Nonlinear Bayesian Estimation with Convex Sets of Probability Densities

Benjamin Noack, Vesa Klumpp, Dietrich Brunn, and Uwe D. Hanebeck
Intelligent Sensor-Actuator-Systems Laboratory
Institute of Computer Science and Engineering
Universität Karlsruhe (TH), Germany
Email: {bnoack,klumpp,brunn}@ira.uka.de, uwe.hanebeck@ieee.org

***Abstract*—This paper presents a theoretical framework for Bayesian estimation in the case of imprecisely known probability density functions. The lack of knowledge about the true density functions is represented by sets of densities. A formal Bayesian estimator for these sets is introduced, which is intractable for infinite sets. To obtain a tractable filter, properties of convex sets in form of convex polytopes of densities are investigated. It is shown that pathwise connected sets and their convex hulls describe the same ignorance. Thus, an exact algorithm is derived, which only needs to process the hull, delivering tractable results in the case of a proper parametrization. Since the estimator delivers a convex hull of densities as output, the theoretical grounds are laid for deriving efficient Bayesian estimators for sets of densities. The derived filter is illustrated by means of an example.**

Keywords: Nonlinear Bayesian estimation, convex set, convex polytope

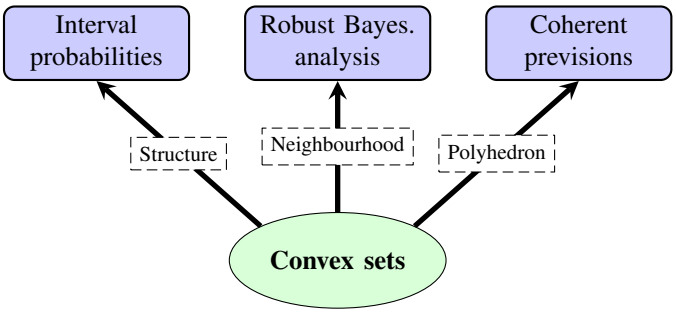


Figure 1. Convex sets of probabilities in the theorey of interval probabilities, robust Bayesian analysis and coherent previsions.

I. Introduction

In many technical applications, unknown quantities are incorporated and have to be dealt with. Typical examples include localization problems, map building, reconstruction of quantities from noisy measurements, etc. A main problem is to determine estimates of inaccessible states from measurements corrupted by noise. In general, Bayesian inference models are used for combining data to reach conclusions and insight. This implies that the herein used probability distributions are assumed to be optimal. The appropriateness of stochastic mechanisms has thus to be questioned, when probabilities are not well known. A proposal to overcome this issue is to process sets of possible probabilities, which are the focus of this paper.

After proposing a generic Bayesian estimator for sets of densities, convex sets parametrized as convex polytopes are investigated. It is shown that a pathwise connected set and its convex hull describe the same ignorance. Thus restricting our investigation to convex sets is sufficient. An exact Bayesian estimator, which only processes the convex hull, is derived and illustrated by means of an example.

First of all the existing approaches are discussed: For stochastic uncertainties a multitude of approaches exist. For linear Gaussian systems, the famous Kalman Filter [1] can be utilized. For the case of non-Gaussian noise, typically appearing in nonlinear systems, closed-form solutions do not exist in general. Additionally, recursive processing being necessary for many technical systems, the complexity of the density representation increases continually. Only in special cases, the complexity is bounded [2]. So, for almost all real-world applications, approximations are inevitable. A common approach is to linearize the system as in the extended Kalman Filter [3] or to assume the joint densities to be Gaussian as in the Unscented Kalman Filter [4].

Widely used are particle filters, where densities are represented by random or pseudo-random samples [5]. Though they shine through their algorithmic simplicity, they do posses some problems in terms of convergence [6], [7]. Another approach is to represent the probability densities by function series: [8] uses Gram-Chalier expansions, [9] derives a filter based on generalized Edgeworth series for continuous time systems with additive Gaussian noise, and [10], [11] use Fourier Expansions. Furthermore, Gaussian Mixtures [12], Dirac Mixtures [13] or combinations of them [14] are also utilized.

Convex sets of probabilities found attention in literature as well. [15] investigates interval-probabilities. This concept pursues the goal of generalizing classical probability theory in order to achieve an extensible description of uncertainty. The key idea is to assign intervals to random events instead of single values. Convex sets have a significant role in this theory: The largest set of distributions yielding a certain interval of probabilities is convex. Such a set is called the structure of an interval probability assessment.

In the theory of robust Bayesian analysis [16], [17] convex sets are used to assess global robustness. Neighbourhood

classes of functions are often employed to study sensitivity to prior distributions. At this, the range of deviations is computed while varying the prior over a class of functions. Such sets are common to be convex.

Lower and upper probabilities are defined as a special case of lower and upper previsions in the subjectivistic and behavioristic theory of imprecise probabilities [18]. These probabilities generally correspond to convex polyhedrons with parallel faces.

As illustrated by Figure 1, convex sets are to be regarded as an underlying concept for a variety of theories, which aspire to extend classical probability theory. Against the background of decision theory, [19] provides an overview and a justification of using convex sets of probabilities. In this context, convex sets are called credal sets. Considering linear loss or utility functions, an arbitrary set and its convex hull effect the same pattern of preferences. So the main reason for using credal sets is that the largest set containing all probability distributions, which induce a specific decision, is convex. Transferring the concepts of conditional probability and independence to the theory of probability sets turned out to be a difficult task. We overcome this issue by identifying all sets inducing the same probability assessment.

For our investigations we will follow Levi's epistemology as suggested in [20], [21]. According to this, stochastic models are intended to express lack of precision. Such a model produces a single estimate, which is considered to be the best one. Due to inappropriate assumptions, incomplete prior knowledge or unknown quantities, there might be more estimates to be taken into account. In this regard, set-valued stochastic models can cope with the incapability to specify appropriate probabilities. In that case, a random variable will be characterized by a set of distributions, each of which is a possible and permissible probabilistic description of that variable. Sets of probability distributions represent a state of ignorance, since we are not capable of making a decision among those distributions. This means that the possible distributions compete for being the best estimate. Adding all convex combinations to the set, i.e., taking the convex hull, is one way to effect a compromise. This accounts for the strong tendency towards convex sets. Existing approaches such as the set-valued Kalman filter [21] or projection-based approaches [22] model only the initial state as a convex set. In contrast, this work provides a theoretical framework for modeling system and measurement noise as convex sets as well.

In this article, we consider systems with a continuous state space and investigate convex sets of probability density functions. The sets of the corresponding cumulative distributions are also convex because of the linearity of integration. The following section reviews the discrete-time Bayesian estimator and explains the concept of processing sets of densities. The remainder of this paper is structured as follows: In Section III, convex polytopes of densities are defined. It is shown that intervals of probabilities and expectation values can easily be computed by using the vertices of the polytopes. Section IV derives a Bayesian estimator for convex polytopes. This

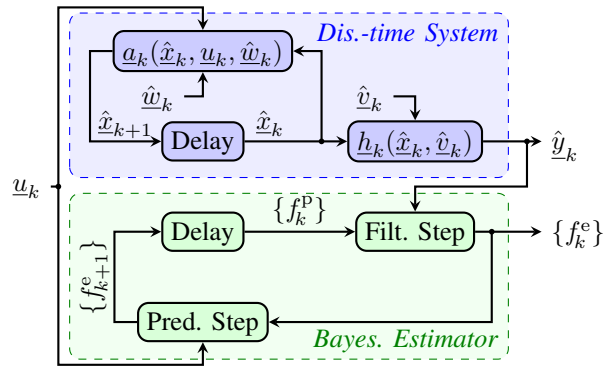


Figure 2. Structure of a discrete-time system and a Bayesian estimator for sets of densities.

estimator processes only the vertices of the polytopes and thereby generates the vertices of a new convex polytope. This resulting polytope is the convex hull of the exact set, which would arise from element-wise processing. The key result is that the exact set and its covering convex polytope yield the same probability intervals, even after multiple processing steps. An example application is investigated in Section V. Here, the measurement uncertainties of a radar altimeter are given by a convex polytope. This work is concluded in Section VI with an outlook to future investigations.

