

# Approximate Nonlinear Bayesian Estimation Based on Lower and Upper Densities

Vesa M. Klumpp, Dietrich Brunn and Uwe D. Hanebeck

Intelligent Sensor-Actuator-Systems Laboratory  
Institute of Computer Science and Engineering

Universität Karlsruhe (TH)

Karlsruhe, Germany

{*vklumpp,brunn*}@ira.uka.de, *uwe.hanebeck@ieee.org*

**Abstract** – *Recursive calculation of the probability density function characterizing the state estimate of a nonlinear stochastic dynamic system in general cannot be performed exactly, since the type of the density changes with every processing step and the complexity increases. Hence, an approximation of the true density is required. Instead of using a single complicated approximating density, this paper is concerned with bounding the true density from below and from above by means of two simple densities. This provides a kind of guaranteed estimator with respect to the underlying true density, which requires a mechanism for ordering densities. Here, a partial ordering with respect to the cumulative distributions is employed. Based on this partial ordering, a modified Bayesian filter step is proposed, which recursively propagates lower and upper density bounds. A specific implementation for piecewise linear densities with finite support is used for demonstrating the performance of the new approach in simulations.*

**Keywords:** Nonlinear Bayesian Estimator, Bounding density, Lower and upper bound, Cubic Sensor Problem

## 1 Introduction

The task of determining unknown quantities through indirect measurements in nonlinear systems is a key problem in digital signal processing. From a theoretical point of view, this problem is solved by the Bayesian Estimator or Bayesian filter. Practical implementations of this algorithm only exist for special systems.

In this paper, we introduce a novel approach to propagate sets of probability density functions through a Bayesian Estimator. This is achieved by defining lower and upper densities for sets of densities, which are passed through the filtering algorithm, to obtain new lower and upper bounds for the resulting estimated density. Several works have been done in processing sets of probability distributions in a Bayesian way. The field of *robust Bayesian analysis* [1] deals with sets of probability distributions processed by Bayesian inference and the dependence of the posterior to the prior distributions and the likelihood function

respectively. Much work has been done on this topic [2] and the theory is well established, but no convincing algorithm has been proposed so far [6]. Furthermore, the problem of state prediction is not solved with these results. Another approach to state estimation is based on interval analysis [3]. The problem here is the discretization of sample space.

By discretization no guaranteed probability distribution bounds of the continuous sample space can be given. The same problem holds for the Dempster-Shafer-Theory [4] and operations on the Dempster-Shafer structures [5]. This will be shown by an example: Consider a Gaussian probability density function  $f(x) = \mathcal{N}(0, 2)$  with mean 0 and variance 2, and a sample space  $\Omega = [-5, 5]$ . Now a discretization of the sample space is given by the intervals  $A := [-5, -2.5)$ ,  $B := [-2.5, 0)$ ,  $C := [0, 2.5)$ , and  $D := [2.5, 5]$ . For the intervals, the probability is given as  $P(A) = 0.1$ ,  $P(B) = 0.4$ ,  $P(C) = 0.4$ , and  $P(D) = 0.1$ . With this information, a probability box as described in [5] can be determined. It is shown in Figure 1. The cumulative distribution function  $F$ , an upper distribution function  $U$ , and a lower distribution function  $L$  are shown. With the intervals and the corresponding probabilities, a probability box can be determined that bounds the cumulative distribution function  $F$ . Performing discrete Bayes' law with the probabilities

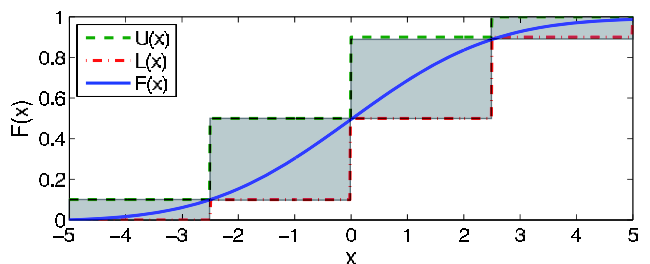


Figure 1: Distribution function  $F$  with its probability box  $U, L$ .

of  $f$  both as prior and likelihood results in a new probability box described by  $U'$  and  $L'$ . The corresponding posterior probability density is  $f'(x) = \mathcal{N}(0, \sqrt{2})$  in continuous space. Discrete Bayes' law agrees with this result in that the probability box bounds the continu-

ous distribution  $F'$ . If the discretization of the sample space is changed in such a way that the intervals have different widths, the result is less satisfactory. As an example the intervals  $E_1 := [0, 1)$  with  $P(E_1) = 0.19$  and  $E_2 := [1, 2.5)$  with  $P(E_2) = 0.21$  are defined. The partitioning of  $\Omega$  is now  $\{A, B, E_1, E_2, D\}$ . Applying Bayes' law leads to a different solution shown in Figure 2. Here the probability box does not bound the function  $F'$  and so no guaranteed upper and lower distribution functions are given. This can be seen at the location 0.5. Furthermore, it can be shown that even

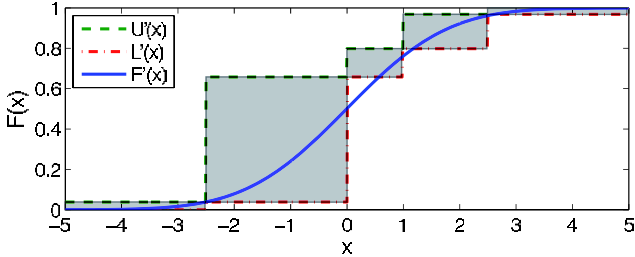


Figure 2: Distribution function  $F'$  with its probability box  $U', L'$ . The probability box does not bound the function  $F'$ .

for equal interval widths the bounding property of the probability boxes is not preserved for continuous densities. The same applies to Dempster-Shafer structures. For that reason they are not adequate for this estimation problem.

