Bounding Linearization Errors with Sets of Densities in Approximate Kalman Filtering

Benjamin Noack, Vesa Klumpp, Nikolay Petkov, and **Uwe D. Hanebeck**
Intelligent Sensor-Actuator-Systems Laboratory (ISAS),
Institute for Anthropomatics,
Karlsruhe Institute of Technology (KIT), Germany.
{noack,klumpp}@kit.edu, nikolay.petkov@student.kit.edu, uwe.hanebeck@ieee.org

Abstract — *Applying the Kalman filtering scheme to linearized system dynamics and observation models does in general not yield optimal state estimates. More precisely, inconsistent state estimates and covariance matrices are caused by neglected linearization errors. This paper introduces a concept for systematically predicting and updating bounds for the linearization errors within the Kalman filtering framework. To achieve this, an uncertain quantity is not characterized by a single probability density anymore, but rather by a set of densities and accordingly, the linear estimation framework is generalized in order to process sets of probability densities. By means of this generalization, the Kalman filter may then not only be applied to stochastic quantities, but also to unknown but bounded quantities. In order to improve the reliability of Kalman filtering results, the last-mentioned quantities are utilized to bound the typically neglected nonlinear parts of a linearized mapping.*

Keywords: Imprecise probabilities, sets of densities, credal sets, Bayesian state estimation, set-theoretic state estimation, linearization errors.

1 Introduction

Many practical applications entail the problem of inferring insight to unknown quantities that cannot directly be measured. Furthermore, the description of the underlying system dynamics and the obtained observations in general involve uncertainties, which are commonly modeled as random quantities. This means that an uncertain state variable is corrupted by stochastic noise with known probability density function. The generic Bayesian inference scheme then embodies an optimal solution to the state estimation problem. Unfortunately, this scheme represents more a theoretical solution than a practical one, as arbitrary densities generally do not allow a finite parameterization and nonlinear state estimation is computationally intractable. To overcome this issue, approximation techniques are employed to the densities or the models in order to be able to perform state estimation on-line. Of course, reliable state estimates then cannot be guaranteed anymore. However, for linear models perturbed by Gaussian noise, the Kalman filter [1] provides an optimal closed-form solution, where the state estimate can uniquely be parameterized by the conditional mean and covariance matrix. The simple processing of these parameters is the reason why the Kalman filter has also become a well-accepted approach for nonlinear systems. For that purpose, the process and measurement models are linearized either by a Taylor series expansion, as it is done for the extended Kalman filter [2], or by a linear regression analysis, for which the unscented Kalman filter [3] is a candidate. Especially, the former approach suffers from inconsistent estimates [4]. The regression-based filters generally calculate an additional linearization noise, in order to preserve consistency at the cost of less informative results [5]. Of course, consistency here strongly depends on properly chosen regression points.

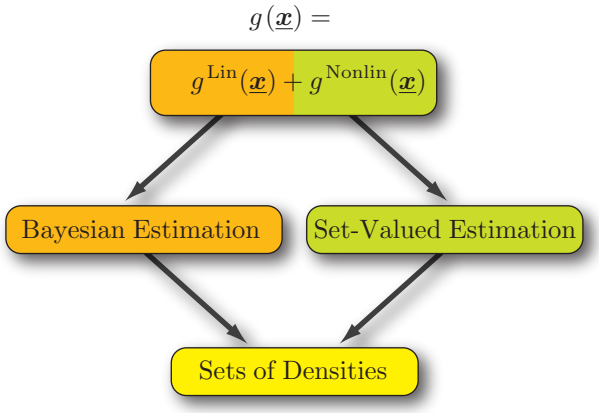

Figure 1: Kalman filter methods only consider linear parts of nonlinear system and observation models. The proposed method will incorporate the nonlinearities by means of set-valued descriptions. This is achieved by combining stochastic and set-valued quantities through sets of densities.

The focus of this paper is to offer a method for considering linearization errors systematically. This will

also enable us to assess the impact of these errors on the obtained state estimates. To achieve this, we will avail ourselves of concepts from set-membership state estimation. These approaches have evolved besides the Bayesian estimation techniques and provide guaranteed state estimates, i.e., sets to which the true state belongs. Essentially, intervals [6] and ellipsoidal sets, as used in [7], [8], and [9], are employed to model these unknown but bounded quantities.

