A Recursive Formulation of Possibilistic Filters

Dominik Hose DOMINIK.HOSE@ITM.UNI-STUTTGART.DE
Michael Hanss MICHAEL.HANSS@ITM.UNI-STUTTGART.DE
Institute of Engineering and Computational Mechanics, University of Stuttgart, Germany

Abstract

We derive a recursive formulation of possibilistic filters that allow inference on the states in non-linear time-discrete dynamical systems in the presence of both aleatory and epistemic uncertainty with an imprecise probabilistic interpretation, and we present a particle-based implementation thereof.

Keywords: Imprecise Probabilities; Possibility Theory; Statistical Inference; Uncertainty Quantification; Systems Theory; State Filtering; Particle Filter

1. Introduction

Consider the problem of estimating the unknown states $\mathbf{x}(t_k) = \mathbf{x}_k \in \mathscr{X} \subseteq \mathbb{R}^{N_x}$ of a time-discrete dynamical system by observing its output $\mathbf{y}(t_k) = \mathbf{y}_k \in \mathscr{Y} \subseteq \mathbb{R}^{N_y}$ in the presence of both process noise $\mathbf{v}(t_k) = \mathbf{v}_k \in \mathscr{V} \subseteq \mathbb{R}^{N_v}$ and measurement error $\mathbf{w}(t_k) = \mathbf{w}_k \in \mathscr{W} \subseteq \mathbb{R}^{N_w}$ at times $t_0 \leq t_1 \leq \ldots \leq t_{N_t}$. The non-linear system dynamics are described by

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}, \mathbf{v}_{k-1}) \tag{1}$$

for $k = 1, \ldots, N_t$, with a given initial value $\mathbf{x}_0$. For simplicity, we do not assume any control input. The output equation is given by

$$\mathbf{y}_k = g(\mathbf{x}_k, \mathbf{w}_k). \tag{2}$$

for $k = 1, \ldots, N_t$. The process noise $\mathbf{v}_0, \mathbf{v}_1, \ldots$ constitutes realizations of the respective random variables $\widetilde{V}_0, \widetilde{V}_1, \ldots$, the measurement errors $\mathbf{w}_1, \mathbf{w}_2, \ldots$ are realizations of the random variables $\widetilde{W}_1, \widetilde{W}_2, \ldots$, and the initial value $\mathbf{x}_0$ is a realization of the random variable $\widetilde{X}_0$. These random variables are assumed to be stochastically independent with the respective identical probability distributions $\mathrm{P}_{\widetilde{V}}$, $\mathrm{P}_{\widetilde{W}}$, and $\mathrm{P}_{\widetilde{X}_0}$. The goal of a filter for such a system is to compute estimators of the current system state $\mathbf{x}_k$ or belief measures thereof which allow for a more thorough description of the uncertainty regarding its true value [31].

The algebraic structure of probability measures allows for simple and intuitive recursive formulations of Bayesian filters of such dynamical systems [29, 30]. However, Bayesian, i.e. additive, belief measures can exhibit unwanted behavior by assigning arbitrarily large degrees of belief to false hypotheses as shown, e.g., by the False Confidence Theorem [2]. Martin and Liu [27] argue that additive belief measures are unsuitable to guarantee their fundamental validity property and instead propose a larger class of inferential models. Here, we intend to promote inferential models based on possibility theory [25].

Apart from the general unsuitedness of additive belief structures, the problematic modeling of epistemic uncertainty by single probability measures shall also be avoided. In filtering problems, the process noise is usually an expression of missing insight into the actual system dynamics, rather than of actual random behavior of the system. In fact, the system may very well be entirely deterministic, and only imperfectly described, a property that is common to the majority of mathematical abstractions and models. By modeling this type of uncertainty with a precise probability distribution, the actual information is not accurately described and any interpretability of the corresponding belief measure, something engineers should seek to provide, is eliminated.

Consequently, in this contribution, we investigate the case when the probability distributions $\mathrm{P}_{\widetilde{V}}$, $\mathrm{P}_{\widetilde{W}}$, and $\mathrm{P}_{\widetilde{X}_0}$ are only imprecisely specified, i.e. when they are only known to belong to certain families of probability distributions. This also includes the case of entirely epistemic uncertainty, where, e.g., part of the deterministic model is only imprecisely specified but not random itself, as in the case of process noise. Possibility theory provides a convenient framework for computing with such imprecise probability distributions

A possibilistic batch filtering technique has already been described [21]. The aim of this contribution is to derive a similar recursive formulation of a filter using possibilistic descriptions of uncertainty for improved computational efficiency. The formulation is, in part, inspired by set-membership filtering techniques [24, 28]. The derived confidence distributions then yield a decidedly frequentist interpretation of belief.

2. Possibility Theory

In this section, we briefly discuss quantitative possibility theory, the branch of fuzzy set theory intended to reason with imprecise probabilities [12]. A detailed account is, e.g., provided by Hose and Hanss [20].

In the following, a tilde denotes a so-called $\mathscr{U}$-valued imprecise probabilistic random variable $\widetilde{U}$ defined on the measurable space $(\mathscr{U},\mathscr{B}(\mathscr{U}))$, where $\mathscr{U}\subseteq\mathbb{R}^N$ and $\mathscr{B}(\mathscr{U})$ is the corresponding Borel σ-field thereon.

A regular function $\phi:\mathscr{U}\to\mathscr{V}$ is lifted to a set-valued argument $U\subseteq\mathscr{U}$ in the standard way, i.e. $\phi(U)=\{\phi(u):u\in U\}$. Similarly, lifting the inverse of ϕ to a set-valued argument $V\subseteq\mathscr{V}$ is given by $\phi^{-1}(V)=\bigcup_{v\in V}\phi^{-1}(v)$, where $\phi^{-1}(v)=\{u\in\mathscr{U}:\phi(u)=v\}$.

2.1. Possibility Distributions

A possibility distribution of $\widetilde{U}$ is any (measurable) function $\pi_{\widetilde{U}}:\mathscr{U}\to[0,1]$ satisfying the normality condition

$$\sup_{\upsilon\in\mathscr{U}}\pi_{\widetilde{U}}(\upsilon)=1.$$

The resulting possibility measure

$$\Pi_{\widetilde{U}}(V)=\sup_{\upsilon\in V}\pi_{\widetilde{U}}(\upsilon)$$

for $V\in\mathscr{B}(\mathscr{U})$, where the supremum of the empty set is defined as zero, is complete because $\Pi_{\widetilde{U}}(\emptyset)=0$, normal because $\Pi_{\widetilde{U}}(\mathscr{U})=1$, and – most importantly – maxitive because

$$\Pi_{\widetilde{U}}\left(\bigcup_{k\geq1}U_k\right)=\sup_{k\geq1}\Pi_{\widetilde{U}}(U_k)$$

for a collection of (not necessarily disjoint) events $(U_k)_{k\geq1}$ with $U_k\in\mathscr{B}(\mathscr{U})$, making it a subadditive monotone plausibility measure (capacity). The dual necessity measure is defined for all $V\in\mathscr{B}(\mathscr{U})$ via the possibility of the corresponding counterevent $\neg V$ by

$$\mathrm{N}_{\widetilde{U}}(V)=1-\Pi_{\widetilde{U}}(\neg V)$$

and is a superadditive (minitive) monotone belief measure.