II. Problem Formulation

In this paper, we focus on probabilistic dynamic discrete-time nonlinear systems, which can be described by

$$\underline{\boldsymbol{x}}_{k+1} = \underline{a}_k(\underline{\boldsymbol{x}}_k, \underline{u}_k, \underline{\boldsymbol{w}}_k) \ , \tag{1}$$

$$\underline{\boldsymbol{y}}_k = \underline{h}_k(\underline{\boldsymbol{x}}_k, \underline{\boldsymbol{v}}_k) \ , \tag{2}$$

where the underlining denotes vectors and the lowercase boldface letters random variables. In the system equation (1), the time update is performed: The system function $\underline{a}_k(\,\cdot\,,\,\cdot\,,\,\cdot\,)$ projects the n-dimensional inaccessible state, characterized by the random vector $\underline{\boldsymbol{x}}_k \in \Omega_{\underline{x}} \subseteq \mathbb{R}^n$, with respect to the input $\underline{u}_k \in \Omega_{\underline{u}} \subseteq \mathbb{R}^k$ and the system noise $\underline{\boldsymbol{w}}_k \in \Omega_{\underline{w}} \subseteq \mathbb{R}^p \sim f_k^w(\underline{w}_k)$ onto the state $\underline{\boldsymbol{x}}_{k+1}$ at the discrete time $k+1$. The measurement equation (2) models the outcome of a measurement $\underline{\boldsymbol{y}}_k \in \Omega_{\underline{y}} \subseteq \mathbb{R}^m$ at time k with respect to the state $\underline{\boldsymbol{x}}_k$ and the measurement noise $\underline{\boldsymbol{v}}_k \in \Omega_{\underline{v}} \subseteq \mathbb{R}^q \sim f_k^v(\underline{v}_k)$ utilizing the measurement function $\underline{h}_k(\,\cdot\,,\,\cdot\,)$.

The real system is a realization of the probabilistic system (1) and (2) as depicted in Figure 2. The realizations of random vectors are denoted by a hat.

Assuming the probability densities of the noise $f_k^w(\underline{w}_k)$, $f_k^v(\underline{v}_k)$ and the initial density $f_0^{\mathrm{p}}(\underline{x}_k)$ as being stochastically independent, an estimation for the state $\underline{\boldsymbol{x}}_k$ at time k can be determined with the Bayesian estimator, which consists of two steps:

The *filtering step* determines an improved estimate $f_{\underline{x}}^{\mathrm{e}}(\underline{x}_k)$ by incorporating a measurement $\hat{\underline{y}}_k$ into a prior estimate

$f_k^{\mathrm{p}}(\underline{x}_k)$ by evaluating

$$f_k^{\mathrm{e}}(\underline{x}_k) = \frac{1}{c_k} f_k^{\mathrm{L}}(\underline{x}_k) f_k^{\mathrm{p}}(\underline{x}_k) \ , \tag{3}$$

$$f_k^{\mathrm{L}}(\underline{x}_k) := f(\hat{\underline{y}}_k|\underline{x}_k) = \int_{\Omega_{\underline{v}}} \delta^q\big(\hat{\underline{y}}_k - \underline{h}_k(\underline{x}_k, \underline{v}_k)\big) f_k^v(\underline{v}_k) \ \mathrm{d}\underline{v}_k \ ,$$

$$c_k = \int_{\Omega_{\underline{x}}} f_k^{\mathrm{L}}(\underline{x}_k) f_k^{\mathrm{p}}(\underline{x}_k) \ \mathrm{d}\underline{x}_k \ .$$

The conditional density $f_k^{\mathrm{L}}(\underline{x}_k) = f(\hat{\underline{y}}_k|\underline{x}_k)$ is called likelihood, c_k is a normalization constant, and $\delta^q(.)$ is the q-dimensional Dirac delta function.

In the *prediction step*, the state estimate is extrapolated from $f_k(\underline{x}_k)$ at discrete time k to time $k+1$ by determining

$$f_k(\underline{x}_{k+1}) = \int_{\Omega_{\underline{x}}} f_k^{\mathrm{T}}(\underline{x}_{k+1}|\underline{x}_k) f_k(\underline{x}_k) \ \mathrm{d}\underline{x}_k \ , \tag{4}$$

$$f_k^{\mathrm{T}}(\underline{x}_{k+1}|\underline{x}_k) = \int_{\Omega_{\underline{w}}} \delta^p \left(\underline{x}_{k+1} - \underline{a}_k(\underline{x}_k, \underline{u}_k, \underline{w}_k)\right) f_k^w(\underline{w}_k) \ \mathrm{d}\underline{w}_k \ .$$

The conditional density $f_k^{\mathrm{T}}(\underline{x}_{k+1}|\underline{x}_k)$ is also called transition density. Since $\underline{u}_k$ is assumed to be known, it is not explicitly denoted in $f_k^{\mathrm{T}}(\underline{x}_{k+1}|\underline{x}_k)$. The prior density normally is the result of a previous filtering step, i. e., $f_k(\underline{x}_k) = f_k^{\mathrm{e}}(\underline{x}_k)$.

Now we turn to investigating sets of densities: As discussed in the introduction, such sets can evolve from imprecise prior knowledge, unknown measurements or unknown system parameters. Since this work focuses on theoretical aspects, it is not investigated how proper sets of densities can be obtained for practical applications.

Here we consider to have a set of system functions $\mathcal{A}_K$ and a set of measurement functions $\mathcal{H}_k$ at time k. The set of transition densities is then given by

$$\boldsymbol{\mathcal{M}}_k^T := \Big\{ f_k^{\mathrm{T}} \big| \underline{x}_k, \underline{x}_{k+1} \in \Omega_{\underline{x}}, \underline{a}_k \in \mathcal{A}_k, \\ f_k^{\mathrm{T}}(\underline{x}_{k+1}|\underline{x}_k) = \int_{\Omega_{\underline{w}}} \delta^p \left(\underline{x}_{k+1} - \underline{a}_k(\underline{x}_k, \underline{u}_k, \underline{w}_k)\right) f_k^w(\underline{w}_k) \ \mathrm{d}\underline{w}_k \Big\}$$

and the set of likelihoods is given by

$$\boldsymbol{\mathcal{M}}_k^L = \Big\{ f_k^{\mathrm{L}} \big| f_k^{\mathrm{L}}(\underline{x}_k) = \int_{\Omega_{\underline{v}}} \delta^q\big(\hat{\underline{y}}_k - \underline{h}_k(\underline{x}_k, \underline{v}_k)\big) f_k^v(\underline{v}_k) \ \mathrm{d}\underline{v}_k, \\ \underline{x}_k \in \Omega_{\underline{x}}, \underline{h}_k \in \mathcal{H}_k \Big\} \ .$$