The use of bounds is able to simplify the Bayesian estimation problem. In general, the complexity of a density representation increases with a rising number of measurements. Furthermore, the expressions usually cannot be calculated in closed form and the density type changes with each processing step. Only linear systems with Gaussian noise can be evaluated analytically with a constant complexity by employing the well-known Kalman Filter [8]. To overcome the problem of increasing complexity, the use of samples instead of whole densities is employed [9, 10]. The so called particle filters are very efficient from an algorithmic point of view. Unfortunately, it is very difficult to guarantee a sufficient precision of the estimate. An alternative approach is the use of generic parameterized density functions, like Gaussian mixtures [11]. The Gaussian mixture filtering algorithms themselves are also very efficient. Approaches exist for approximating arbitrary density functions [12], but they are computationally very costly. Our key idea is to sacrifice a bit of precision by introducing lower and upper density bounds. In return, we obtain much simpler density functions we can handle. Furthermore, if a parameterized representation of the bounding densities is used, the complexity of the representation can be easily kept at a constant level.

Another aspect for the use of density sets is that in technical systems the exact likelihood is often unknown or cannot be obtained with reasonable effort. Numerical deviations, sensor drift or complex measurement equipment are some common examples. Since these

errors are typically of a mixed stochastic-deterministic nature, a set based approach has to be pursued. The use of bounds would be an efficient method for dealing with such stochastic-deterministic errors.

The following section reviews the time-discrete Bayesian Estimator and discusses the structure of filter employing bounds. The rest of this paper is structured as follows: In Section 3, a more rigorous definition of lower and upper densities will be given. Then a construction algorithm of a set containing a given or partially known probability density function will be derived. It is shown what information is required and how to derive lower and upper probability density functions. Section 4 describes the modifications of the Bayesian filter step to propagate such bounds: The lower and upper bounds of two sets are combined to obtain new bounds. The second part of the estimator, the prediction step is shown in Section 5. Section 6 investigates the cubic sensor problem as an example application. This paper closes with a conclusion and an outlook on future work.

## 2 Problem Formulation

For the sake of clarity and brevity, merely scalar stochastic variables will be considered which are denoted by boldface, e.g.  $\mathbf{x}$ .

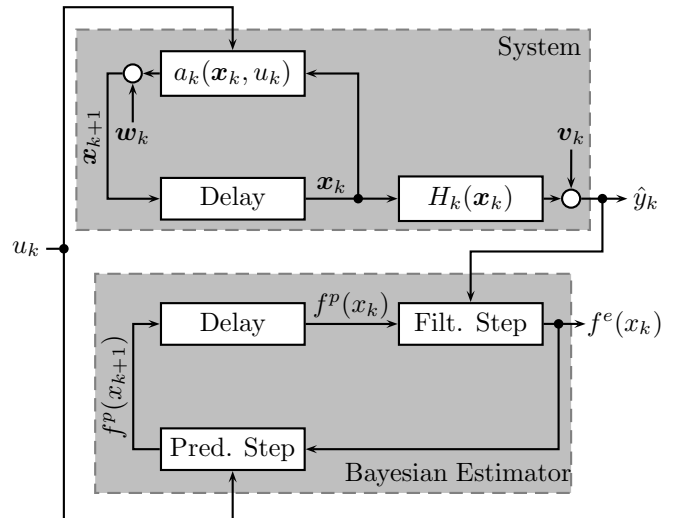


Figure 3: Structure of a time discrete nonlinear system with input  $u_k$ , output  $\hat{y}_k$  and a generic Bayesian Estimator determining  $f^e(x)$  for the state variable  $\mathbf{x}$ .

We consider the following nonlinear time discrete system

$$\begin{aligned} \mathbf{x}_{k+1} &= a_k(\mathbf{x}_k, u_k) + \mathbf{w}_k \\ \mathbf{y}_k &= h_k(\mathbf{x}_k) + \mathbf{v}_k \end{aligned} \quad (1)$$

with the not directly observable state variable  $\mathbf{x}_k$ , the deterministic input  $u_k$ , the scalar measurement  $\mathbf{y}_k$ , and the additive noise terms  $\mathbf{v}_k, \mathbf{w}_k$  which have the densities  $f_v(v_k), f_w(w_k)$ . Note that an realization of  $\mathbf{y}_k$ , i.e. an actual measurement, is denoted by  $\hat{y}_k$ .

The Bayesian Estimator generates an estimate for  $\mathbf{x}_k$  in the form of a density  $f^e(x_k)$  for each discrete

time  $k$ . As shown in Figure 3, the estimator consists of the *filter step* and the *prediction step*.

The purpose of the *filter step* is to enhance a given estimate  $f^p(x_k)$  with the information of the measurement value  $\hat{y}_k$ . This is accomplished by determining the likelihood [8]

$$f^L(x_k) = f_v(\hat{y}_k - h_k(x_k))$$

and then using Bayes' law

$$f^e(x_k) = \frac{f^L(x_k)f^p(x_k)}{\int_{\mathbb{R}} f^L(\xi_k)f^p(\xi_k)d\xi_k} \quad (2)$$

to receive the desired estimated density  $f^e(x_k)$ . It is obvious that by recursive application  $f^e(x_k)$  gets more and more complicated for almost any type of density function. The prominent exception is the Gaussian density with a linear measurement equation.

The *prediction step* is used to propagate a current estimate  $f^e(x_k)$  to the next time step  $k+1$ . It can be shown that

$$f^p(x_{k+1}) = \int_{\mathbb{R}} f_w(x_{k+1} - a_k(\xi_k))f^e(\xi_k)d\xi_k$$

is an optimal predicted density with respect to equation (1). A solution for value discrete densities can be found in [13].