2.2. Imprecise Probabilities

The concept of consistency [7] is crucial for the imprecise probability notion in possibility theory. A possibility distribution is an imprecise predictive distribution of $\widetilde{U}$ if the true (additive) probability distribution $\mathrm{P}_{\widetilde{U}}$ of $\widetilde{U}$ is consistent with $\Pi_{\widetilde{U}}$, i.e. if the probability of every event $V\in\mathscr{B}(\mathscr{U})$ is bounded from above by $\mathrm{P}_{\widetilde{U}}(V)\leq\Pi_{\widetilde{U}}(V)$, and from below by $\mathrm{N}_{\widetilde{U}}(V)=1-\Pi_{\widetilde{U}}(\neg V)\leq1-\mathrm{P}_{\widetilde{U}}(\neg V)=\mathrm{P}_{\widetilde{U}}(V)$ making the possibility an upper probability and the necessity a lower probability [11]. We write $\mathrm{P}_{\widetilde{U}}\preceq\pi_{\widetilde{U}}$. Alternatively, the concept of consistency may be explained more illustratively via the level sets of $\pi_{\widetilde{U}}$. In particular, a probability and a possibility measure are consistent if and only if the probabilities of the sublevel sets $S^{\alpha}_{\pi_{\widetilde{U}}}=\left\{\upsilon\in\mathbb{R}:\pi_{\widetilde{U}}(\upsilon)\leq\alpha\right\}$ for all $\alpha\in[0,1]$ are bounded via

$$\mathrm{P}_{\widetilde{U}}\left(S^{\alpha}_{\pi_{\widetilde{U}}}\right)\leq\alpha$$

or, conversely, if the probability on the corresponding superlevel sets $C^{\alpha}_{\pi_{\widetilde{U}}}=\neg S^{\alpha}_{\pi_{\widetilde{U}}}$, the so-called α-cuts, is bounded by

$$\mathrm{P}_{\widetilde{U}}\left(C^{\alpha}_{\pi_{\widetilde{U}}}\right)=1-\mathrm{P}_{\widetilde{U}}\left(S^{\alpha}_{\pi_{\widetilde{U}}}\right)\geq1-\alpha.$$

Refer e.g. to Hose and Hanss [20] for a proof of this claim. In particular, this implies that the superlevel sets form prediction sets of $\widetilde{U}$ that are guaranteed to contain at least $1-\alpha$ probability mass [6]. We call a possibility distribution convex if these superlevel sets are convex.

In order to be able to express imprecise probabilistic knowledge about $\widetilde{U}$ in a possibilistic manner, an outer approximation of a set of probability distributions $\mathfrak{P}_{\widetilde{U}}$, which could, e.g., be given by a general credal set [1], a p-box [14], a precise probability distribution $\mathfrak{P}_{\widetilde{U}}=\{\mathrm{P}_{\widetilde{U}}\}$, or else, must be found – even though it is usually not possible to find an exact possibilistic description [10]. However, given an arbitrarily chosen candidate[1] possibility distribution $q_{\widetilde{U}}$ of $\widetilde{U}$, Hose and Hanss [20] argue that the possibility distribution $\pi_{\widetilde{U}}$ obtained by the Imprecise Probability-to-Possibility Transform

$$\pi_{\widetilde{U}}(u)=\sup_{\mathrm{P}_{\widetilde{U}}\in\mathfrak{P}_{\widetilde{U}}}\mathrm{P}_{\widetilde{U}}\left(\{\upsilon\in\mathscr{U}:q_{\widetilde{U}}(\upsilon)\leq q_{\widetilde{U}}(u)\}\right)\tag{3}$$

for $u\in\mathscr{U}$ is consistent with all $\mathrm{P}_{\widetilde{U}}\in\mathfrak{P}_{\widetilde{U}}$, i.e. it provides an outer approximation of $\mathfrak{P}_{\widetilde{U}}$. Most importantly, it is also plausibility-conform to $q_{\widetilde{U}}$, i.e. from $q_{\widetilde{U}}(u_1)\leq q_{\widetilde{U}}(u_2)$ it follows that $\pi_{\widetilde{U}}(u_1)\leq\pi_{\widetilde{U}}(u_2)$ for all $u_1,u_2\in\mathscr{U}$ (but not vice versa), and it is maximally specific[2]. That is, even though $\pi_{\widetilde{U}}$ does not generally exactly describe $\mathfrak{P}_{\widetilde{U}}$, one may not find a 'better' plausibility-conform possibility distribution that is also an outer approximation. Of course, much of the quality of the resulting $\pi_{\widetilde{U}}$ depends on the candidate possibility distribution $q_{\widetilde{U}}$, the obvious degree of freedom in this approach, but Hose and Hanss provide several reasonable options for how to choose it.

2.3. Statistical Inference

As discussed further below, the problem of filtering is connected to that of statistical inference where, in general, one seeks to infer information about the possible values of an unknown parameter $\vartheta\in\Theta$. It is assumed that this parameter is connected to an observable output, the $\mathscr{Z}$-valued imprecise probabilistic random variable $\widetilde{Z}$, an additional unknown nuisance parameter $\rho\in\mathscr{R}$ and a $\mathscr{U}$-valued imprecise probabilistic auxiliary random variable $\widetilde{U}$ with a known predictive distribution $\pi_{\widetilde{U}}$ via the implicit structural equation

$$0=\Psi(\vartheta,\rho,\widetilde{Z},\widetilde{U}).\tag{4}$$

1. Hose and Hanss call this a (subjective) plausibility function.
2. The concept of specificity is not discussed here. Refer, e.g., to Dubois and Prade [9] for further details.

Therein, all the variables, including 0, may be multivariate. The following lemma describes a confidence procedure to determine confidence sets of ϑ while marginalizing over ρ.

Lemma 1 (Possibilistic Inference Principle) *For all confidence levels $\alpha \in [0,1]$, the corresponding confidence set of $\vartheta \in \Theta$*

$$\mathscr{K}^{\alpha}(\widetilde{Z}) = \left\{ \theta \in \Theta : 0 \in \Psi\left(\theta, \mathscr{R}, \widetilde{Z}, C^{\alpha}_{\pi_{\widetilde{U}}}\right) \right\} \tag{5}$$

has the coverage probability $\mathrm{P}(\vartheta \in \mathscr{K}^{\alpha}(\widetilde{Z})) \geq 1-\alpha$.

Proof From $0 = \Psi(\vartheta, \rho, \widetilde{Z}, \widetilde{U})$, it follows that $\mathscr{K}^{\alpha}(\widetilde{Z}) = \{\theta \in \Theta : \Psi(\vartheta, \rho, \widetilde{Z}, \widetilde{U}) \in \Psi(\theta, \mathscr{R}, \widetilde{Z}, C^{\alpha}_{\pi_{\widetilde{U}}})\}$. The set membership criterion for $\vartheta \in \mathscr{K}^{\alpha}(\widetilde{Z})$, i.e. $\Psi(\vartheta, \rho, \widetilde{Z}, \widetilde{U}) \in \Psi(\vartheta, \mathscr{R}, \widetilde{Z}, C^{\alpha}_{\pi_{\widetilde{U}}})$, is trivially fulfilled if $\widetilde{U} \in C^{\alpha}_{\pi_{\widetilde{U}}}$, since, by assumption, we also know that $\rho \in \mathscr{R}$. Hence, the probability of the former condition being fulfilled is greater than the probability of the latter condition being fulfilled, yielding $\mathrm{P}(\vartheta \in \mathscr{K}^{\alpha}(\widetilde{Z})) \geq \mathrm{P}(\widetilde{U} \in C^{\alpha}_{\pi_{\widetilde{U}}}) \geq 1-\alpha$. ■

Having observed $\widetilde{Z} = z$, we can define a confidence distribution $\gamma_{\vartheta|z} : \Theta \to [0,1]$ whose values for $\theta \in \Theta$ are obtained by means of the Decomposition Theorem, exhibited e.g. in [15],

$$\gamma_{\vartheta|z}(\theta) = \sup\{\alpha \in [0,1] : \theta \in \mathscr{K}^{\alpha}(z)\},$$

which implies that the confidence sets $\mathscr{K}^{\alpha}(z)$ are simply the superlevel sets $C^{\alpha}_{\gamma_{\vartheta|z}}$ of $\gamma_{\vartheta|z}$ for all $\alpha \in [0,1]$. It is easily verified that this confidence distribution may also be computed directly from the predictive distribution of $\widetilde{U}$ via

$$\gamma_{\vartheta|z}(\theta) = \sup_{u \in \mathscr{U}, r \in \mathscr{R} : 0 = \Psi(\theta, r, z, u)} \pi_{\widetilde{U}}(u). \tag{6}$$

A very similar result can be derived independently in the framework of inferential models [25].

2.4. Joint Predictions

The formulations in Equations (5) and (6) are of limited use: If $\widetilde{U}$ is a multi-dimensional vector, only marginal predictive distributions $\pi_{\widetilde{U}_1}, \ldots, \pi_{\widetilde{U}_D}$ on their respective marginal domains $\mathscr{U}_1, \ldots, \mathscr{U}_D$ are typically available, whereas the Inference Principle requires knowledge of the joint distribution $\pi_{\widetilde{U}_1, \ldots, \widetilde{U}_D}$ on the joint domain $\mathscr{U}_1 \times \ldots \times \mathscr{U}_D$. This joint structure has to first be constructed from the marginal distributions.