The Bayesian estimator for sets of densities is defined as the element-wise processing of the sets. The filtering step can be written as

$$\boldsymbol{\mathcal{M}}_k^e = \Big\{ f_k^{\mathrm{e}} \big| f_k^{\mathrm{e}}(\underline{x}_k) = \frac{f_k^{\mathrm{p}}(\underline{x}_k) f_k^{\mathrm{L}}(\underline{x}_k)}{\int_{\Omega_{\underline{x}}} f_k^{\mathrm{p}}(\underline{x}_k) f_k^{\mathrm{L}}(\underline{x}_k) \ \mathrm{d}\underline{x}_k}, \\ \underline{x}_k \in \Omega_{\underline{x}}, f_k^{\mathrm{L}} \in \boldsymbol{\mathcal{M}}_k^L, f_k^{\mathrm{p}} \in \boldsymbol{\mathcal{M}}_k^p \Big\} \ ,$$

with $\boldsymbol{\mathcal{M}}_k^p$ being the set of priors. The prediction step results in

$$\boldsymbol{\mathcal{M}}_{k+1}^p = \Big\{ f_k^{\mathrm{p}} \big| f_k^{\mathrm{p}}(\underline{x}_{k+1}) = \int_{\Omega_{\underline{x}}} f_k^{\mathrm{T}}(\underline{x}_{k+1}|\underline{x}_k) f_k(\underline{x}_k) \ \mathrm{d}\underline{x}_k, \\ \underline{x}_{k+1} \in \Omega_{\underline{x}}, f_k^{\mathrm{T}} \in \boldsymbol{\mathcal{M}}_k^T, f_k^{\mathrm{e}} \in \boldsymbol{\mathcal{M}}_k^e \Big\} \ .$$

This formulation is more of formal importance, because for infinite sets the latter expressions are intractable. Therefore, a parametrization based on convex polytopes is investigated in the following sections.

III. Convex Sets of Probability densities

As a special case of convex sets we consider convex polytopes of probability densities, which can easily be obtained by taking the convex hull of a finite number of density functions. The reason for choosing convex polytopes is clarity and mathematical brevity. Since the number of density functions may be arbitrarily large, any convex set can be approximated with an arbitrarily small error. A generalization to general convex sets is a laborious task but does not promise additional insights.

A convex polytope is defined by

$$\boldsymbol{\mathcal{P}} := \left\{ \sum_{i=1}^n \alpha_i \cdot f_i \ \Big| \ \sum_{i=1}^n \alpha_i = 1, \alpha_i \geq 0 \right\} , \tag{5}$$

where $f_1, \ldots, f_n$ are arbitrary probability density functions. For every $f = \sum_i \alpha_i f_i \in \boldsymbol{\mathcal{P}}$ the probability of an event A can be written as the convex combination

$$P(\mathrm{A}) = \int_{\mathrm{A}} f(\underline{x}) \ \mathrm{d}\underline{x} = \sum_{i=1}^n \alpha_i \int_{\mathrm{A}} f_i(\underline{x}) \ \mathrm{d}\underline{x} \ ,$$

where the integrals in the latter expression are the probabilities computed from the vertices $f_1, \ldots, f_n$. Hence, $\boldsymbol{\mathcal{P}}$ yields a probability interval $\mathcal{I}_{\boldsymbol{\mathcal{P}}}(\mathrm{A}) := \big[\underline{P}(\mathrm{A}), \overline{P}(\mathrm{A})\big]$ for every event A, where

$$\underline{P}(\mathrm{A}) = \min_{i=1}^n \left\{ \int_{\mathrm{A}} f_i(\underline{x}) \ \mathrm{d}\underline{x} \right\}$$

and

$$\overline{P}(\mathrm{A}) = \max_{i=1}^n \left\{ \int_{\mathrm{A}} f_i(\underline{x}) \ \mathrm{d}\underline{x} \right\} \ .$$

The probability interval of the complementary event A^c is given by $\big[1 - \overline{P}(\mathrm{A}), 1 - \underline{P}(\mathrm{A})\big]$.

Interestingly, a pathwise connected set $\boldsymbol{\mathcal{M}}$ which contains $f_1, \ldots, f_n$ and which has the convex hull $\boldsymbol{\mathcal{P}}$ as depicted in Figure 3 provides the same intervals of probabilities as $\boldsymbol{\mathcal{P}}$. This relationship is proven in the following theorem. Note, that the operator $\mathrm{conv}\{\boldsymbol{\mathcal{M}}\}$ denotes the convex hull of $\boldsymbol{\mathcal{M}}$.

Theorem 1

Consider a polytope $\boldsymbol{\mathcal{P}}$ as defined in equation (5) *and a pathwise connected set $\boldsymbol{\mathcal{M}}$ with the following properties:*

$$\{f_1, \ldots, f_n\} \in \boldsymbol{\mathcal{M}} \qquad \text{and} \qquad \boldsymbol{\mathcal{P}} = \mathrm{conv}\{\boldsymbol{\mathcal{M}}\} \ .$$

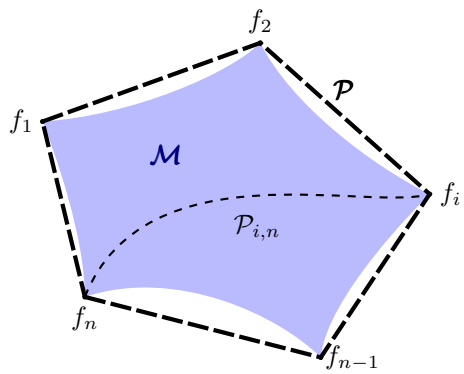


Figure 3. The set $\boldsymbol{\mathcal{M}}$ contains the vertices $f_1, \ldots, f_n$ of $\boldsymbol{\mathcal{P}}$ and is pathwise connected.

Then, for an event A*, the set of probabilities*

$$\mathcal{I}_{\boldsymbol{\mathcal{M}}}(\mathrm{A}) := \left\{ P(\mathrm{A}) \middle| P(\mathrm{A}) = \int_{\mathrm{A}} f(\underline{x})\ \mathrm{d}\underline{x},\ f \in \boldsymbol{\mathcal{M}} \right\}$$

is an interval identical to $\mathcal{I}_{\boldsymbol{\mathcal{P}}}(\mathrm{A})$.

PROOF. Since $\boldsymbol{\mathcal{M}} \subseteq \boldsymbol{\mathcal{P}}$, the set of probabilities $\mathcal{I}_{\boldsymbol{\mathcal{M}}}(\mathrm{A})$ is a subset of $\mathcal{I}_{\boldsymbol{\mathcal{P}}}(\mathrm{A})$.

Let f_i and f_j define the endpoints of the probability interval $\mathcal{I}_{\boldsymbol{\mathcal{P}}}(\mathrm{A})$. Since $\boldsymbol{\mathcal{M}}$ is pathwise connected, there exists a continuous path $\mathcal{P}_{i,j} : [0,1] \mapsto \boldsymbol{\mathcal{M}}$ with $\mathcal{P}_{i,j}(0) = f_i$ and $\mathcal{P}_{i,j}(1) = f_j$. Due to the linearity of the integral, the interval

$$\mathcal{I}_{i,j}(\mathrm{A}) = \left\{ P(\mathrm{A}) \middle| P(\mathrm{A}) = \int_{\mathrm{A}} f(\underline{x})\ \mathrm{d}\underline{x},\ f \in \mathcal{P}_{i,j}\left([0,1]\right) \right\}$$

is pathwise connected as well and has the same endpoints as $\mathcal{I}_{\boldsymbol{\mathcal{P}}}(\mathrm{A})$. Thus we have

$$\mathcal{I}_{\boldsymbol{\mathcal{P}}}(\mathrm{A}) = \mathcal{I}_{i,j}(\mathrm{A}) \subseteq \mathcal{I}_{\boldsymbol{\mathcal{M}}}(\mathrm{A})$$

and the proof is complete. □

This Theorem is a strong argument for the utilization of convex sets: It does not make any difference if a pathwise connected set or its convex hull is considered. The probability intervals of any event will not differ.