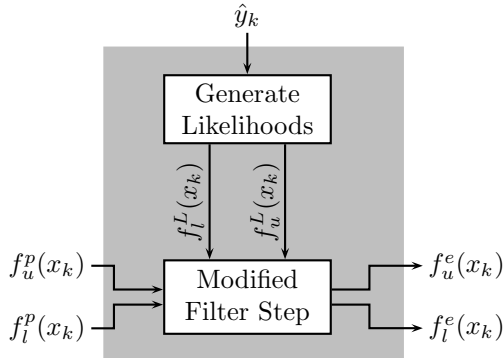


Figure 4: Modified filter step for the use with lower and upper bounds.

Our approach is to modify the filter step for propagating a set of densities. As shown in Figure 2, two likelihoods, one lower and one upper, are generated. The generation of these likelihood depend on the non-stochastic errors of the system. Efficient methods of deriving density bounds from real systems are another point of future research. These two likelihoods plus the lower and upper bounds of the prior density  $f_{u,l}^p(x)$  are used to calculate a lower and upper bounding estimate density  $f_{u,l}^e(x)$ .

### 3 Lower and Upper Densities

For defining lower and upper bounds a partial ordering over the cumulative probability distributions is introduced. This approach is much more intuitive than using the density function, since cumulative distributions

are monotonous and intervals for probabilities can easily be derived.

DEFINITION 3.1 *Given are two probability density functions  $f_1(x) \neq f_2(x)$ . If*

$$F_1(x) = \int_{-\infty}^x f_1(\xi)d\xi \leq \int_{-\infty}^x f_2(\xi)d\xi = F_2(x)$$

*holds for every  $x$ , then  $f_2(x)$  is called an upper density of  $f_1(x)$ .  $f_1(x)$  is obviously a lower density of  $f_2(x)$ .*  $\square$

For investigating sets of probabilities, bounding densities have to be defined.

DEFINITION 3.2 *Given is a set of densities  $\Phi$  over  $x$ . A density  $f_u$  is called upper bounding density on  $\Phi$ , if for every density  $f_\Phi \in \Phi$  and for every  $x \in \Omega$*

$$F_\Phi(x) = \int_{-\infty}^x f_\Phi(\xi)d\xi \leq \int_{-\infty}^x f_u(\xi)d\xi = F_u(x)$$

*holds. A lower bounding density is defined analogously.*  $\square$

The ideal thing would be to know the density which tightly bounds a given set.

DEFINITION 3.3 *Given is an upper bounding density  $f_u(x)$  on  $\Phi$ .  $f_u(x)$  is called an upper tight bounding density, if there exists for every  $x_0 \in \mathbb{R}$  a  $f_\Phi \in \Phi$  so that*

$$f_\Phi(x_0) = f_u(x_0)$$

*holds. A lower tight bounding density is defined analogously.*  $\square$

It is not possible to derive tight bounding densities for arbitrary sets of densities. But we are able to show that bounding densities can be derived for a large class of continuous density functions which are close enough the tight bounding density to be used for nonlinear estimation.

#### 3.1 A Specific Class of Boundable Probability Density Functions

This work considers one-dimensional probability densities. The following assumptions are made for all density functions  $f$ :

- Two scalars  $\alpha$  and  $\beta$ ,  $\alpha < \beta$ , exist with  $f(x) = 0$  for all  $x < \alpha$  and  $x > \beta$ . So  $f$  only has values larger than 0 inside the interval  $[\alpha, \beta]$ .
- The density functions must be continuous. This means that maximum and minimum values of  $f$  exist for every set in  $[\alpha, \beta]$ . The same assumption applies to its derivative  $\frac{\partial}{\partial x}f(x)$  also.

In order to compute the bounds the density functions need not to be completely known. The following information is required:

- A partitioning of  $[\alpha, \beta]$  in intervals  $[x_i, x_{i+1}]$  with  $x_1 = \alpha$ ,  $x_{n+1} = \beta$  and  $x_i < x_{i+1}$  for all  $i = 1 \dots n$ . So all intervals  $[x_i, x_{i+1}]$  are disjunct except their borders  $x_i$  and their union is  $[\alpha, \beta]$ .

- The function values of  $f$  at all locations  $x_i$ ,  $x = 1 \dots n + 1$ . This assumption can be loosened by assuming an interval for the function values  $f(x_i)$ . This is part of future work.
- The maximum and minimum bounds of the considered probability density function  $f$  within  $[x_i, x_{i+1}]$

$$c_i \geq \max_{x \in [x_i, x_{i+1}]} \{f(x)\} ,$$

$$d_i \leq \min_{x \in [x_i, x_{i+1}]} \{f(x)\} .$$

- The maximum and minimum bounds of the derivatives of the probability density functions:

$$a_i \geq \max_{x \in [x_i, x_{i+1}]} \left\{ \frac{\partial}{\partial x} f(x) \right\} ,$$

$$b_i \leq \min_{x \in [x_i, x_{i+1}]} \left\{ \frac{\partial}{\partial x} f(x) \right\} .$$

First, only one interval  $[x_i, x_{i+1}]$  is considered. The idea is to determine an upper bound  $f_u$  for the density function  $f$  with the known parameters  $f(x_i)$ ,  $f(x_{i+1})$ ,  $a_i$ ,  $b_i$ ,  $c_i$ , and  $d_i$ . In this work it is bounded with 3 straight lines. Other functions like higher order polynomials or Gaussian distribution functions are also possible. The first one goes through the point  $(x_i | f(x_i))$  with maximum slope  $a_i$ . The second one has a constant value of  $c_i$ . The third straight line goes through  $(x_{i+1} | f(x_{i+1}))$  and has the minimum slope  $b_i$ . The upper estimation  $f_u$  of  $f$  is now the minimum of these three straight lines.

$$f_u(x) = \min \left\{ \begin{array}{l} a_i \cdot (x - x_i) + f(x_i), \quad c_i, \\ b_i \cdot (x - x_{i+1}) + f(x_{i+1}) \end{array} \right\} \quad (3)$$

$f_u$  is displayed in Figure 5.