This topic is closely connected to the different concepts of independence for imprecise probabilities, refer e.g. to [5]. Here, we restrict ourselves to the case of strong independence, i.e. we construct joint possibility distributions, such that all combinations of stochastically independent probability distributions from the respective marginal credal sets are consistent with the joint distribution. Furthermore, it suffices to restrict ourselves to the case of $D = 2$. That is, we seek a maximally specific[3] possibility distribution $\pi_{\widetilde{U}_1, \widetilde{U}_2}$ such that for all $\mathrm{P}_{\widetilde{U}_1} \preceq \pi_{\widetilde{U}_1}$ and $\mathrm{P}_{\widetilde{U}_2} \preceq \pi_{\widetilde{U}_2}$, the joint probability distribution under the assumption of stochastic independence of $\widetilde{U}_1$ and $\widetilde{U}_2$, which we write as $\mathrm{P}_{\widetilde{U}_1, \widetilde{U}_2} = \mathrm{P}_{\widetilde{U}_1} \cdot \mathrm{P}_{\widetilde{U}_2}$, also fulfills $\mathrm{P}_{\widetilde{U}_1, \widetilde{U}_2} \preceq \pi_{\widetilde{U}_1, \widetilde{U}_2}$.

To construct such joint distributions, several options are available [20]. We use the minimum-based rule proposed by Hose and Hanss [18, Theorem 28]. According to this rule, the joint possibility distribution of strongly independent imprecise probabilistic random variables is given by

$$\pi_{\widetilde{U}_1, \widetilde{U}_2}(u_1, u_2) = \min_{i=1,2} 1 - \left(1 - \pi_{\widetilde{U}_i}(u_i)\right)^2 \tag{7}$$

for all $u_1 \in \mathscr{U}_1$ and $u_2 \in \mathscr{U}_2$. This joint distribution is also applicable in the case of random set independence [8]. Moreover, if the α-cuts of $\pi_{\widetilde{U}_1}$ and $\pi_{\widetilde{U}_2}$ are convex to begin with, this rule guarantees the convexity of the α-cuts of $\pi_{\widetilde{U}_1, \widetilde{U}_2}$ as a Cartesian product of two convex sets, i.e.

$$C^{\alpha}_{\pi_{\widetilde{U}_1, \widetilde{U}_2}} = C^{1-\sqrt{1-\alpha}}_{\pi_{\widetilde{U}_1}} \times C^{1-\sqrt{1-\alpha}}_{\pi_{\widetilde{U}_2}}, \tag{8}$$

which is a computationally desirable feature.

Without proof, we furthermore argue that for a vacuous distribution $\pi_{\widetilde{U}_2}$, i.e. $\pi_{\widetilde{U}_2} \equiv 1$ everywhere on $\mathscr{U}_2$, a joint distribution is given by

$$\pi_{\widetilde{U}_1, \widetilde{U}_2}(u_1, u_2) = \pi_{\widetilde{U}_1}(u_1) \tag{9}$$

for all $u_1 \in \mathscr{U}_1$ and $u_2 \in \mathscr{U}_2$ which becomes apparent by reviewing the proof of Theorem 28 by Hose and Hanss [18]. Naturally, this rule also preserves convexity.

Of course, such methods for constructing joint possibility distributions are only recommended when no other information on the joint probability distribution is available in the first place. Generally, much expressiveness is lost by Equation (7), and, if one is able to specify a more expressive family of joint probability distributions $\mathfrak{P}_{\widetilde{U}_1, \widetilde{U}_2}$, e.g. from multivariate moments, explicit dependencies, etc., then one should convert this knowledge into a possibility distribution by means of the Imprecise Probability-to-Possibility Transform in Equation (3).

2.5. Joint Confidence

If the structural equation (4) admits a decomposition

$$0 = \Psi(\rho, \vartheta, \widetilde{Z}, \widetilde{U}) = \begin{pmatrix} \Psi_1(\rho, \vartheta, \widetilde{Z}_1, \widetilde{U}_1) \\ \Psi_2(\rho, \vartheta, \widetilde{Z}_2, \widetilde{U}_2) \end{pmatrix} \tag{10}$$

for the independent auxiliary variables $\widetilde{U}_1$ and $\widetilde{U}_2$ with the respective possibility distributions $\pi_{\widetilde{U}_1}$ and $\pi_{\widetilde{U}_2}$, then,

3. See footnote 2.

Equations (6) and (7) yield

$$\gamma_{\vartheta|z}(\theta) = \sup_{\substack{r\in\mathscr{R},u_1\in\mathscr{U}_1,u_2\in\mathscr{U}_2:\\ 0=\Psi_1(\theta,r,z_1,u_1)\\ 0=\Psi_2(\theta,r,z_2,u_2)}} \min_{i=1,2} 1-\left(1-\pi_{\widetilde{U}_i}(u_i)\right)^2$$

for $\theta \in \Theta$. It is readily verified that this is equivalent to first computing a confidence distribution of ϑ and ρ from both $\widetilde{Z}_1$ and $\widetilde{Z}_2$ independently, i.e. for $i = 1,2$

$$\gamma_{\rho,\vartheta|z_i}(r,\theta) = \sup_{u_i\in\mathscr{U}_i:0=\Psi_i(r,\theta,z_i,u_i)} \pi_{\widetilde{U}_i}(u_i), \tag{11}$$

combining them according to

$$\bar{\gamma}_{\rho,\vartheta|z_1,z_2}(r,\theta) = \min_{i=1,2} \gamma_{\rho,\vartheta|z_i}(r,\theta), \tag{12}$$

marginalizing over ρ via

$$\bar{\gamma}_{\vartheta|z_1,z_2}(\theta) = \sup_{r\in\mathscr{R}} \bar{\gamma}_{\rho,\vartheta|z_1,z_2}(r,\theta), \tag{13}$$

and finally (monotonously) re-scaling by

$$\gamma_{\vartheta|z_1,z_2}(\theta) = 1-\left(1-\bar{\gamma}_{\vartheta|z_1,z_2}(\theta)\right)^2, \tag{14}$$

by inserting Equations (11), (12), (13) and (14) into one another.[4] Due to the monotonicity of $\alpha \mapsto 1-(1-\alpha)^2$ on $[0,1]$, the re-scaling operation in Equation (14) may be performed last – or anywhere between the first three steps in Equations (11), (12) and (13). Additionally, the minimum over the (monotone functions of) $\pi_{\widetilde{U}_i}$ and the supremum over u_i have been swapped because the choice of the optimal u_1 does not depend on u_2, and vice versa. The constraints $0 = \Psi_i(\theta,r,z_i,u_i)$ for $i = 1,2$ are only coupled via the values of r and θ, and the optimal values of u_1 and u_2 are uniquely determined by them, which is expressed in Equations (11), (12) and (13). Equation (11) simply tells us how the optimal values of u_1 and u_2 are chosen for some combination (r,θ). But the final choice of u_1 does not constrain the value of u_2 and vice versa. Equation (12) finds the minimum of these two independent functions, and Equation (13) finds the corresponding optimal value of r that maximizes this minimum, implicitly fixing the optimal values of u_1 and u_2. Notice that it would not be permissible to also swap the supremum over $r \in \mathscr{R}$, the connection of the otherwise decoupled constraints Ψ_1 and Ψ_2, with the minimum.

Similar to Equation (7), Equation (12) also preserves the potential convexity of the independent confidence sets obtained in Equation (11).

The inference procedure described by the decomposition in Equation (10) and the subsequent evaluation of the steps in Equations (11), (12), (13) and (14) form the basis for the following recursive formulation of the possibilistic filter.

4. Considering that $\gamma_{\vartheta|z_i}(\theta)$ may be understood as a p-value for the hypothesis $\mathscr{H}_0 : \vartheta = \theta$, Equations (12), (13) and (14) correspond to a simple procedure for their combination under the assumption of independence [23].

3. Recursive Possibilistic Filtering

State filtering may be understood as a special case of statistical inference [31], which, most importantly, requires a sensible model of the involved uncertainties.

If a random variable, indeed, exhibits entirely aleatory behavior and its true probability distribution is known, then standard probability-to-possibility transformations are appropriate [13]. This is typically well-warranted for the measurement error $\widetilde{W}$.