Analogously, we can derive an interval for the expectation values induced by $\boldsymbol{\mathcal{P}}$, i.e.,

$$\left[\min_{i=1}^{n} \mathrm{E}_{f_i}\{\boldsymbol{x}\}\,,\ \max_{i=1}^{n} \mathrm{E}_{f_i}\{\boldsymbol{x}\} \right] \quad .$$

In a similar manner, the preceding observations can be adopted for every linear functional, since sets of convex combinations of scalars result in intervals.

In contrast, specifying intervals for the variances cannot be achieved by determining minimal and maximal variance of the vertex densities. A lengthy, but straightforward, calculation yields

$$\begin{aligned} \mathrm{V}_f\{\boldsymbol{x}\} = &\sum_{i=1}^{n} \alpha_i\, \mathrm{V}_{f_i}\{\boldsymbol{x}\} \\ &+ \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \left[\mathrm{E}_{f_i}\{\boldsymbol{x}\} - \mathrm{E}_{f_j}\{\boldsymbol{x}\} \right]^2 \end{aligned}$$

for $f = \sum_i \alpha_i f_i \in \boldsymbol{\mathcal{P}}$. So the variances of $\boldsymbol{\mathcal{P}}$ depend both on the variances of the vertices $f_1, \ldots, f_n$ and the squared distances of their expected values.

IV. Bayesian Estimation with Convex Sets

The previously explained Bayesian estimator for sets of densities performs filter and prediction steps elementwise. Consequently, an appropriate representation of infinite sets is inevitable in order to realize filter and prediction step with bounded complexity. We have to ascertain that we are able to retain that representation after filtering and prediction for further processing steps. In this section, we present a Bayesian state estimator for convex polytopes of density functions. We assume that not only the initial state but also measurement and system noise are modelled as convex polytopes.

A. Filtering with Convex Polytopes

First, we consider the filter step. Let

$$\boldsymbol{\mathcal{P}}^{\mathrm{p}} := \mathrm{conv}\{f_1^{\mathrm{p}}, \ldots, f_m^{\mathrm{p}}\}$$

be the convex polytope of prior or predicted probability density functions and

$$\boldsymbol{\mathcal{P}}^{\mathrm{L}} := \mathrm{conv}\left\{f_1^{\mathrm{L}}, \ldots, f_n^{\mathrm{L}}\right\}$$

be the convex polytope of likelihoods. For the sake of clarity, we omit the time index k. Here, the subscripts denote the different vertices of a polytope. The elementwise application of filter step (3) yields the *exact* set

$$\boldsymbol{\mathcal{M}}^{\mathrm{e}} := \left\{ \frac{f^{\mathrm{p}} \cdot f^{\mathrm{L}}}{\int_{\Omega} f^{\mathrm{p}}(\underline{\nu}) \cdot f^{\mathrm{L}}(\underline{\nu})\ \mathrm{d}\underline{\nu}} \,\middle|\, f^{\mathrm{p}} \in \boldsymbol{\mathcal{P}}^{\mathrm{p}},\ f^{\mathrm{L}} \in \boldsymbol{\mathcal{P}}^{\mathrm{L}} \right\} \quad .$$

The following theorem reveals a way of calculating the convex hull of $\boldsymbol{\mathcal{M}}^{\mathrm{e}}$. Let $\boldsymbol{\mathcal{V}}\left(\boldsymbol{\mathcal{P}}^{\mathrm{p}}\right)$ and $\boldsymbol{\mathcal{V}}\left(\boldsymbol{\mathcal{P}}^{\mathrm{L}}\right)$ denote the vertices of $\boldsymbol{\mathcal{P}}^{\mathrm{p}}$ and $\boldsymbol{\mathcal{P}}^{\mathrm{L}}$, respectively. We have $\boldsymbol{\mathcal{V}}\left(\boldsymbol{\mathcal{P}}^{\mathrm{p}}\right) = \{f_1^{\mathrm{p}}, \ldots, f_m^{\mathrm{p}}\}$ and $\boldsymbol{\mathcal{V}}\left(\boldsymbol{\mathcal{P}}^{\mathrm{L}}\right) = \left\{f_1^{\mathrm{L}}, \ldots, f_n^{\mathrm{L}}\right\}$.

Theorem 2 (Filtering Step for Convex Polytopes)
The vertices of the smallest convex polytope covering $\boldsymbol{\mathcal{M}}^{\mathrm{e}}$ *are obtained by elementwise filtering of* $\boldsymbol{\mathcal{V}}\left(\boldsymbol{\mathcal{P}}^{\mathrm{p}}\right)$ *with* $\boldsymbol{\mathcal{V}}\left(\boldsymbol{\mathcal{P}}^{\mathrm{L}}\right)$. *According to this, we define* $\boldsymbol{\mathcal{P}}^{\mathrm{e}}$ *by*

$$\mathrm{conv}\left\{ \frac{f_i^{\mathrm{p}} \cdot f_j^{\mathrm{L}}}{\int_{\Omega} f_i^{\mathrm{p}}(\underline{\nu}) \cdot f_j^{\mathrm{L}}(\underline{\nu})\ \mathrm{d}\underline{\nu}} \,\middle|\, f_i^{\mathrm{p}} \in \boldsymbol{\mathcal{V}}\left(\boldsymbol{\mathcal{P}}^{\mathrm{p}}\right), f_j^{\mathrm{L}} \in \boldsymbol{\mathcal{V}}\left(\boldsymbol{\mathcal{P}}^{\mathrm{L}}\right) \right\} \quad .$$

Then we have

$$\boldsymbol{\mathcal{P}}^{\mathrm{e}} = \mathrm{conv}\{\boldsymbol{\mathcal{M}}^{\mathrm{e}}\} \quad .$$

PROOF. Suppose that f^{p} is an element of $\boldsymbol{\mathcal{P}}^{\mathrm{p}}$ and f^{L} lies in $\boldsymbol{\mathcal{P}}^{\mathrm{L}}$. Then there are weights $\alpha_1, \ldots, \alpha_m$ and $\beta_1, \ldots, \beta_n$, such that

$$f^{\mathrm{p}} = \sum_{i=1}^{m} \alpha_i \cdot f_i^{\mathrm{p}} \quad \text{and} \quad f^{\mathrm{L}} = \sum_{j=1}^{n} \beta_i \cdot f_j^{\mathrm{L}} \quad ,$$

where $f_i^{\mathrm{p}} \in \boldsymbol{\mathcal{V}}(\boldsymbol{\mathcal{P}}^{\mathrm{p}})$ and $f_j^{\mathrm{L}} \in \boldsymbol{\mathcal{V}}(\boldsymbol{\mathcal{P}}^{\mathrm{L}})$. For a shorter notation, we define $\Gamma_{i,j} := \int_\Omega f_i^{\mathrm{p}}(\nu) \cdot f_j^{\mathrm{L}}(\nu)\ \mathrm{d}\nu$. Applying (3) to f^{p} and f^{L} gives

$$\begin{aligned} f^{\mathrm{e}}(\underline{x}_k) &= \frac{f^{\mathrm{p}}(\underline{x}_k) \cdot f^{\mathrm{L}}(\underline{x}_k)}{\int_\Omega f^{\mathrm{p}}(\underline{\nu}) \cdot f^{\mathrm{L}}(\underline{\nu})\ \mathrm{d}\underline{\nu}} \\ &= \frac{\sum_{i=1,j=1}^{m,n} \alpha_i f_i^{\mathrm{p}}(\underline{x}_k) \cdot \beta_j f_j^{\mathrm{L}}(\underline{x}_k)}{\sum_{k=1,l=1}^{m,n} \alpha_k \beta_l \int_\Omega f_k^{\mathrm{p}}(\underline{\nu}) \cdot f_l^{\mathrm{L}}(\underline{\nu})\ \mathrm{d}\underline{\nu}} \\ &= \sum_{i=1,j=1}^{m,n} \left[\frac{\alpha_i \beta_j}{\sum_{k=1,l=1}^{m,n} \alpha_k \beta_l \Gamma_{k,l}} \right] f_i^{\mathrm{p}}(\underline{x}_k) \cdot f_j^{\mathrm{L}}(\underline{x}_k) \\ &= \sum_{i=1,j=1}^{m,n} \left[\frac{\alpha_i \beta_j \Gamma_{i,j}}{\sum_{k=1,l=1}^{m,n} \alpha_k \beta_l \Gamma_{k,l}} \right] \frac{f_i^{\mathrm{p}}(\underline{x}_k) \cdot f_j^{\mathrm{L}}(\underline{x}_k)}{\Gamma_{i,j}} \ . \end{aligned}$$