As illustrated in Figure 1, the key idea of this paper consists in a combination of Bayesian and set-valued estimation concepts, where the former technique, i.e., the Kalman filter, is applied to the linear parts of a mapping and the latter concept is meant to account for the usually neglected nonlinear parts. In doing so, a state estimate will then be given as a combination of a stochastic and an unknown but bounded quantity. Such a state estimate can neither be described by a single probability function nor by a single set. To solve this issue, we make use of the theory of imprecise probabilities [10], from which generalizations of Bayesian state estimation have, for example, been derived in [11, 12]. For our purpose of bounding linearization errors within the Kalman filtering framework, sets of probability densities [13, 14] will serve as proper characterizations of uncertain quantities. We will start this paper with a thorough derivation of linear state estimation with sets of densities, which lays the ground for processing guaranteed bounds for linearization errors through the Kalman filter.

2 Linear State Estimation with Sets of Densities

Before we will explicate how to increase the reliability of Kalman filtering techniques when dealing with nonlinear system and measurement models, some preparatory work is required. This section is intended to lay a sound ground for combining stochastic and set-membership descriptions of uncertain quantities. Quite intuitively, this will lead us to imprecise probabilities. In the course of this generalization of classical probability theory, an uncertain quantity will not be characterized by a single probability density function anymore, but rather by a set of densities [12, 13]. Initially, this concept has been employed to allow for a simultaneous treatment of stochastic and systematic errors [14], whereas the latter ones are meant to subsume unknown but bounded disturbances. The extension of this concept to linearization errors then lies in the focus of Section 3.

Even though state estimation with sets of densities has been established for arbitrary densities and arbitrary models in [14], we here restrict our discussion to Gaussian densities and linear models. This means more precisely that we consider the Kalman filtering scheme and extend it to unknown but bounded uncertainties. Then, a quantity $\underline{\hat{x}}$ can additively be affected by both a zero-mean Gaussian random noise $\underline{\boldsymbol{w}} \sim \mathcal{N}(\underline{0}, \mathbf{C})$ with covariance matrix $\mathbf{C} \in \mathbb{R}^{n \times n}$ and an unknown but bounded perturbation $\underline{d} \in \mathcal{D} \subset \mathbb{R}^n$, i.e.,

$$\underline{\boldsymbol{x}} = \underline{\hat{x}} + \underline{\boldsymbol{w}} + \underline{d} \; .$$

Throughout this paper, vectors are indicated by underlining and random variables are denoted by boldface letters. In order to simplify further considerations, we assume that the statistics of $\underline{\boldsymbol{w}}$ do not depend on the outcomes of $\underline{d}$. For a moment, let $\underline{d}$ be fixed, then the uncertain quantity $\underline{\boldsymbol{x}}$ is normally distributed with $\mathcal{N}(\underline{\hat{x}}+\underline{d}, \mathbf{C})$. But, since $\underline{d}$ is unknown and only specified by its membership to the set $\mathcal{D}$, we cannot identify a certain distribution for $\underline{\boldsymbol{x}}$. Instead, we obtain a set

$$\{\mathcal{N}(\underline{\hat{x}} + \underline{d}, \mathbf{C}) \mid \underline{d} \in \mathcal{D}\} \tag{1}$$

of distributions, each of which is a seriously possible candidate for the true distribution of $\underline{\boldsymbol{x}}$. This justifies and even requires to consider the entire set as a characterization of $\underline{\boldsymbol{x}}$. For the following discussions in this paper, we refer to the set of Gaussian densities that corresponds to the set (1) of distributions.

Apparently, the vector $\underline{d}$ only affects the expected value, hence the set consists of translated Gaussian densities with the same covariance matrix. This implies that this set can be parameterized by the related set $\mathcal{X} = \{\underline{\hat{x}} + \underline{d} \mid \underline{d} \in \mathcal{D}\}$ of means and the covariance matrix $\mathbf{C}$. In order to generalize the Kalman filter to sets of means and to preserve computational tractability, ellipsoidal sets have proven to be particularly appropriate for this purpose. An ellipsoid

$$\mathcal{E}(\underline{\hat{c}}, \mathbf{X}) := \left\{\underline{x} \in \mathbb{R}^n \mid (\underline{x} - \underline{\hat{c}})^{\mathrm{T}} \mathbf{X}^{-1} (\underline{x} - \underline{\hat{c}}) \leq 1\right\} \tag{2}$$