For instance, if $\widetilde{W}$ follows a known multivariate normal probability distribution with mean $\mu \in \mathbb{R}^{N_w}$ and covariance matrix $\Sigma \in \mathbb{R}^{N_w\times N_w}$, then the optimal transform proposed by Dubois et al., a special case of the Imprecise Probability-to-Possibility Transform in Equation (3), yields the predictive distribution

$$\pi_{\widetilde{W}}(\omega) = 1 - F_{\chi^2_{N_w}}\left((\omega-\mu)^{\mathrm{T}}\Sigma^{-1}(\omega-\mu)\right) \tag{15}$$

for all $\omega \in \mathbb{R}^{N_w}$ where $F_{\chi^2_{N_w}}$ is the cumulative distribution function of the χ^2-distribution with N_w degrees of freedom. See Figure 1 for a visualization of the samples $\mathbf{w}_i$ from a multivariate normal distribution with

$$\mu = \begin{pmatrix}1\\1\end{pmatrix} \qquad \text{and} \qquad \Sigma = \begin{pmatrix}2 & 1\\1 & 1\end{pmatrix}$$

and the corresponding possibilities $\pi_{\widetilde{W}}(\mathbf{w}_i)$ indicated by the respective colors.

Should the true probability distribution be only imprecisely specified, then the more general Imprecise Probability-to-Possibility Transform must be employed. This could, e.g., be applied for the initial conditions where, due to the varying nature of the information about them, general recommendations may not be given.

Finally, it has already been said that the process noise is usually an expression of the lack of knowledge regarding the actual system dynamics. This uncertainty is regarded as reducible – one would expect it to vanish with a better understanding of the system – and does not typically exhibit stochastic behavior. In this extreme case of ignorance, we model the process noise by a vacuous predictive distribution $\pi_{\widetilde{V}} \equiv 1$ on the set of possible values $\mathscr{V}$ which has been shown to be appropriate in this case [20].

A batch filtering technique for the corresponding filtering problem as described in Section 1 has already been derived for linear time-invariant systems [21]. Here, a recursive formulation based on prediction, inversion and updating shall be pursued. The general feasibility of such schemes in a possibilistic setting has so far been considered by few scholars, e.g. from a perspective of possibilistic conditioning [4]. The remainder of this section contains an inductive construction of a simple recursive filter in Section 3.1, a formulation which is oriented towards general filter formulations, in particular set-membership filtering [24, 28], in

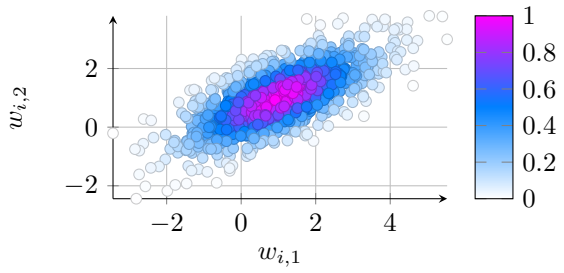


Figure 1: Multivariate normal distribution.

Section 3.2, and the description of a particle-based implementation of this filter in Section 3.3.

In the following, assume that we are at time t_k and wish to do inference on the current system state $\vartheta = \mathbf{x}_k$ given the output observations $\widetilde{Z} = (\mathbf{y}_1, \ldots, \mathbf{y}_k)$. The structural equation Ψ is composed of Equations (1) and (2) and the past system states $\rho = (\mathbf{x}_1, \ldots, \mathbf{x}_{k-1})$ are considered to be nuisance parameters. The auxiliary variable $\widetilde{U}$ is composed of all measurement errors $\widetilde{W}_1, \ldots, \widetilde{W}_k$ and the process noise $\widetilde{V}_0, \ldots, \widetilde{V}_{k-1}$, as well as the initial value $\widetilde{X}_0$, which are assumed to be unknown; but we have possibilistic information.[5] In particular, we assume that $\pi_{\widetilde{V}_0} = \pi_{\widetilde{V}_1} = \ldots$ are all vacuous on $\mathscr{V}$ and that the measurement errors are all described by the same predictive possibility distribution $\pi_{\widetilde{W}} = \pi_{\widetilde{W}_1} = \pi_{\widetilde{W}_2} = \ldots$

Readers who are only interested in the formulation of the possibilistic filter may skip the following section and continue reading in Section 3.2.

3.1. Inductive Derivation

Consider the following decomposition of the statistical inference problem described above. The output observations are decomposed into $\widetilde{Z}_1 = (\mathbf{y}_1, \ldots, \mathbf{y}_{k-1})$ and $\widetilde{Z}_2 = \mathbf{y}_k$, and the auxiliary variable is composed of the imprecise probabilistic random variables $\widetilde{U}_1 = (\widetilde{X}_0, \widetilde{V}_0, \ldots, \widetilde{V}_{k-2}, \widetilde{W}_1, \ldots, \widetilde{W}_{k-1})$, relevant to the past system states, and the imprecise probabilistic random variables $\widetilde{U}_2 = (\widetilde{V}_{k-1}, \widetilde{W}_k)$, relevant for the current system state. Similarly, the structural equation is decomposed into

$$0 = \Psi_1 = \left\{ \begin{array}{l} \mathbf{x}_i - f(\mathbf{x}_{i-1}, \widetilde{V}_{i-1}) \\ \mathbf{y}_i - g(\mathbf{x}_i, \widetilde{W}_i) \end{array} \right\} \text{ for } i = 1, \ldots, k-1$$

and

$$0 = \Psi_2 = \left\{ \begin{array}{l} \mathbf{x}_k - f(\mathbf{x}_{k-1}, \widetilde{V}_{k-1}) \\ \mathbf{y}_k - g(\mathbf{x}_k, \widetilde{W}_k). \end{array} \right.$$

Notice that Ψ_1 is not actually a function of $\vartheta = \mathbf{x}_k$ and must therefore yield a vacuous confidence distribution with respect to $\mathbf{x}_k$. That is, $\gamma_{\mathbf{x}_0,\ldots,\mathbf{x}_k|\mathbf{y}_1,\ldots,\mathbf{y}_{k-1}}(\xi_0, \ldots, \xi_k) = \gamma_{\mathbf{x}_0,\ldots,\mathbf{x}_k|\mathbf{y}_1,\ldots,\mathbf{y}_{k-1}}(\xi_0, \ldots, \xi_{k-1})$ for all $\xi_0, \ldots, \xi_k \in \mathscr{X}$. Similarly, Ψ_2 only depends on one component of ρ, namely $\mathbf{x}_{k-1}$, and must yield a vacuous confidence distribution with respect to $\mathbf{x}_0, \ldots, \mathbf{x}_{k-2}$.

This decomposition allows us to recursively infer a confidence distribution of $\mathbf{x}_k$ from a previously computed confidence distribution of $\mathbf{x}_{k-1}$, the current output $\mathbf{y}_k$ and the predictive distributions of $\widetilde{V}$ and $\widetilde{W}$. Below, we present a constructive proof of this claim by induction in k.

Initialization At time t_0, only the initial value $\mathbf{x}_0$ is relevant. In the absence of measurements, the predictive distribution of $\widetilde{X}_0$ becomes the confidence distribution $\gamma_{\mathbf{x}_0|\emptyset} = \pi_{\widetilde{X}_0}$. The key difference between the two distributions is that the predictive distribution is postulated before the filtering process has started. At t_0, the initial condition $\widetilde{X}_0 = \mathbf{x}_0$ has been realized but is still an unknown, and the confidence distribution provides the appropriate confidence sets and coverage probabilities thereof.

Recursion In the following, we assume that $\gamma_{\mathbf{x}_0,\ldots,\mathbf{x}_{k-1}|\mathbf{y}_1,\ldots,\mathbf{y}_{k-1}}$ is the confidence distribution obtained from $0 = \Psi$ in the previous iteration, and that it is vacuous with respect to the earlier system states $\mathbf{x}_0, \ldots, \mathbf{x}_{k-2}$. In particular, this assumption is fulfilled by $\gamma_{\mathbf{x}_0|\emptyset}$ going from $k = 0$ to $k = 1$.

Next, we define $\gamma_{\mathbf{x}_0,\ldots,\mathbf{x}_k|\mathbf{y}_1,\ldots,\mathbf{y}_{k-1}}$ as the vacuous extension of $\gamma_{\mathbf{x}_0,\ldots,\mathbf{x}_{k-1}|\mathbf{y}_1,\ldots,\mathbf{y}_{k-1}}$ with respect to $\mathbf{x}_k$, i.e.

$$\gamma_{\mathbf{x}_0,\ldots,\mathbf{x}_k|\mathbf{y}_1,\ldots,\mathbf{y}_{k-1}}(\xi_0, \ldots, \xi_k) = \gamma_{\mathbf{x}_{k-1}|\mathbf{y}_1,\ldots,\mathbf{y}_{k-1}}(\xi_{k-1}) \quad (16)$$

for all $\xi_0, \ldots, \xi_k \in \mathscr{X}$. This coincides with the successful evaluation of Equation (11) for $i = 1$ of the proposed decomposed inference procedure, and only leaves the steps described in Equation (11) for $i = 2$ and in Equations (12), (13) and (14) to be solved.