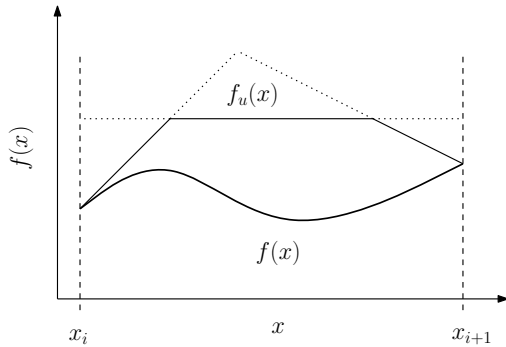


Figure 5: Upper bound  $f_u$  of a probability density function  $f$  with three straight lines.

The lower estimation  $f_l$  of  $f$  is given in a similar way again with 3 straight lines. In comparison to the upper estimation the slopes  $a_i$  and  $b_i$  are now interchanged and the second straight line has a constant value of the minimum  $d_i$ . To obtain the lower bound of  $f$  the maximum of the 3 lines is needed:

$$f_l(x) := \max \left\{ \begin{array}{l} b_i \cdot (x - x_i) + f(x_i), \quad d_i, \\ a_i \cdot (x - x_{i+1}) + f(x_{i+1}) \end{array} \right\} \quad (4)$$

$f_l$  is displayed in Figure 6.

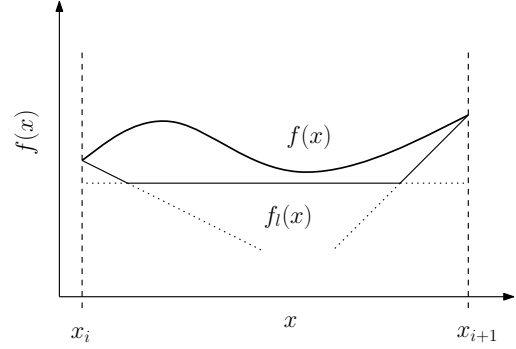


Figure 6: Lower bound  $f_l$  of a probability density function  $f$  with three straight lines.

Now the estimates  $f_l$  and  $f_u$  can easily be calculated for all  $x$  in  $[\alpha, \beta]$ .  $f_u$  and  $f_l$  are not valid probability density functions because none of their integrals is guaranteed to be equal to 1. Only  $\int f_u(\xi) d\xi \geq \int f(\xi) d\xi = 1$  and  $\int f_l(\xi) d\xi \leq \int f(\xi) d\xi = 1$  are fulfilled. How to get valid probability density functions is discussed in the next Subsection.

### 3.2 Deriving Valid Lower and Upper Bounding Densities

To obtain valid density and distribution functions only the upper estimation  $f_u$  is needed. If we compare  $\int f_u(\xi) d\xi$  to  $\int f(\xi) d\xi$  we can say that

$$\int_{-\infty}^x f_u(\xi) d\xi \geq \int_{-\infty}^x f(\xi) d\xi \quad \forall x \quad (5)$$

and especially

$$\int_{-\infty}^{\infty} f_u(\xi) d\xi \geq \int_{-\infty}^{\infty} f(\xi) d\xi$$

are fulfilled. To convert the integral of  $f_u$  to a valid probability distribution function it will be bounded to the maximum of 1. A scalar value  $v$  is determined which satisfies the equation

$$\int_{-\infty}^v f_u(\xi) d\xi = 1 , \quad (6)$$

which exists because  $\int f_u(\xi) d\xi \geq 1$ . Now the corresponding valid probability density function  $u$  is

$$u(x) := \begin{cases} f_u(x) & x \leq v \\ 0 & x > v \end{cases} . \quad (7)$$

The probability distribution function is then  $\int_{-\infty}^x u(\xi) d\xi$  which is an upper bound of the distribution function of  $f$ .

The lower boundary is determined as follows: A scalar value  $w$  with

$$\int_w^{\infty} f_u(\xi) d\xi = 1 \quad (8)$$

is defined. The resulting probability density function  $l$ , the lower bound, is then

$$l(x) := \begin{cases} 0 & x < w \\ f_u(x) & x \geq w \end{cases} . \quad (9)$$

Its probability distribution function is  $\int_{-\infty}^x l(\xi) d\xi$ .

We will prove that this term is actually the lower bound of the distribution function  $\int_{-\infty}^x f(\xi) d\xi$ , so

$$\int_{-\infty}^x l(\xi) d\xi \leq \int_{-\infty}^x f(\xi) d\xi \quad (10)$$

applies for all  $x$ . For  $x < w$  equation (10) is obviously fulfilled. For  $x \in [w, \beta]$ ,  $x$  is substituted by  $y - \beta$  and

$$\int_{-\infty}^{\beta-y} l(\xi) d\xi \leq \int_{-\infty}^{\beta-y} f(\xi) d\xi \quad (11)$$

is obtained. Differentiating both sides of equation (11) with respect to  $y$  we obtain

$$\frac{\partial}{\partial y} \left( \int_{-\infty}^{\beta-y} l(\xi) d\xi \right) = -l(\beta - y)$$

and

$$\frac{\partial}{\partial y} \left( \int_{-\infty}^{\beta-y} f(\xi) d\xi \right) = -f(\beta - y) .$$

Comparing both derivatives leads to

$$-l(\beta - y) \leq -f(\beta - y) \quad (12)$$

$$\Leftrightarrow l(x) \geq f(x)$$

$$\Leftrightarrow f_u(x) \geq f(x), \quad x \in [w, \beta] . \quad (13)$$

Inequality (13) is true. For  $y = 0$  with  $\int_{-\infty}^{\beta} l(\xi) d\xi = \int_{-\infty}^{\beta} f(\xi) d\xi = 1$  and equation (12) follows that equation (11) is fulfilled. For any  $x > \beta$  equation (10) is fulfilled using equation (9).