The coefficients in the latter sum are non-negative and satisfy

$$\sum_{i=1,j=1}^{m,n} \left[\frac{\alpha_i \beta_j \Gamma_{i,j}}{\sum_{k=1,l=1}^{m,n} \alpha_k \beta_l \Gamma_{k,l}} \right] = 1 \ .$$

Hence, f^{e} is to be regarded as a convex combination of the density functions

$$\frac{f_i^{\mathrm{p}} \cdot f_j^{\mathrm{L}}}{\int_\Omega f_i^{\mathrm{p}}(\underline{\nu}) \cdot f_j^{\mathrm{L}}(\underline{\nu})\ \mathrm{d}\underline{\nu}} \ ,$$

which result from elementwise filtering of $\boldsymbol{\mathcal{V}}(\boldsymbol{\mathcal{P}}^{\mathrm{p}})$ and $\boldsymbol{\mathcal{V}}(\boldsymbol{\mathcal{P}}^{\mathrm{L}})$ and are themselves elements of $\boldsymbol{\mathcal{M}}^{\mathrm{e}}$. In particular, f^{e} lies in the convex hull of these functions. This states that $\boldsymbol{\mathcal{M}}^{\mathrm{e}} \subseteq \boldsymbol{\mathcal{P}}^{\mathrm{e}}$, which finishes the proof. □

If $\boldsymbol{\mathcal{P}}^{\mathrm{p}}$ or $\boldsymbol{\mathcal{P}}^{\mathrm{L}}$ consists of only one element, then convexity will be conserved, as stated in the next theorem.

Theorem 3
The exact set $\boldsymbol{\mathcal{M}}^{\mathrm{e}}$ is convex, if $\boldsymbol{\mathcal{P}}^{\mathrm{p}}$ or $\boldsymbol{\mathcal{P}}^{\mathrm{L}}$ is a singleton set. In particular, the following equation applies:

$$\boldsymbol{\mathcal{P}}^{\mathrm{e}} = \boldsymbol{\mathcal{M}}^{\mathrm{e}}$$

PROOF. We will show $\boldsymbol{\mathcal{M}}^{\mathrm{e}} \supseteq \boldsymbol{\mathcal{P}}^{\mathrm{e}}$. Without loss of generality, we assume that $\boldsymbol{\mathcal{P}}^{\mathrm{L}} = \{f_1^{\mathrm{L}}\}$. Let $f^{\mathrm{e}} \in \boldsymbol{\mathcal{P}}^{\mathrm{e}}$ be arbitrary. Then this function is a convex combination

$$f^{\mathrm{e}} = \sum_{i=1}^{m} \alpha_i \cdot \frac{f_i^{\mathrm{p}} \cdot f_1^{\mathrm{L}}}{\int_\Omega f_i^{\mathrm{p}}(\underline{\nu}) \cdot f_1^{\mathrm{L}}(\underline{\nu})\ \mathrm{d}\underline{\nu}}$$

of the vertices $\boldsymbol{\mathcal{V}}(\boldsymbol{\mathcal{P}}^{\mathrm{e}})$, where $\sum_{i=1}^m \alpha_i = 1,\ \alpha_i \geq 0$. In order to prove that f^{e} lies in $\boldsymbol{\mathcal{M}}^{\mathrm{e}}$ we have to determine $f^{\mathrm{p}} = \sum_{i=1}^m \gamma_i \cdot f_i^{\mathrm{p}} \in \boldsymbol{\mathcal{P}}^{\mathrm{p}}$, such that

$$\begin{aligned} f^{\mathrm{e}} &= \sum_{i=1}^{m} \alpha_i \cdot \frac{f_i^{\mathrm{p}}(\underline{x}_k) \cdot f_1^{\mathrm{L}}(\underline{x}_k)}{\int_\Omega f_i^{\mathrm{p}}(\underline{\nu}) \cdot f_1^{\mathrm{L}}(\underline{\nu})\ \mathrm{d}\underline{\nu}} \\ &= \frac{f^{\mathrm{p}}(\underline{x}_k) \cdot f_1^{\mathrm{L}}(\underline{x}_k)}{\int_\Omega f^{\mathrm{p}}(\underline{\nu}) \cdot f_1^{\mathrm{L}}(\underline{\nu})\ \mathrm{d}\underline{\nu}} \\ &= \frac{\left(\sum_{i=1}^m \gamma_i \cdot f_i^{\mathrm{p}}(\underline{x}_k)\right) \cdot f_1^{\mathrm{L}}(\underline{x}_k)}{\int_\Omega \left(\sum_{k=1}^m \gamma_k \cdot f_i^{\mathrm{p}}(\underline{\nu})\right) \cdot f_1^{\mathrm{L}}(\underline{\nu})\ \mathrm{d}\underline{\nu}} \\ &= \sum_{i=1}^{m} \gamma_i \cdot \frac{f_i^{\mathrm{p}}(\underline{x}_k) \cdot f_1^{\mathrm{L}}(\underline{x}_k)}{\sum_{k=1}^m \gamma_k \int_\Omega f_k^{\mathrm{p}}(\underline{\nu}) \cdot f_1^{\mathrm{L}}(\underline{\nu})\ \mathrm{d}\underline{\nu}} \in \boldsymbol{\mathcal{M}}^{\mathrm{e}} \end{aligned} \tag{6}$$

holds. A comparison of coefficients gives

$$\frac{\alpha_i}{\Gamma_i} = \frac{\gamma_i}{\sum_{k=1}^m \gamma_k \cdot \Gamma_k} \ . \tag{7}$$

As before, Γ_i denotes the integral $\int_\Omega f_k^{\mathrm{p}}(\underline{\nu}) \cdot f_1^{\mathrm{L}}(\underline{\nu})\ \mathrm{d}\underline{\nu}$. We require that the weights γ_i sum to one and therefore we have

$$1 = \sum_{i=1}^m \gamma_i = \sum_{i=1}^m \frac{\alpha_i}{\Gamma_i} \cdot \left(\sum_{k=1}^m \gamma_k \cdot \Gamma_k \right) \ .$$

Together with (7) we now obtain

$$\gamma_i = \frac{\alpha_i / \Gamma_i}{\sum_{i=1}^m \alpha_i / \Gamma_i} \ .$$

The weights γ_i are obviously non-negative and result thus in the demanded function $f^{\mathrm{p}} = \sum_{i=1}^m \gamma_i \cdot f_i^{\mathrm{p}}$ for (6).

□

B. Prediction with Convex Polytopes

In line with Theorem 2, a corresponding conclusion holds for the prediction step. We define the convex polytope of estimated densities by $\boldsymbol{\mathcal{P}}^{\mathrm{e}} = \mathrm{conv}\{f_1^{\mathrm{e}}, \ldots, f_m^{\mathrm{e}}\}$ and accordingly the convex polytope of transition densities by $\boldsymbol{\mathcal{P}}^{\mathrm{T}} = \mathrm{conv}\{f_1^{\mathrm{T}}, \ldots, f_m^{\mathrm{T}}\}$. As before, $\boldsymbol{\mathcal{M}}^{\mathrm{p}}$ denotes the resulting set after elementwise processing with (4).