Considering $\widetilde{U}_2$, only the marginal possibility distributions of $\widetilde{V}_{k-1}$ and $\widetilde{W}_k$ are available and need to be aggregated. Since the process noise has a vacuous predictive distribution on $\mathscr{V}$, the joint distribution is given by $\pi_{\widetilde{V}_{k-1},\widetilde{W}_k}(\nu_{k-1}, \omega_k) = \pi_{\widetilde{W}}(\omega_k)$ for $\nu_{k-1} \in \mathscr{V}$ and $\omega_k \in \mathscr{W}$, and for $i = 2$, Equation (11) reduces to

$$\gamma_{\mathbf{x}_0,\ldots,\mathbf{x}_k|\mathbf{y}_k}(\xi_0, \ldots, \xi_k) = \sup_{\substack{\nu_{k-1} \in \mathscr{V}, \omega_k \in \mathscr{W}: \\ 0 = \xi_k - f(\xi_{k-1}, \nu_{k-1}) \\ 0 = \mathbf{y}_k - g(\xi_k, \mathbf{w}_k)}} \pi_{\widetilde{W}}(\omega_k) \quad (17)$$

for all $\xi_0, \ldots, \xi_k \in \mathscr{X}$. Since the right-hand side of this expression does not contain $\xi_0, \ldots, \xi_{k-2}$, it is clear that

$$\gamma_{\mathbf{x}_0,\ldots,\mathbf{x}_k|\mathbf{y}_k}(\xi_0, \ldots, \xi_k) = \gamma_{\mathbf{x}_{k-1},\mathbf{x}_k|\mathbf{y}_k}(\xi_{k-1}, \xi_k)$$

only describes a confidence distribution of $\mathbf{x}_{k-1}$ and $\mathbf{x}_k$, and is vacuous with respect to $\mathbf{x}_0, \ldots, \mathbf{x}_{k-2}$.

5. For improved results, $\widetilde{U}$ could also be a function of the auxiliary variables – similar to pivotal quantities.

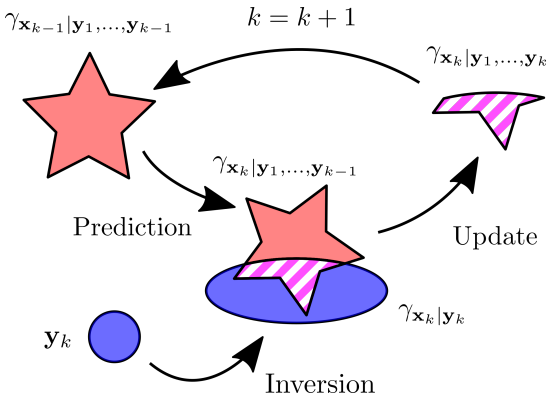


Figure 2: Sequence of the recursive filtering process.

The combination and marginalization steps described in Equations (12) and (13), furthermore, yield

$$\bar{\gamma}_{\mathbf{x}_k|\mathbf{y}_1,\ldots,\mathbf{y}_k}(\xi_k) = \sup_{\xi_{k-1}\in\mathscr{X}} \min\left(\gamma_{\mathbf{x}_{k-1}|\mathbf{y}_1,\ldots,\mathbf{y}_{k-1}}(\xi_{k-1}), \gamma_{\mathbf{x}_{k-1},\mathbf{x}_k|\mathbf{y}_k}(\xi_{k-1},\xi_k)\right) \tag{18}$$

for $\xi_k \in \mathscr{X}$, and, finally, this expression can be re-scaled according to Equation (14) in order to obtain the final confidence distribution

$$\gamma_{\mathbf{x}_k|\mathbf{y}_1,\ldots,\mathbf{y}_k}(\xi_k) = 1 - \left(1 - \bar{\gamma}_{\mathbf{x}_k|\mathbf{y}_1,\ldots,\mathbf{y}_k}(\xi_k)\right)^2. \tag{19}$$

In the next recursion step, this confidence distribution is used in place of $\gamma_{\mathbf{x}_0,\ldots,\mathbf{x}_{k+1}|\mathbf{y}_1,\ldots,\mathbf{y}_k}$, i.e. we define the vacuous extension

$$\gamma_{\mathbf{x}_0,\ldots,\mathbf{x}_{k+1}|\mathbf{y}_1,\ldots,\mathbf{y}_k}(\xi_0,\ldots,\xi_{k+1}) = \gamma_{\mathbf{x}_k|\mathbf{y}_1,\ldots,\mathbf{y}_k}(\xi_k) \tag{20}$$

for $\xi_0,\ldots,\xi_{k+1} \in \mathscr{X}$. This confidence distribution, again, fulfills the assumption in Equation (16) and enables us to repeat the described procedure for a new measurement $\mathbf{y}_{k+1}$.

An implementation of recursive filtering procedure as outlined above would be characterized by the repeated evaluation of Equations (17), (18) and (19). This description is useful for demonstrating the feasibility of recursive filters via induction; however, a much simpler formulation can be found, which is more in line with standard formulations, such as the Kalman filter or more advanced filtering techniques [31].

3.2. Recursive Formulation

In order to implement the recursion in a simpler fashion, we may describe it as the multi-step recursive process depicted in Figure 2 and detailed below. The general idea is to express all the available information in the confidence distributions with respect to only the current system state $\mathbf{x}_k$. In order to do this, $0 = \Psi$ is decomposed – more finely than in Section 3.1 – as visualized in Figure 3. Nodes I and II

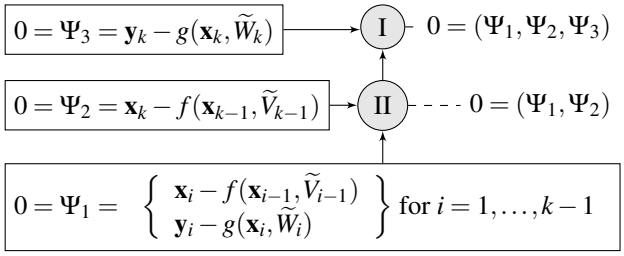


Figure 3: Two-fold decomposition of $0 = \Psi$.

indicate the two required decomposition steps, and inference is then carried out by subsequently applying the steps in Equations (11), (12), (13) and (14), first by considering the joint confidence obtained from $0 = \Psi_1$ and $0 = \Psi_2$; and then again from the joint confidence for $0 = (\Psi_1, \Psi_2)$, and $0 = \Psi_3$. The inductive proof of this implementation is similar to Section 3.1, but unnecessarily lengthy and technical, and shall not be discussed here.

Initialization The filter is initialized with $\gamma_{\mathbf{x}_0|\emptyset} = \pi_{\widetilde{X}_0}$ for the same reasons as in Section 3.1.

Prediction In the prediction step, the information about $\mathbf{x}_{k-1}$ obtained from the past measurements $\mathbf{y}_1,\ldots,\mathbf{y}_{k-1}$ is propagated according to the system dynamics in order to obtain a prediction of the confidence distribution of $\mathbf{x}_k$ taking the effect of the process noise $\widetilde{V}_{k-1}$, which potentially affects the predicted state, into account, not yet incorporating the information obtained from the next measurement $\mathbf{y}_k$. Computing the confidence distribution

$$\gamma_{\mathbf{x}_k|\mathbf{y}_1,\ldots,\mathbf{y}_{k-1}}(\xi_k) = \sup_{\substack{\xi_{k-1}\in\mathscr{X}_{k-1}, v_{k-1}\in\mathscr{V}: \\ \xi_k = f(\xi_{k-1}, v_{k-1})}} \gamma_{\mathbf{x}_{k-1}|\mathbf{y}_1,\ldots,\mathbf{y}_{k-1}}(\xi_{k-1}) \tag{21}$$

for $\xi_k \in \mathscr{X}$ provides this prediction of the new system state and is equivalent to

$$C^{\alpha}_{\gamma_{\mathbf{x}_k|\mathbf{y}_1,\ldots,\mathbf{y}_{k-1}}} = f\left(C^{\alpha}_{\gamma_{\mathbf{x}_{k-1}|\mathbf{y}_1,\ldots,\mathbf{y}_{k-1}}}, \mathscr{V}\right)$$

for all $\alpha \in [0,1]$. Both these expressions correspond to the well-known fuzzy extension principle. By treating confidence distributions as fuzzy membership functions, it may, therefore, be solved by fuzzy arithmetical methods [15].