An example of the densities  $u$ ,  $l$ , and their distribution functions is shown in Figure 7. The density  $f(x)$  at the top of Figure 7 is a Gaussian density  $\mathcal{N}(0, 1)$ . 5 intervals have been defined with their borders  $-5$ ,  $-3$ ,  $-1$ ,  $1$ ,  $3$ , and  $5$  to calculate the bounds. The line segments of the lower density bound  $l$  and the upper density bound  $u$  can be seen clearly. The positions  $v$  and  $w$ , where the densities  $u$  and  $l$  are cut-off, are at about  $-0.7$  and  $0.7$ . At the bottom of Figure 7 the probability distributions  $F$ ,  $L$ , and  $U$  are shown. Here the order of the densities can be seen. Neither the distribution function  $L$  corresponding to the lower density bound  $l$  nor the distribution function  $U$  corresponding to the upper density bound  $u$  subtend the distribution function  $F$ .

## 4 Deriving a Modified Bayesian Filter Step

With the bounds introduced in the preceding section, it is possible to derive an algorithm to obtain guaranteed lower and upper bounds of the estimated resulting distribution function after a Bayesian filter step.

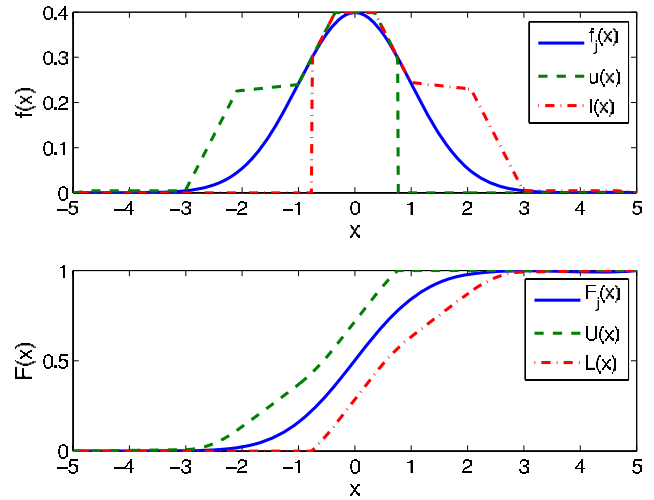


Figure 7: Top: density function  $f$  and corresponding lower and upper density bounds with 5 intervals. Bottom: distribution functions.

Passing the lower and upper bound directly to the filter would not produce the desired result because the order of the densities would not be conserved. The normalization factor given in the denominator of equation (2) would destroy the previously valid order. In the following a modified Bayesian filter step is introduced.

We assume that the parameters introduced in Section 3 of the functions  $f_1$  and  $f_2$  are known. First an approximation of the product  $f_1 \cdot f_2$  will be shown. All functions and variables relating to this product are denoted by an additional index 12.

Now the parameters  $a_{i,12}$ ,  $b_{i,12}$ ,  $c_{i,12}$ ,  $d_{i,12}$ , and the function values at the interval borders are needed to create the bounds. The function values at  $x_i$  are given as  $f_1(x_i) \cdot f_2(x_i)$  and are therefore known. The upper and lower bounds of the product can be estimated by  $c_{i,12} := c_{i,1} \cdot c_{i,2}$  and  $d_{i,12} := d_{i,1} \cdot d_{i,2}$  because they are all nonnegative. The bounds  $a_{i,12}$  and  $b_{i,12}$  of the derivative of  $f_1 \cdot f_2$  must comply with

$$a_{i,12} \geq \max_{x \in [x_i, x_{i+1}]} \{f_1'(x) \cdot f_2(x) + f_1(x) \cdot f_2'(x)\}$$

and

$$b_{i,12} \leq \min_{x \in [x_i, x_{i+1}]} \{f_1'(x) \cdot f_2(x) + f_1(x) \cdot f_2'(x)\} .$$

The parameters are then calculated as

$$\begin{aligned} a_{i,12} &:= \max\{a_{i,1} \cdot c_{i,2}, a_{i,1} \cdot d_{i,2}\} \\ &\quad + \max\{a_{i,2} \cdot c_{i,1}, a_{i,2} \cdot d_{i,1}\} \text{ and} \\ b_{i,12} &:= \min\{b_{i,1} \cdot c_{i,2}, b_{i,1} \cdot d_{i,2}\} \\ &\quad + \min\{b_{i,2} \cdot c_{i,1}, b_{i,2} \cdot d_{i,1}\} . \end{aligned}$$

With this information the bounds of a Bayesian filter step of the two functions  $f_1$  and  $f_2$ , labelled  $f_{12}$ , can be given. The lower and upper bounds will also be denoted by the additional index 12. To calculate the bounds an upper limit of the fraction in equation 2 is determined. This is done by calculating a lower

limit for the denominator and an upper limit for the numerator.

$L := \int_{-\infty}^{\infty} f_{l,12}(\xi) d\xi$  is an approximation of the normalization constant

$$\int_{-\infty}^{\infty} f_1(\xi) \cdot f_2(\xi) d\xi$$

with  $L \leq \int_{-\infty}^{\infty} f_1(\xi) \cdot f_2(\xi) d\xi$ .  $L$  can easily be computed. So an upper estimate of  $f_{12}$  is

$$\frac{f_1(x) \cdot f_2(x)}{\int_{-\infty}^{\infty} f_1(\xi) \cdot f_2(\xi) d\xi} \leq \frac{f_{u,12}(x)}{L}$$

and the parameters are  $a_{i,12} = \frac{a_{i,12}}{L}$ ,  $b_{i,12} = \frac{b_{i,12}}{L}$ ,  $c_{i,12} = \frac{c_{i,12}}{L}$ ,  $d_{i,12} = \frac{d_{i,12}}{L}$ , and the assumed function values at all locations  $x_i$  are  $\frac{f_1(x_i) \cdot f_2(x_i)}{L}$ . With these parameters lower and upper functions  $f_{l,12}$  and  $f_{u,12}$  can be calculated.