Theorem 4 (Prediction Step for Convex Polytopes)
The prediction step for sets of densities with the vertex sets $\boldsymbol{\mathcal{V}}(\boldsymbol{\mathcal{P}}^{\mathrm{e}})$ and $\boldsymbol{\mathcal{V}}(\boldsymbol{\mathcal{P}}^{\mathrm{T}})$ yields the vertices of $\boldsymbol{\mathcal{P}}^{\mathrm{p}}$, which is related to the exact set $\boldsymbol{\mathcal{M}}^{\mathrm{p}}$ by means of the equation

$$\boldsymbol{\mathcal{P}}^{\mathrm{p}} = \mathrm{conv}\{\boldsymbol{\mathcal{V}}(\boldsymbol{\mathcal{P}}^{\mathrm{p}})\} = \mathrm{conv}\{\boldsymbol{\mathcal{M}}^{\mathrm{p}}\} \ .$$

PROOF. Every $f^{\mathrm{e}} \in \boldsymbol{\mathcal{P}}^{\mathrm{e}}$ and every $f^{\mathrm{T}} \in \boldsymbol{\mathcal{P}}^{\mathrm{T}}$ can be written as

$$f^{\mathrm{e}} = \sum_{i=1}^m \alpha_i \cdot f_i^{\mathrm{e}} \quad \text{and} \quad f^{\mathrm{T}} = \sum_{j=1}^n \beta_i \cdot f_j^{\mathrm{T}} \ ,$$

respectively. Due to the linearity of integration, it follows directly that

$$\begin{aligned} f^{\mathrm{p}} &= \int_\Omega f^{\mathrm{T}}(\,\cdot\,|\underline{x}_k) \cdot f^{\mathrm{e}}(\underline{x}_k)\ \mathrm{d}\underline{x}_k \\ &= \sum_{i=1,j=1}^{m,n} \alpha_i \beta_j \int_\Omega f_j^{\mathrm{T}}(\,\cdot\,|\underline{x}_k) \cdot f_i^{\mathrm{e}}(\underline{x}_k)\ \mathrm{d}\underline{x}_k \ , \end{aligned} \tag{8}$$

where $\sum_{i=1,j=1}^{m,n} \alpha_i \beta_j = 1$ and $\int_\Omega f_j^{\mathrm{T}}(\,\cdot\,|\underline{x}_k) \cdot f_i^{\mathrm{e}}(\underline{x}_k)\ \mathrm{d}\underline{x}_k$ are the vertices of $\boldsymbol{\mathcal{P}}^{\mathrm{p}}$. Hence, the predicted density f^{p} lies in $\boldsymbol{\mathcal{P}}^{\mathrm{p}}$ as anticipated. □

In the following theorem, we again consider the situation of processing only one probility density function f^{e} or f^{T} instead of $\boldsymbol{\mathcal{P}}^{\mathrm{e}}$ or $\boldsymbol{\mathcal{P}}^{\mathrm{T}}$.

Theorem 5
If $\boldsymbol{\mathcal{P}}^{\mathrm{e}}$ or $\boldsymbol{\mathcal{P}}^{\mathrm{T}}$ is singleton, then $\boldsymbol{\mathcal{M}}^{\mathrm{p}}$ is convex and equal to $\boldsymbol{\mathcal{P}}^{\mathrm{p}}$.

PROOF. Considering (8), one can easily accept that convexity will be preserverd, when the cardinality of $\boldsymbol{\mathcal{P}}^{\mathrm{e}}$ or $\boldsymbol{\mathcal{P}}^{\mathrm{T}}$ amounts to one. □

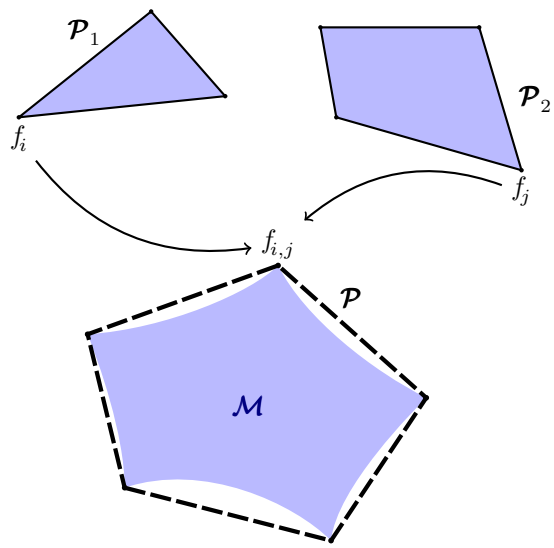


Figure 4. Elementwise processing of the vertices $\mathcal{V}(\mathcal{P}_1)$ and $\mathcal{V}(\mathcal{P}_2)$ yields the vertices of $\mathcal{P}$ which is the convex hull of $\mathcal{M}$. The exact set $\mathcal{M}$ arises from elementwise processing of $\mathcal{P}_1$ and $\mathcal{P}_2$.

C. Discussion

Elementwise processing of two convex sets of density functions usually leads to a non–convex set $\mathcal{M}$. With Bayesian estimation, this occurs when combining sets of prior densities and sets of likelihoods, or transition densities. It was shown that after each filter step or prediction step the convex hull of the exact set can easily be found by restricting oneself to the vertices of the polytopes.

Especially (8) points out that in general the exact set $\mathcal{M}$ is a proper subset of the corresponding polytope $\mathcal{P}$. This is due to the fact that this equation can be interpreted as a quadratic function of the weights. Considering the convex hull $\mathcal{P}$ instead of the non–convex resulting set $\mathcal{M}$, implies that density functions will be introduced after each processing step, which are not part of the exact set, as illustrated in Figure 4. Fortunately, the use of the covering polytope for further processing is appropriate, since it leads to exactly the same probability assessment. This is due to the linearity of the integral, which yields the probability from the densities.

One key statement of this paper is that after multiple recursive processing steps the result of the proposed estimator is the convex hull of the exact set given by elementwise processing. This means that applying k processing steps to $\mathcal{P}$ and taking the convex hull afterwards results in the same set as processing the vertices $\mathcal{V}(\mathcal{P})$ and taking the convex hull of them at all time steps until k. This is mainly due to the fact that the resulting vertices are images of the vertices of $\mathcal{P}$.

Other approaches working on sets of densities commonly just use sets of prior probabilities, whereas likelihoods and transition densities are fixed. In this particular case convexity will be preserved due to Theorem 3 and 5. In terms of convex polytopes, we have proposed a method of modelling system and measurement noise as sets as well.

An important point to emphasize is the fact that the vertices of the resulting polytope are in general affinely independent. Then, the polytope is a simplex, which cannot be simplified to a polytope with less vertices. The example in the following section shows a special case, where some of the vertices after each processing step coincide.

In conclusion, this section shows that only the finite sets of vertices are required to implement a Bayesian estimator for convex polytopes, though it can be interpreted as element-wise processing of an infinite number of densities. Unfortunately, the number of vertices needed to characterize the convex set increases exponentially over the number of processing steps, depending on the number of vertices of the prior sets. Processing two convex sets with n and m vertices usually leads to a set whose convex polytope is characterized by $n \cdot m$ vertices.