Inversion Having expressed the past information about the current system state $\mathbf{x}_k$ by computing $\gamma_{\mathbf{x}_k|\mathbf{y}_1,\ldots,\mathbf{y}_{k-1}}$ in the prediction step, the inversion step is concerned with projecting the information obtained from the current measurement $\mathbf{y}_k$ onto the state space in order to derive a second confidence distribution of $\mathbf{x}_k$, namely

$$\gamma_{\mathbf{x}_k|\mathbf{y}_k}(\xi_k) = \sup_{\omega_k\in\mathscr{W}:\mathbf{y}_k = g(\xi_k,\omega_k)} \pi_{\widetilde{W}}(\omega_k)$$

for $\xi_k \in \mathscr{X}$. In the common case of additive noise $\mathbf{y}_k = g(\mathbf{x}_k, \mathbf{w}_k) = h(\mathbf{x}_k) + \mathbf{w}_k$, this yields the explicit formula

$$\gamma_{\mathbf{x}_k|\mathbf{y}_k}(\xi_k) = \pi_{\widetilde{W}}(\mathbf{y}_k - h(\xi_k)) \tag{22}$$

for $\xi_k \in \mathscr{X}$, or

$$C^{\alpha}_{\gamma_{\mathbf{x}_k|\mathbf{y}_k}} = h^{-1}\left(\mathbf{y}_k - C^{\alpha}_{\pi_{\widetilde{W}}}\right)$$

for $\alpha \in [0,1]$, which may readily be solved via inverse fuzzy arithmetic as discussed in [16, 17, 19, 26].

Update Finally, the two confidence distributions are combined in the updating step by means of the familiar minimum-based rule

$$\gamma_{\mathbf{x}_k|\mathbf{y}_1,\ldots,\mathbf{y}_k}(\xi_k) = \min\left(1 - \left(1 - \gamma_{\mathbf{x}_k|\mathbf{y}_1,\ldots,\mathbf{y}_{k-1}}(\xi_k)\right)^2, 1 - \left(1 - \gamma_{\mathbf{x}_k|\mathbf{y}_k}(\xi_k)\right)^2\right) \tag{23}$$

This is nothing more than the fuzzy set intersection, i.e. the intersection of the respective confidence sets and, subsequently, re-scaling their marginal confidence levels, i.e. for all $\alpha \in [0,1]$

$$C^{1-(1-\alpha)^2}_{\gamma_{\mathbf{x}_k|\mathbf{y}_1,\ldots,\mathbf{y}_k}} = C^{\alpha}_{\gamma_{\mathbf{x}_k|\mathbf{y}_1,\ldots,\mathbf{y}_{k-1}}} \cap C^{\alpha}_{\gamma_{\mathbf{x}_k|\mathbf{y}_k}}.$$

If any of these intersections happens to be empty, then a subnormal confidence distribution is obtained in Equation (23). Contrary to predictive possibility distributions, which then lose their coherence property and do not avoid sure loss anymore, this is not uncommon for confidence distributions [3]. Even though a universally accepted interpretation and treatment of empty confidence sets has not been agreed upon, this issue does not seem to yield any practical problems in the proposed scheme, and, for this reason, we do not propose any form of re-normalization.

3.3. Particle-Based Implementation

A great advantage of the possibilistic filtering approach presented here is that it enables an implementation of Equations (21), (22) and (23) based on particles, which is very computationally efficient. The fundamental idea (without rigorous verification) is the following.

Similar to Bayesian particle filters, the confidence distributions of $\mathbf{x}_k$ at every time step k can be represented by a set of N_p particles with position $\mathbf{x}_k^{(i)}$ and the corresponding membership

$$\mu_k^{(i)} \approx \gamma_{\mathbf{x}_k|\mathbf{y}_1,\ldots,\mathbf{y}_k}\left(\mathbf{x}_k^{(i)}\right) \tag{24}$$

for $i = 1,\ldots,N_p$. However, this particle distribution does not represent a probability density. Informally, the corresponding belief measure on the state space is not obtained via 'integrating' over the (weights of the) particles, but by interpolating from them. That is, the positions and memberships of the particles must allow one to reconstruct the underlying confidence distribution from the nodes provided by the particles.

Accordingly, the particles must be initialized from suitable nodes for the interpolation of $\pi_{\widetilde{X}_0} = \gamma_{\mathbf{x}_0|\emptyset}$. In the easiest case, where this distribution is a box-shaped quasi-vacuous distribution on some subset $\mathscr{X}_0 \subseteq \mathscr{X}$, i.e. $\pi_{\widetilde{X}_0}(\xi_0) = 1$ for $\xi_0 \in \mathscr{X}_0$ and zero otherwise, this may, e.g., be achieved via Latin hypercube sampling.

Subsequently, the prediction, inversion and updating step must be performed recursively on the particles and their memberships in every time step $k > 0$.

Similar to sampling-based fuzzy arithmetic [15], the prediction step corresponds to a simple sample propagation

$$\mathbf{x}_k^{(i)} = f\left(\mathbf{x}_{k-1}^{(i)}, \mathbf{v}_k^{(i)}\right). \tag{25}$$

Therein, the particles $\mathbf{v}_k^{(i)}$ constitute a particle-based representation of $\pi_{\widetilde{V}}$ with unit membership on $\mathscr{V}$, which is recommended to be re-drawn in every time step in order to avoid systematic errors. The past memberships of $\mathbf{x}_{k-1}^{(i)}$ are also the prediction memberships of $\mathbf{x}_k^{(i)}$ because by Equation (21) it follows that

$$\gamma_{\mathbf{x}_k|\mathbf{y}_1,\ldots,\mathbf{y}_{k-1}}\left(\mathbf{x}_k^{(i)}\right) \geq \gamma_{\mathbf{x}_k|\mathbf{y}_1,\ldots,\mathbf{y}_k}(\mathbf{x}_k^{(i)}) \approx \mu_{k-1}^{(i)},$$

i.e. $\mu_k^{(i),\text{pred.}} = \mu_{k-1}^{(i)}$ provides approximate lower bounds of the respective values of $\gamma_{\mathbf{x}_k|\mathbf{y}_1,\ldots,\mathbf{y}_{k-1}}$. Of course, the quality of this approximation depends critically on many other factors, such as the number and density of the particles, the sampling scheme for $\widetilde{V}$, the regularity of f, etc., and requires further investigation. Still, numerical experiments as discussed below suggest good convergence properties.

Subsequently, for the inversion step, a simple evaluation of Equation (22) at the particle location, i.e.

$$\mu_k^{(i),\text{inv.}} = \pi_{\widetilde{W}}\left(\mathbf{y}_k - h(\mathbf{x}_k^{(i)})\right) \tag{26}$$

provides an exact interpolation node of $\gamma_{\mathbf{x}_k|\mathbf{y}_k}$ at the respective particle position. If these positions are well distributed for interpolation purposes, no issues are to be expected.

Finally, the membership must be updated according to Equation (23), i.e. the prediction and inversion memberships are combined and then re-scaled via

$$\mu_k^{(i)} = 1 - \left(1 - \min\left(\mu_k^{(i),\text{pred.}}, \mu_k^{(i),\text{inv.}}\right)\right)^2 \tag{27}$$

in order to obtain a representation of $\gamma_{\mathbf{x}_k|\mathbf{y}_1,\ldots,\mathbf{y}_k}$. In particular, if the inversion memberships are very high due to good agreement with the present measurement (a high p-value), then the minimum of the prediction and the inversion membership is usually given by the former, and the membership of such particles increases because

$$\mu_k^{(i)} \geq \min\left(\mu_k^{(i),\text{pred.}}, \mu_k^{(i),\text{inv.}}\right) = \mu_k^{(i),\text{pred.}} = \mu_{k-1}^{(i)},$$

which follows from $1-(1-\min(\cdot))^2 \geq \min(\cdot)$ for $\min(\cdot) \in [0,1]$. Conversely, particles with poor measurement agreement indicated by low inversion memberships will exhibit decreasing overall memberships.