An example is given in Figure 8. The top figure shows the resulting density  $f_{12}$  after the Bayesian filter step with its lower and upper bounds  $l_{12}$  and  $u_{12}$ . The functions  $f_1$  and  $f_2$  are probability densities with

$$f_1(x) = \begin{cases} \left[ \frac{e^{-0.5x^2}}{\sqrt{2\pi}} + \frac{\sin^2(3x) \cdot 0.03}{5-|x|} \right] \cdot k & , x \in [-5, 5] \\ 0 & , x \notin [-5, 5] \end{cases}$$

a normalization constant  $k$ , and  $f_2 = \mathcal{N}(2, 1)$ .  $f_1$  is a composite of a Gaussian density and a sine curve and therefore rather complex and elaborate to handle. The bottom of Figure 8 shows the corresponding probability distributions  $F_{12}$ ,  $U_{12}$  and  $L_{12}$ .

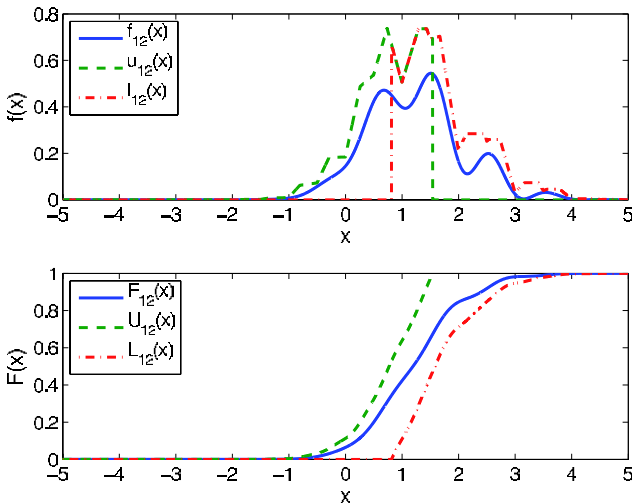


Figure 8: Resulting estimation of a Bayesian filter step of 2 probability densities with 20 intervals. Top: density functions, Bottom: distribution functions.

## 5 Deriving a Modified Prediction Step

With the information of the parameters  $a$ ,  $b$ ,  $c$ ,  $d$ , and  $f(x_i)$  of two probability distributions, the prediction

step described in Section 2 can be performed. The required computation is given by a convolution of the probability densities  $f_1$  and  $f_2$ .

A special case with  $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{w}_k$  is described here. Other system model equations can be handled by estimating the transitional density function  $f_w(\hat{y}_k - h_k(x_k))$  and processing it with  $f_k^e$ .

Regarding the segmentation of the sample space  $f_{1*2} = \int_{-\infty}^{\infty} f_1(\xi) \cdot f_2(x - \xi) d\xi$  can be written as

$$f_{1*2}(x) = \sum_{j=1}^n \int_{x_j}^{x_{j+1}} f_1(\xi) \cdot f_2(x - \xi) d\xi .$$

Maximum function values  $c_{1*2}$  can be approximated by  $\max_{x \in [x_i, x_{i+1}]} \{f_1(x) * f_2(x)\} \leq c_{i,1*2}$  and it holds

$$\begin{aligned} & \max_{x \in [x_i, x_{i+1}]} \{f_1(x) * f_2(x)\} \\ & \leq \sum_{j=1}^n \int_{x_j}^{x_{j+1}} f_{u,1}(\xi) \cdot \max_{x \in [x_i, x_{i+1}]} \{f_{u,2}(x - \xi)\} d\xi . \end{aligned}$$

The factor  $\max_{x \in [x_i, x_{i+1}]} \{f_{u,2}(x - \xi)\}$  can be bounded by a constant value  $c_{i,2} \geq \max_{x \in [x_i - x_{j+1}, x_{i+1} - x_j]} \{f_2(x)\}$  in the interval  $[x_i - x_{j+1}, x_{i+1} - x_j]$ . So the convolution is performed by integrating  $f_{u,1}$  multiplied by a constant factor. Because of the simple structure of  $f_{u,1}$  it is easy to integrate. In an analogous way the minimum function values  $d_{i,1*2}$  can be computed.

The minimal and maximal derivatives can be computed by considering

$$\frac{\partial}{\partial x} (f_1(x) \cdot f_2(x)) = \int_{-\infty}^{\infty} f_1(\xi) \cdot \frac{\partial}{\partial x} (f_2(x - \xi)) d\xi$$

and replacing the factor  $\frac{\partial}{\partial x} (f_2(x - \xi))$  by the corresponding upper or lower derivation  $a_i$  or  $b_i$ .

Assumed function values  $(f_1 * f_2)(x_i)$  at the interval borders can be estimated by computing the convolution interval directly. The product  $f_1 \cdot f_2$  is a polynomial of order 2 and is easily integrated.

These are all parameters needed in order to perform the prediction step.

## 6 Example: Cubic Sensor Problem

The cubic sensor problem is a well known problem which is extensively discussed in literature. It was introduced by Bucy in [7]. This work investigates the discrete-time cubic sensor problem with the measurement model equation

$$\mathbf{y}_k = \mathbf{x}_k^3 + \mathbf{v}_k \quad (14)$$

where  $\mathbf{v}_k$  is zero mean noise consisting of a Gaussian and a sine part and the system model equation

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{w}_k \quad (15)$$

at a discrete time step  $k$ .  $\mathbf{w}_k$  is also zero mean noise and constructed like  $\mathbf{v}_k$ . In order to perform backward

inference to estimate the system state  $\mathbf{x}_k$  at time step  $k$  given the measurement  $\hat{y}_k$  and a prior density function of the system state, the normalized product

$$f_k^e(x_k) = f(x_k|\hat{y}_k) = \frac{f^p(x_k) \cdot f_v(\hat{y}_k - x_k^3)}{\int_{-\infty}^{\infty} f^p(\xi) \cdot f_v(\hat{y}_k - \xi^3) d\xi} \quad (16)$$

analogous to equation (2) needs to be calculated. This problem can be reduced to estimating the likelihood  $f_v(\hat{y}_k - x_k^3)$  and then applying the modified Bayesian filter step discussed above.