V. Example

To illustrate the idea of an estimator for sets of densities, a simple example is discussed in this section. [23] presents a problem, where a radar altimeter is utilized for measuring the ground clearance of a plane. The measurements differ significantly when flying over trees in contrast to flying over clear ground. The likelihood

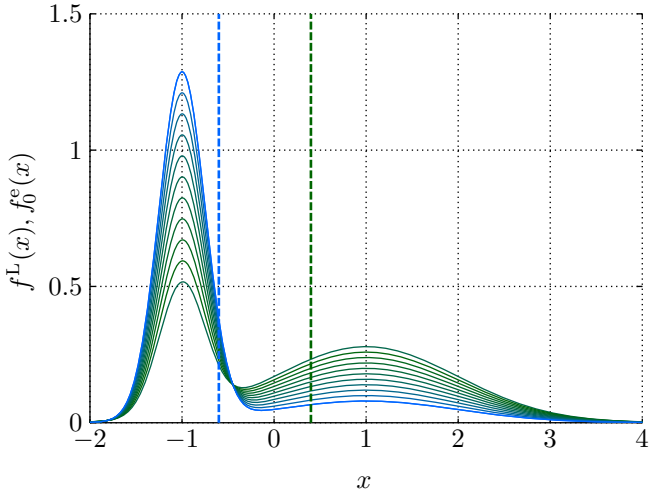
$$f_k^{\mathrm{L}} = \pi \mathcal{N}(\hat{x}_1, \sigma_1^2) + (1-\pi)\mathcal{N}(\hat{x}_2, \sigma_2^2) \tag{9}$$

is suggested, where $\mathcal{N}(\hat{x}, \sigma^2)$ denotes the probability density funcion of the normal distribution with standard deviation σ and expected value $\hat{x}$. When flying over a tree-covered area, $\mathcal{N}(\hat{x}_1, \sigma_1^2)$ describes the likelihood of an echo $\boldsymbol{x}$ over clear ground and $\mathcal{N}(\hat{x}_2, \sigma_2^2)$ the echo over tree tops. In contrast to [23], we assume the probability π of being over clear grounds to be uncertain in a sense of ignorance.

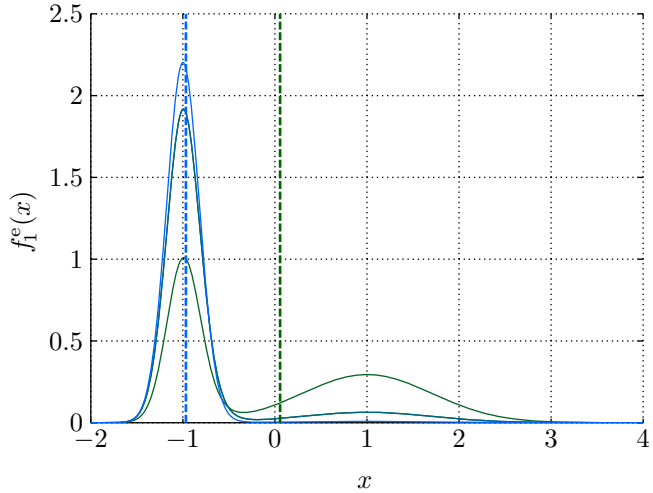
For our example, we assume having three identical measurements, which should be fused. The likelihood has the parameters $\hat{x}_1 = -1$, $\sigma_1^2 = 1/4$, $\hat{x}_2 = 1$, $\sigma_1^2 = 1$ and the parameter π lies in the interval $[0.3, 0.8]$. According to (9) the set is defined as

$$\begin{aligned}\mathcal{P}^{\mathrm{L}} = \Big\{ f_k^{\mathrm{L}} \big| f_k^{\mathrm{L}}(\,\cdot\,) = \lambda \left(0.3 \cdot \mathcal{N}(-1, \frac{1}{4}) + 0.7 \cdot \mathcal{N}(1,1) \right) + \\ (1-\lambda) \left(0.8 \cdot \mathcal{N}(-1, \frac{1}{4}) + 0.2 \cdot \mathcal{N}(1,1) \right), \lambda \in [0,1] \Big\} .\end{aligned}$$

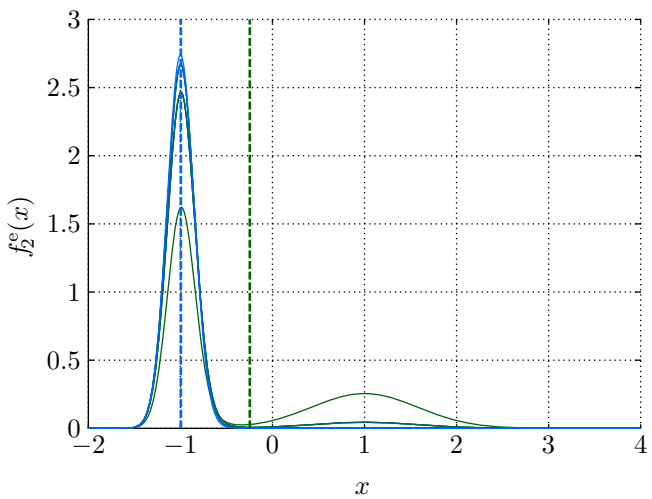
In the first step, the likelihood set $\mathcal{P}^{\mathrm{L}}$ is fused with a uniform distribution. The result is naturally the same set, which is depicted in Figure 5(a). The multiplication with itself results in a set with four vertices of four-component Gaussian mixtures as plotted in Figure 5(b). In the plot, only three components can be seen, since two are identical. This is because of the symmetries within the chosen sets due to the simplicity of the problem. Figure 5(c) shows the second fusion step. Now the set of vertices consists of eight six-component Gaussian mixtures. Again, only four different densities can be distinguished, since some of the densities are identical. The



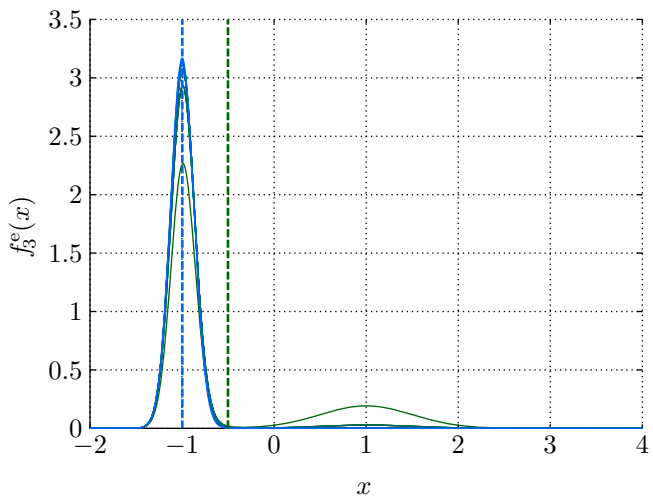
(a) The initial set of likelihood densities. The two vertex densities of the polytope and nine densities in-between are plotted. The polytope was fused with a uniform density, so the result after the first step is identical with the likelihood set. The interval of the expectation, depicted by the vertical dashed lines, is $[-0.6, 0.4]$.



(b) After the first fusion step. Now only the resulting three vertex densities are plotted and the expectation interval, depicted by the vertical dashed lines, is $[-0.97, 0.06]$.



(c) After the second fusion step. Only the resulting four vertex densities are plotted and the expectation interval, depicted by the vertical dashed lines, is $[-0.99, -0.25]$.



(d) After the third fusion step. Only the resulting five vertex densities are plotted and the expectation interval, depicted by the vertical dashed lines, is $[-1.0, -0.5]$.

Figure 5. Four fusion steps with $\boldsymbol{\mathcal{P}}^{\mathrm{L}}$. The prior distribution is uniform. The number of vertices increases by one after each step, but in general the number increases exponentially.

final step ist depicted in Figure 5(d). The resulting polytope has five vertices.