The particle distributions may, furthermore, be marginalized similar to Equation (13). For instance, in order to obtain a particle distribution of the d-th system state, one marginalizes over all other system states as 'nuisance parameters', which reduces to simply dropping all but the d-th coordinate of the particles.

Generally, the reconstruction of (the superlevel sets of) a confidence distribution from particle distributions is not trivial [32] and certainly requires further research. Here, we postulate that reasonable empirical approximations of the superlevel sets are given by the convex hull of all particles with a membership higher than α, i.e.

$$\hat{C}^{\alpha}_{\gamma_{\mathbf{x}_k|\mathbf{y}_1,\ldots,\mathbf{y}_k}} = \text{convhull}\left(\{\mathbf{x}_k^{(i)} : \mu_k^{(i)} > \alpha\}\right). \tag{28}$$

The appropriateness of the convex hull is directly linked to the continuity of f and the convexity-preserving properties of joint possibility and confidence distributions. By the intermediate value theorem, convex sets are mapped to connected sets.[6] And if the the initial distribution $\pi_{\widetilde{X}_0}$ is convex, and the inversion step in Equation (22) produces convex confidence sets, which depends only on the convexity of $\pi_{\widetilde{W}}$ in the case of additive measurement noise, then the superlevel sets of $\gamma_{\mathbf{x}_k|\mathbf{y}_1,\ldots,\mathbf{y}_k}$ are also convex. Under these very mild assumptions, the convex hull in Equation (28) is well-warranted. In order to check whether some system state $\xi_k \in \mathscr{X}$ belongs to this convex hull, it suffices to consider the feasibility problem of finding linear combination coefficients $\beta_i \in [0,1]$ with $\sum_{i:\mu_k^{(i)}>\alpha} \beta_i = 1$ and

$$\xi_k = \sum_{i:\mu_k^{(i)}>\alpha} \beta_i \mathbf{x}_k^{(i)}. \tag{29}$$

4. Simulation Study

We adopt the standard two-state batch reaction example studied by Leong and Nair [24] and many others. This problem is interesting because Bayesian belief measures may exhibit multimodal probability densities which often lead to unphysical state estimates [30]. In its time-discretized form, it is described by the system dynamics $\mathbf{x}_k = f(\mathbf{x}_{k-1}, \mathbf{v}_k) = \mathbf{x}_{k-1} + (1+\mathbf{v}_k)\delta_t r \mathbf{c}^{\mathrm{T}}$, where the reaction rate $r = d\mathbf{x}_{k,1}^2$ is a function of the first element of the system state $\mathbf{x}_k \in \mathbb{R}^2$ and the constant $d = 0.16$. Furthermore, $\delta_t = 0.1\,\mathrm{s}$ is the sampling time, and $\mathbf{c} = [-2,1]$ is a stoichiometric matrix describing the chemical reaction. The experiment is simulated for $N_t = 60$ time steps with the initial conditions $\mathbf{x}_0 = [3,1]^{\mathrm{T}}$, but we assume that our lack of knowledge about it is described by the quasi-vacuous distribution $\pi_{\widetilde{X}_0}$ on $\mathscr{X}_0 = [0,5]\times[0,5]$. Similarly, the process noise is described by the quasi-vacuous distribution $\pi_{\widetilde{V}}$ on $\mathscr{V} = [-0.1, 0.1]$ and may be interpreted both as an expression of the model uncertainty due to the explicit sampling scheme and allowing for a ten-percent deviation in the reaction ratio. To illustrate the capabilities of the filter, we perturb the system with an extreme case, the constant signal $\mathbf{v}_k^{\text{exp.}} = 0.1$ for all k, in the simulated experiment. Finally, it is assumed that the states are directly observable via the measurements $\mathbf{y}_k = \mathbf{x}_k + \mathbf{w}_k$, which are corrupted by additive noise $\mathbf{w}_k$ following a multivariate Gaussian probability distribution $\mathcal{N}(\mu, \Sigma)$ with

$$\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \qquad \text{and} \qquad \Sigma = \begin{pmatrix} 10^{-2} & 0 \\ 0 & 10^{-2} \end{pmatrix},$$

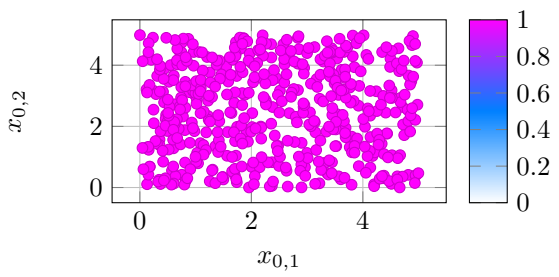
which is well described by the possibility distribution $\pi_{\widetilde{W}}$ given in Equation (15). Still, we sample the $\mathbf{w}_k^{\text{exp.}}$ from the original probability distribution $\mathcal{N}(\mu, \Sigma)$.

In Figure 4, some exemplary particle positions and memberships (indicated by colors) in the initialization, prediction, inversion and updating step of an implementation with a total of $N_p = 10^5$ particles at time steps $k = 0$ and $k = 1$ are depicted. The particle positions in the prediction step result from applying the system dynamics to the initial particles obtained from a Latin hypercube sampling on the initial box $\mathscr{X}_0$ in the initialization step. Since $\pi_{\widetilde{X}_0}$ is quasi-vacuous and the memberships are retained in the prediction step, the predicted particles constitute a representation of the quasi-vacuous distribution $\gamma_{\mathbf{x}_1|\emptyset}$ on $f(\mathscr{X}_0, \mathscr{V})$. In the inversion step, a second set of memberships for the same particle positions is computed according to Equation (22), and finally, both the prediction and the inversion memberships are combined. Clearly, because the prediction memberships are one for all particles, the inversion memberships are always lower, and consequently their values, which are re-scaled in Equation (23), determine the updated memberships.

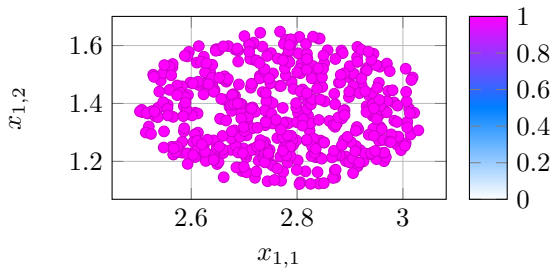
Of course, it is crucial to verify that the proposed implementation works as intended, i.e. that Lemma 1 holds, and that the empirical superlevel sets exhibit the guaranteed coverage probability. Consider, for instance, the marginal empirical α-cuts of the second system state visualized in Figure 5, which are inferred from the simulated experimental data (exp.) after the respective updating steps, and show the possible system trajectories along with the associated degree of confidence. At each time instant k, the a-priori probability that the true reference (ref.) is at a confidence level $\alpha \in [0,1]$ or higher, i.e. that it is in the corresponding empirical superlevel (the area with a darker blue than α), is $1-\alpha$. This claim is supported by the empirical coverage probabilities

$$\hat{p}(\alpha) = \frac{1}{N_t+1}\left|\left\{k = 0,\ldots,N_t : \mathbf{x}_k \in \hat{C}^{\alpha}_{\gamma_{\mathbf{x}_k|\mathbf{y}_1,\ldots,\mathbf{y}_k}}\right\}\right|$$

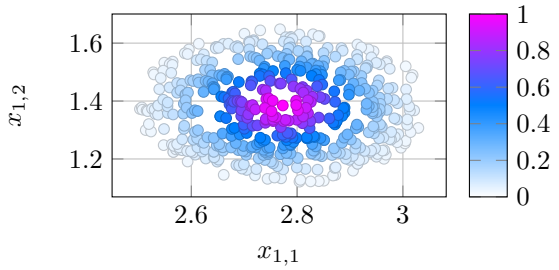
6. In many practical applications, the image of a convex set under a continuous function is actually convex, too. Nevertheless, the convex hull of a connected set is an outer approximation.



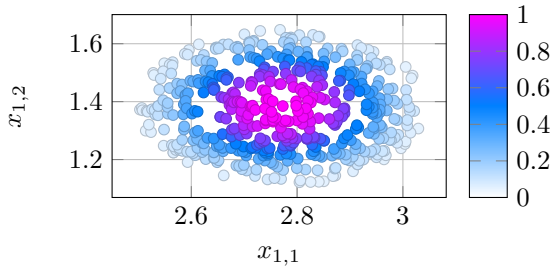
(*a*) Initialization step at $k = 0$.



(*b*) Prediction step at $k = 1$.



(*c*) Inversion step at $k = 1$.