First the parameters of the measurement noise function  $f_v$  need to be specified. They are denoted by the index  $v$ . Now a function  $f_{v,y-x^3}(x)$  is defined with  $f_{v,y-x^3}(x) := f_v(\hat{y} - x^3)$ . The parameters  $a_{i,v,y-x^3}$ ,  $b_{i,v,y-x^3}$ ,  $c_{i,v,y-x^3}$ ,  $d_{i,v,y-x^3}$ , and their function values at the interval borders are determined now.

The function values at the interval borders are given by  $f_v(\hat{y} - x_i^3)$  for all  $i = 1 \dots n + 1$ . The maximum and minimum function values can easily be determined by evaluating  $f_{u,v}$  between  $\hat{y}_k - x_i^3$  and  $\hat{y}_k - x_{i+1}^3$  for all  $i$ . The derivative

$$\frac{\partial}{\partial x} f_v(\hat{y} - x^3) = -3x^2 \cdot \frac{\partial}{\partial x} (f_v(\hat{y} - x^3))$$

is estimated as follows: The maximum and minimum slope of  $f_{u,v}$  within  $[\hat{y}_k - x_{i+1}^3, \hat{y}_k - x_i^3]$  are multiplied with the maximum and minimum values of  $-3x^2$ . In this case only the borders  $x_i$  and the root of  $-3x^2$  at  $x = 0$  have to be considered as extremal values. The derivatives of  $f_v$  are only known interval-wise with a lower and an upper bound, named  $a_{i,v}$  and  $b_{i,v}$ . All intervals  $j$  defined as  $[y - x_{j+1}^3, y - x_j^3]$  which have elements of  $[x_i, x_{i+1}]$  in common are in the set  $J_i$ .

We define the set

$$D_{i,j} := \left\{ -3x_i^2 \cdot a_{j,v}, -3x_{i+1}^2 \cdot a_{j,v}, -3x_i^2 \cdot b_{j,v}, -3x_{i+1}^2 \cdot b_{j,v} \right\},$$

which gives

$$a_{i,v,y-x^3} = \max_{\substack{x \in [x_i, x_{i+1}] \\ \wedge j \in J_i}} \{D_{i,j}\}$$

and

$$b_{i,v,y-x^3} = \min_{\substack{x \in [x_i, x_{i+1}] \\ \wedge j \in J_i}} \{D_{i,j}\},$$

whereas the value 0 has to be considered for both  $a_{i,v,y-x^3}$  and  $b_{i,v,y-x^3}$  if  $0 \in [x_i, x_{i+1}]$ .

An example upper estimation  $f_{u,v,y-x^3}(x)$  of the unnormalized density  $f_v(\hat{y} - x^3)$  with  $\hat{y} = 0.5$  and  $\sigma_v^2 = 1$  is shown in Figure 9. Note that the integrals of both functions are below 1. If they would be normalized individually the condition  $f_v(\hat{y} - x^3) \leq f_{u,v,y-x^3}(x)$  would not be met any more and invalid results could appear. Normalization is done in the modified filter step.

Figure 10 shows a complete measurement step with the parameters given above and a prior normal density  $\mathcal{N}(0,1)$ . It is clearly seen that the lower and upper probability density functions are cut-off and are normalized. Iterative estimations are shown in Figure 11.

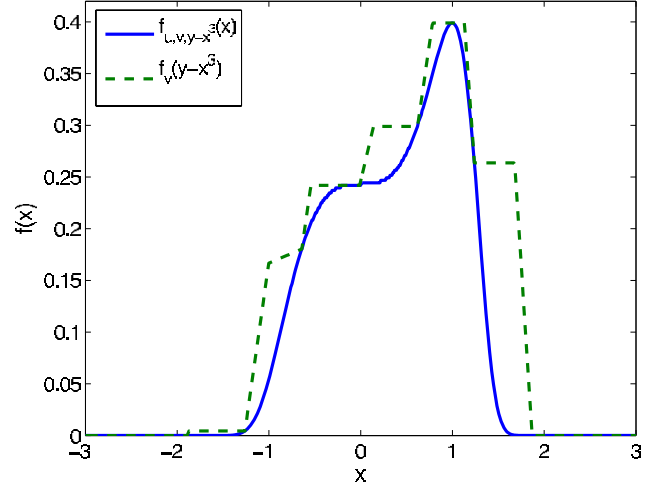


Figure 9: Unnormalized density  $f_v(\hat{y} - x^3)$  and the corresponding upper estimate  $f_{u,v,y-x^3}$  with 16 intervals.

The figure shows the system state after 20 iterations. It can be seen that the density functions are rather irregular. This is mainly due to the cubic measurement equation and the complicated noise density.

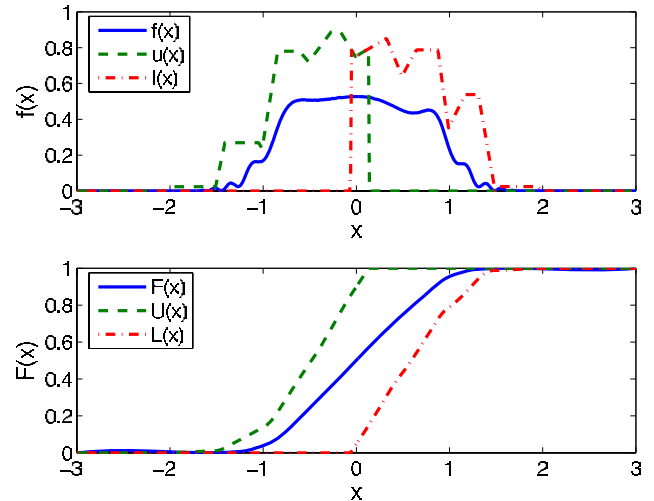


Figure 10: Estimated system state after the cubic measurement. Top: normalized density functions of the real system state and the upper and lower bounds using 20 intervals. Bottom: distribution functions.

