, pp. 401–422.

[2] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation," *IEE Proceedings-F*, vol. 140, no. 2, pp. 107–113, 1993.

[3] M. F. Huber and U. D. Hanebeck, "The Hybrid Density Filter for Nonlinear Estimation based on Hybrid Conditional Density Approximation," in *Proceedings of the 10th International Conference on Information Fusion (FUSION)*, 2007.

[4] Vesa Klumpp, Felix Sawo, Uwe D. Hanebeck, and Dietrich Fränken, "The Sliced Gaussian Mixture Filter for Efficient Nonlinear Estimation," in *Proceedings of the 11th International Conference on Information Fusion (Fusion 2008)*, Cologne, Germany, July 2008.

[5] P. S. Maybeck, *Stochastic Models, Estimation, and Control, Volume 1*, Academic Press, 1979.

[6] Fabio Gagliardi Cozman, "A Brief Introduction to the Theory of Sets of Probability Measures," Tech. Rep., Robotics Institute, Carnegie Mellon University, 1999.

[7] Benjamin Noack, Vesa Klumpp, Dietrich Brunn, and Uwe 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, July 2008.

[8] Peter Walley, *Statistical Reasoning with Imprecise Probabilities*, vol. Monographs on Statistics and Applied Probability 42, London: Chapman and Hall, 1991.

[9] Sanjib Basu, "Ranges of Posterior Probability over a Distribution Band," *Journal of Statistical Planning and Inference*, vol. 44, pp. 149–166, 1995.

[10] J. O. Berger, D. Ríos Insua, and F. Ruggeri, *Robust Bayesian Analysis (Lecture Notes in Statistics)*, vol. 152, chapter Bayesian Robustness, pp. 1–32, Springer, 2000.

[11] M. Berliner, "Hierarchical Bayesian Time Series Models," in *Maximum Entropy and Bayesian Methods*, K. Hanson and R. Silver, Eds. Kluwer, 1996, pp. 15–22.

[12] Francois Caron, Manuel Davy, Arnaud Doucet, Emmanuel Duflos, and Philippe Vanheeghe, "Bayesian Inference for Dynamic Models with Dirichlet Process Mixtures," in *Proceedings of the 9th International Conference on Information Fusion (Fusion 2006)*, 2006.

[13] J. M. Mendel, "Type-2 Fuzzy Sets and Systems: An Overview," *IEEE Computational Intelligence Magazine*, vol. 2, no. 1, pp. 20–29, 2007.