(*d*) Updating step at $k = 1$.

Figure 4: Filtering steps.

in Figure 6, which are computed by checking which fraction of the true system states $\mathbf{x}_k$ belongs to the respective empirical α-cuts in Equation (28), i.e. by checking the corresponding feasibility in Equation (29). The additional 'slack' with respect to the dashed reference coverage probability also indicates a certain degree of conservatism in this approach, i.e. a lack of efficiency, which may be explained by the very general, and therefore conservative, methods employed for constructing joint possibility distributions.

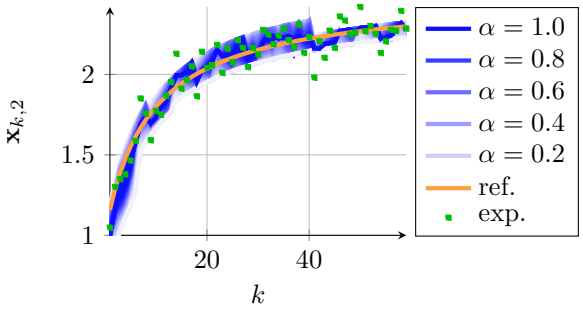


Figure 5: Time evolution of marginal empirical α-cuts of the second system state.

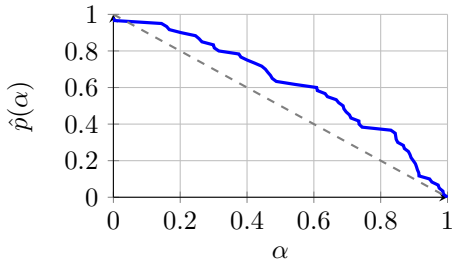


Figure 6: Empirical coverage probabilities.

5. Conclusion

As we intend to demonstrate a general proof of concept of a recursive formulation of possibilistic filters, we do not discuss issues related to observability and/or stability requirements of the filtered system, a topic which should not be dismissed in the future. If the filter is, furthermore, to be used in more challenging applications, both the implementation strategy – apart from the naive particle-based solution employed here, optimization-based implementations, as in [21], and interval-based implementations, which have not yet been investigated, seem to be promising directions of further research – and strategies for decision making based on the information provided by the filter, which could, e.g., be approached from the point of robust optimization as in [22], need to be addressed.

As a final remark, the Possibilistic Inference Principle, in particular Equation (6), seems to be a general version of the fuzzy extension principle. It is, therefore, of great interest to find more commonalities between confidence and predictive possibility distributions as both appear to adhere to the same possibilistic calculus, which has also been observed by Liu and Martin [25].

Acknowledgments

The authors gratefully acknowledge the support provided by the German Research Foundation (DFG) in the framework of the research project no. 319924547 (HA2798/9-2). They would also like to express their sincere appreciation of the reviewers' helpful comments and suggestions for improving the initial version of this manuscript.

References

[1] Alessandro Antonucci and Fabio Cuzzolin. Credal sets approximation by lower probabilities: application to credal networks. In *Proceedings of the 13th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU)*, pages 716–725. Springer, 2010.

[2] Michael S Balch, Ryan Martin, and Scott Ferson. Satellite conjunction analysis and the false confidence theorem. *Proceedings of the Royal Society A*, 475 (2227):20180565, 2019.

[3] Frank G Ball, Tom Britton, and Philip D O'Neill. Empty confidence sets for epidemics, branching processes and brownian motion. *Biometrika*, 89(1):211–224, 2002.

[4] Salem Benferhat, Didier Dubois, and Henri Prade. Kalman-like filtering and updating in a possibilistic setting. In *Proceedings of the 14th European Conference on Artificial Intelligence (ECAI)*, pages 8–12. IOS Press, 2000.

[5] Inés Couso, Serafín Moral, and Peter Walley. A survey of concepts of independence for imprecise probabilities. *Risk, Decision and Policy*, 5(2):165–181, 2000.

[6] Inés Couso, Susana Montes, and Pedro Gil. The necessity of the strong α-cuts of a fuzzy set. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(02):249–262, 2001.

[7] Miguel Delgado and Serafín Moral. On the concept of possibility-probability consistency. *Fuzzy Sets and Systems*, 21(3):311–318, 1987.

[8] Sébastien Destercke, Didier Dubois, and Eric Chojnacki. A consonant approximation of the product of independent consonant random sets. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 17(06):773–792, 2009.

[9] Didier Dubois and Henri Prade. The principle of minimum specificity as a basis for evidential reasoning. In *Proceedings of the 1st International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU)*, pages 75–84. Springer, 1986.

[10] Didier Dubois and Henri Prade. Consonant approximations of belief functions. *International Journal of Approximate Reasoning*, 4(5):419–449, 1990.

[11] Didier Dubois and Henri Prade. When upper probabilities are possibility measures. *Fuzzy Sets and Systems*, 49(1):65–74, 1992.

[12] Didier Dubois and Henri Prade. Possibility theory: qualitative and quantitative aspects. In *Quantified representation of uncertainty and imprecision*, pages 169–226. Springer, 1998.

[13] Didier Dubois, Laurent Foulloy, Gilles Mauris, and Henri Prade. Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. *Reliable Computing*, 10(4):273–297, 2004.

[14] Scott Ferson, Vladik Kreinovich, Lev Grinzburg, Davis Myers, and Kari Sentz. Constructing probability boxes and Dempster-Shafer structures. Technical report, Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), 2015.

[15] Michael Hanss. *Applied Fuzzy Arithmetic – An Introduction with Engineering Applications*. Springer, 2005.

[16] Dominik Hose and Michael Hanss. On inverse fuzzy arithmetical problems in uncertainty analysis. In *Proceedings of the 7th International Conference on Uncertainties in Structural Dynamics (USD)*, 2018.

[17] Dominik Hose and Michael Hanss. Consistent inverse probability and possibility propagation. In *Proceedings of the 2019 Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology (EUSFLAT)*. Atlantis Press, 2019.

[18] Dominik Hose and Michael Hanss. Possibilistic calculus as a conservative counterpart to probabilistic calculus. *Mechanical Systems and Signal Processing*, 133:106290, 2019.

[19] Dominik Hose and Michael Hanss. On data-based estimation of possibility distributions. *Fuzzy Sets and Systems*, 399:77–94, 2020.

[20] Dominik Hose and Michael Hanss. A universal approach to imprecise probabilities in possibility theory. *International Journal of Approximate Reasoning*, 133: 133–158, 2021.

[21] Dominik Hose and Michael Hanss. On the solution of forward and inverse problems in possibilistic uncertainty quantification for dynamical systems. In *Proceedings of the 9th International Workshop on Reliable Engineering Computing (REC) (accepted for publication, OPUS preprint available)*, 2021. doi: 10.18419/opus-10941.

[22] Dominik Hose, Markus Mäck, and Michael F Hanss. Robust optimization in possibility theory. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*, 5(4), 06 2019.

[23] James T Kost and Michael P McDermott. Combining dependent p-values. *Statistics & Probability Letters*, 60(2):183–190, 2002.

[24] Pei Hua Leong and Girish N Nair. Set-membership filtering using random samples. In *19th International Conference on Information Fusion (FUSION)*, pages 1087–1094. IEEE, 2016.

[25] Chuanhai Liu and Ryan Martin. Inferential models and possibility measures. *Researchers.One*, 20.08.00004v1, 2020.

[26] Weldon A Lodwick and Didier Dubois. Interval linear systems as a necessary step in fuzzy linear systems. *Fuzzy Sets and Systems*, 281:227–251, 2015.

[27] Ryan Martin and Chuanhai Liu. *Inferential Models: Reasoning with Uncertainty*. CRC Press, 2015.

[28] Mario Milanese and Carlo Novara. Unified set membership theory for identification, prediction and filtering of nonlinear systems. *Automatica*, 47(10):2141–2151, 2011.

[29] Simo Särkkä. *Bayesian Filtering and Smoothing*. Cambridge University Press, 2013.

[30] Xinguang Shao, Biao Huang, and Jong Min Lee. Constrained Bayesian state estimation – a comparative study and a new particle filter based approach. *Journal of Process Control*, 20(2):143–157, 2010.

[31] Arun K Tangirala. *Principles of System Identification: Theory and Practice*. CRC Press, 2014.

[32] Nico-Philipp Walz. *Fuzzy Arithmetical Methods for Possibilistic Uncertainty Analysis*. PhD thesis. Shaker, 2016.