## 7 Conclusion and Future Work

In this work lower and upper bounds for sets of probability densities have been introduced. For that purpose a partial ordering of densities has been defined. This allows to encase complex densities by simpler bounding densities. It was proposed to only propagate the bounds instead of the whole set through a modified Bayesian filter step which results in new bounding densities.

A kind of guaranteed estimator for complex densities can be designed to cope with numerical deviation

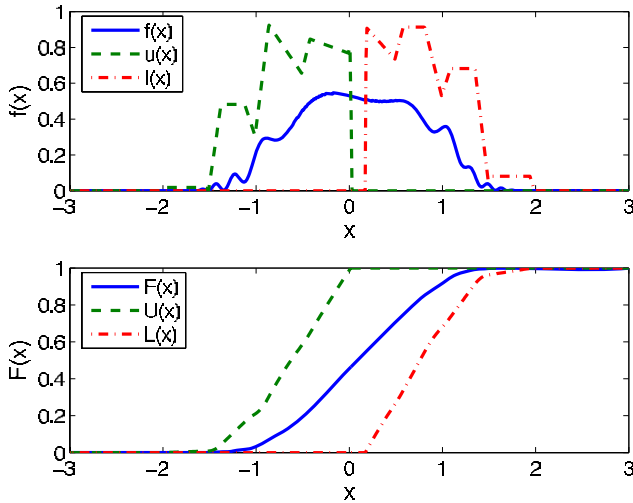


Figure 11: Estimated system state after 20 measurements with upper and lower bounds, 20 intervals. Top: density functions, Bottom: distribution functions.

and other deterministic error sources. By choosing the appropriate bounding density type, it can be ensured that the true density stays within determined bounds, while the complexity of the filter step does not increase. Furthermore, in technical systems the exact likelihood often cannot be obtained with adequate effort. Using bounds can be an efficient method to incorporate these kinds of mixed deterministic-stochastic uncertainties.

To demonstrate the practicability of our approach, lower and upper bounds for a class of density functions were constructed. It was shown that for continuous piecewise limited densities with piecewise limited derivatives, the modified Bayesian filter step can be derived. It was demonstrated on the cubic sensor problem, that these bounds are tight enough to conduct efficient nonlinear filtering.

Further research includes other classes of probabilities, e.g. higher order polynomials or piecewise Gaussians.

On the theoretical side, a derivation of a measure of the tightness of bounds would be helpful to identify new classes of boundable probability density functions.

In conclusion, we believe that the use of lower and upper bounds can substantially improve the quality of nonlinear estimators and simplify their design for a wide range of applications.

## References

- [1] J. O. Berger, D.R. Insula, *Bayesian Robustness in Robust Bayesian Analysis*, New York: Springer-Verlag, pp. 1–32, 2000.
- [2] L. R. Pericchi, *Sets of Prior Probabilities and Bayesian Robustness*, Imprecise Probabilities Project, <http://ippserv.rug.ac.be>, 2000.
- [3] L. Jaulin, M. Kieffer, O. Didrit, and É. Walter, *Applied Interval Analysis*, London: Springer-Verlag, 2001.
- [4] G. Shafer, *A Mathematical Theory of Evidence* Princeton University Press, 1976.
- [5] S. Ferson, V. Kreinovich, L. Ginzburg, D. S. Myers, and K. Sentz, *Constructing Probability Boxes and Dempster-Shafer Structures*, Sandia National Laboratories, 2003.
- [6] B. Betrò, A. Guglielmi, *Methods for Global Prior Robustness under Generalized Moment Conditions in Robust Bayesian Analysis*, New York: Springer-Verlag, pp. 273–293, 2000.
- [7] R. S. Bucy, *Bayes theorem and digital realizations of nonlinear filter*, J.A.A.S., Vol. 2, pp. 80–94 1969.
- [8] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, Second Edition, McGraw Hill Book Company, 1984.
- [9] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. *A Tutorial on Particle Filters for Online Non-linear/Non-Gaussian Bayesian Tracking*. IEEE Transactions of Signal Processing, Vol. 50, No. 2, pp. 174–188, 2002.
- [10] I. M. Rekleitis, *A Particle Filter Tutorial for Mobile Robot Localization*. Technical report, Centre for Intelligent Machines, McGill University, 2004.
- [11] D. L. Alspach, H. W. Sorenson, *Nonlinear Bayesian Estimation Using Gaussian Sum Approximation*. IEEE Transactions on Automatic Control, Vol. AC17, No. 4, pp. 439–448, 1972.
- [12] U. D. Hanebeck, O. Feiermann, *Progressive Bayesian Estimation for Nonlinear Discrete-Time Systems: The Filter Step for Scalar Measurements and Multidimensional States*, Proceedings of the 2003 IEEE Conference on Decision and Control (CDC'03), pp. 5366–5371, Maui, Hawaii, 2003.
- [13] D. Berleant, L. Xie, J. Zhang, *Statool: a tool for Distribution Envelope Determination (DEnv), an interval-based algorithm for arithmetic on random variables*. Reliable Computing 9 (2), pp. 91–108, 2003.
- [14] F. G. Cozman, C. P. de Campos, J. S. Ide, J. C. F. da Rocha, *Propositional and Relational Bayesian Networks Associated with Imprecise and Qualitative Probabilistic Assessments*. Conference on Uncertainty in Artificial Intelligence, pp. 104–111, 2004.
- [15] Joseph Y. Halpern, *Reasoning about Uncertainty*, The MIT Press, 2003.