Though some of the vertex densities are projected onto each other, it can be seen that the use of convex polytopes yields the same key problem as standard stochastic nonlinear filters: The complexity of the density representation typically increases exponentially with each filtering step. Here, the number of vertices and the number of Gaussian components increase with each step. Coping with this increase of complexity is part of future work.

Considering the intervals of the expectations, which are depicted as vertical lines, it can be seen that the width stays constant from the initial to the first step and decreases from the first to the second step. This implies that ignorance decreases at a different rate as the stochastic uncertainty, which typically decreases strictly monotonically and is not discussed here. The investigation of the convergence of Bayesian estimators with sets of densities is another interesting point of further research.

VI. Conclusions and Future Work

In this work, a theoretical framework for processing sets of probability densities in the context of Bayesian estimation was presented. Since a generic Bayesian estimator for density sets is intractable, convex sets were examined. To keep the mathematical expressions simple, only convex polytopes were discussed. It was shown that enlarging a non-convex set to

its convex hull does not increase the extent of ignorance, i. e., the probability intervals of arbitrary events. The convex hull is merely the larger set, which constitutes the same probability assessment. Therefore the utilization of convex sets is sufficient.

A main drawback of using convex polytopes lies in the fact that the number of vertices grows exponentially during processing, which is a direct consequence of Theorems 2 and 4. There exists no straightforward method to simplify the resulting polytopes, since the function space is infinite-dimensional and the polytopes are thus typically simplices. A direction of future work might be to employ Gaussian mixture or Fourier representations. Due to approximation, vertices can therefore become affinely dependend, which allows reducing the number of vertices. It should be mentioned that the herein used sets are then sets of approximate functions and thus do not necessarily contain the true probability density.

Another important point to emphasize is that convex polytopes are sets of mixture densities. Every function inside a polytope is a convex combination of the vertices, which can themselves be mixture densities. So for some applications the number of vertex densities tend to be infinite. An example is a density set where the mean value is unknown but assumed to lie in a specific interval. In this regard we desire to model convex set of translated densities, which have no polytopic representation.

Another application is decentralized data fusion. At this, information fusion is performed on several processing nodes, which usually are spatially distributed. Information in terms of measurements is incorporated into an estimate in each processing node. This estimate is transmitted to the other nodes that incorporate this data into their own estimates. Due to the lack of knowledge about which measurements are incorporated into the estimate, multiple processing of the same measurement can occur, which leads to unknown correlations. Approaches handling these unknown correlations are, e. g., covariance hulls and covariance intersection filters [24], [25], where marginal densities with unknown correlation coefficients are predestinated for a set-based model. An examination in the light of the results of this work seems promising.

Further research will focus on more general models of convex sets to which the presented results can directly be applied.

References

[1] R. E. Kalman, "A new Approach to Linear Filtering and Prediction Problems," *Transactions of the ASME, Journal of Basic Engineering*, vol. 82, pp. 35–45, 1960.

[2] R. Kulhavy, "Can We Preserve the Structure of Recursive Bayesian Estimation in a Limited-Dimensional Implementation?" in *International Symposium on the Mathematical Theory of Networks and Systems*, 1993, pp. 1–20.

[3] H. W. Sorenson, "Least-Squares Estimation: From Gauss to Kalman," *IEEE Spectrum*, vol. 7, pp. 63–68, 1970.

[4] S. J. Julier and J. K. Uhlmann, "Unscented Filtering and Nonlinear Estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, 2004.

[5] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A Tutorial on Particle Filters for On-line Non-linear/Non-Gaussian Bayesian Tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.

[6] F. Daum and J. Huang, "Curse of Dimensionality and Particle Filters," *Aerospace Conference, 2003. Proceedings. 2003 IEEE*, vol. 4, pp. 1979–1993, Mar. 2003.

[7] M. Šimandl and O. Straka, "Sampling Densities of Particle Filter: A Survey and Comparison," in *American Control Conference (ACC '07)*, 2007, pp. 4437–4442.

[8] K. Srinivasan, "State Estimation by Orthogonal Expansion of Probability Distributions," *IEEE Transactions on Automatic Control*, vol. 15, no. 1, pp. 3–10, 1970.

[9] S. Challa, Y. Bar-Shalom, and V. Krishnamurthy, "Nonlinear Filtering via Generalized Edgeworth Series and Gauss-Hermite Quadrature," *IEEE Transactions on Signal Processing*, vol. 48, no. 6, pp. 1816–1820, Jun. 2000.

[10] R. Kronmal and M. Tarter, "The Estimation of Probability Densities and Cumulatives by Fourier Series Methods," *Journal of the American Statistical Association*, vol. 63, no. 323, pp. 925–952, Sep. 1968. [Online]. Available: http://links.jstor.org/sici?sici=0162-1459%28196809%2963%3A323%3C925%3ATEOPDA%3E2.0.CO%3B2-2

[11] 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.

[12] D. L. Alspach and H. W. Sorenson, "Nonlinear Bayesian Estimation using Gaussian Sum Approximations," *IEEE Transactions on Automatic Control*, vol. 17, no. 4, pp. 439–448, Aug. 1972.

[13] O. C. Schrempf, D. Brunn, and U. D. Hanebeck, "Density Approximation Based on Dirac Mixtures with Regard to Nonlinear Estimation and Filtering," in *Proceedings of the 2006 IEEE Conference on Decision and Control (CDC 2006)*, San Diego, California, Dec. 2006.

[14] 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)*, Quebec, Canada, Jul. 2007.

[15] K. Weichselberger, *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung I*. Heidelberg: Physica-Verlag, 2001.

[16] J. O. Berger, "An Overview of Robust Bayesian Analysis," *Test*, vol. 3, pp. 5–124, 1994.

[17] J. O. Berger, D. R. Insula, and F. Ruggeri, "Bayesian Robustness," in *Robust Bayesian Analysis*, ser. Lecture Notes in Statistics. New York: Springer-Verlag, 2000, no. 152, ch. 1, pp. 1–32.

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

[19] F. G. Cozman, "A Brief Introduction to the Theory of Sets of Probability Measures," Robotics Institute, Carnegie Mellon University, Tech. Rep., 1999. [Online]. Available: http://www.cs.cmu.edu/~fgcozman/qBayes.html

[20] I. Levi, "Imprecision and Indeterminacy in Probability Judgment," *Philosophy of Science*, vol. 52, no. 3, pp. 390–409, sep 1985. [Online]. Available: http://links.jstor.org/sici?sici=0031-8248%28198509%2952%3A3%3C390%3AIAIIPJ%3E2.0.CO%3B2-H

[21] D. R. Morrel and W. C. Stirling, "Set-valued filtering and smoothing," *IEEE Transactions on systems, man and cybernetics*, vol. 21, no. 1, pp. 184–193, Jan. 1991.

[22] J. Kenney and W. Stirling, "Nonlinear Filtering of Convex Sets of Probability Distributions," *Journal of statistical planning and inference*, vol. 105, no. 1, pp. 123–147, 2002. [Online]. Available: http://www.ee.byu.edu/faculty/wynns/jspi3.ps

[23] T. Schön, F. Gustafsson, and P.-J. Nordlund, "Marginalized Particle Filters for Nonlinear State-space Models," Division of Communication Systems, Department of Electrical Engineering, Linköpings University, Tech. Rep., Oct. 2003.

[24] U. D. Hanebeck, K. Briechle, and J. Horn, "A Tight Bound for the Joint Covariance of Two Random Vectors with Unknown but Constrained Cross–Correlation," in *Proceedings of the 2001 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2001)*, Baden–Baden, Germany, Aug. 2001, pp. 85–90.

[25] L. Chen, P. O. Arambel, and R. K. Mehra, "Fusion under Unknown Correlation – Covariance Intersection as a Special Case," in *Proceedings of The Fifth International Conference on Information Fusion*, 2002.