is defined by a midpoint $\underline{\hat{c}} \in \mathbb{R}^n$ and a nonnegative definite shape matrix $\mathbf{X} \in \mathbb{R}^{n \times n}$. So, by using an ellipsoid $\mathcal{D} = \mathcal{E}(\underline{0}, \mathbf{X})$ to describe the unknown but bounded errors, the set of Gaussian densities is uniquely parameterized by the ellipsoid $\mathcal{X} = \mathcal{E}(\underline{\hat{x}}, \mathbf{X})$ of means and the covariance matrix $\mathbf{C}$. This provides us with an intuitive interpretation: The matrix $\mathbf{X}$ subsumes all unknown but bounded uncertainties and $\mathbf{C}$ encapsulates all stochastic uncertainties. We will see in Subsections 2.1 and 2.2 that these uncertainties remain distinguishable after propagating and filtering.

Ellipsoidal representations especially involve the advantage that affine transformations can easily be computed by

$$\mathbf{A}\mathcal{E}(\underline{\hat{c}}, \mathbf{X}) + \underline{b} = \mathcal{E}(\mathbf{A}\,\underline{\hat{c}} + \underline{b}, \mathbf{A}\mathbf{X}\mathbf{A}^{\mathrm{T}}) \tag{3}$$

for any $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\underline{b} \in \mathbb{R}^m$. For our further considerations, it will also be necessary to calculate the Minkowski sum of two ellipsoids, which means the elementwise summation. Unfortunately, this sum does in general not yield an ellipsoid anymore. However, it can

again easily be enclosed by an ellipsoid. Such an outer approximation is given by

$$\mathcal{E}(\hat{\underline{c}}_1, \mathbf{X}_1) \oplus \mathcal{E}(\hat{\underline{c}}_2, \mathbf{X}_2) \subseteq \mathcal{E}(\hat{\underline{c}}_1 + \hat{\underline{c}}_2, \mathbf{X}(p)) \ , \quad (4)$$

where the matrix $\mathbf{X}(p)$ is calculated from

$$\mathbf{X}(p) = (1 + p^{-1})\mathbf{X}_1 + (1 + p)\mathbf{X}_2$$

for any $p > 0$ [9]. This scalar can for example be set to

$$p = \mathrm{trace}(\mathbf{X}_1)^{\frac{1}{2}} \cdot \mathrm{trace}(\mathbf{X}_2)^{-\frac{1}{2}} \ . \quad (5)$$

Then, the enclosing ellipsoid has minimal sum of squares of semi-axes, i.e., the trace of $\mathbf{X}(p)$ is minimal. In [9], other optimality criteria for p can be found. With this brief survey of ellipsoidal calculus, we are now in a position to depict the Kalman prediction and filtering steps generalized to ellipsoidal sets of means.

2.1 Prediction

The Kalman prediction step requires linear system dynamics

$$\underline{x}_{k+1} = \mathbf{A}_k \, \underline{x}_k + \mathbf{B}_k \, \underline{u}_k \ , \quad (6)$$

which characterize the time evolution of the system state $\underline{x}_k$ given an input $\underline{u}_k$. In this paper, we confine ourselves to discrete-time models. In order to account for stochastic uncertainties, the standard Kalman filter now processes a normally distributed state estimate $\underline{\boldsymbol{x}}^{\mathrm{e}}_k \sim \mathcal{N}(\hat{\underline{x}}^{\mathrm{e}}_k, \mathbf{C}^{\mathrm{e}}_k)$ through this model (6), where the noise-corrupted signal $\underline{u}_k$ is composed of the known input $\hat{\underline{u}}_k$ and an additive zero-mean white Gaussian noise $\underline{\boldsymbol{w}}_k \sim \mathcal{N}(\underline{0}, \mathbf{C}^w_k)$. The mapping (6) is thereby applied to the means $\hat{\underline{x}}^{\mathrm{e}}_k$ and $\hat{\underline{u}}_k$, in order to obtain the predicted mean $\hat{\underline{x}}^{\mathrm{p}}_{k+1}$. The covariance matrix of the predicted state estimate $\underline{\boldsymbol{x}}^{\mathrm{p}}_{k+1}$ is computed from

$$\mathbf{C}^{\mathrm{p}}_{k+1} = \mathbf{A}_k \mathbf{C}^{\mathrm{e}}_k \mathbf{A}^{\mathrm{T}}_k + \mathbf{B}_k \mathbf{C}^w_k \mathbf{B}^{\mathrm{T}}_k \ . \quad (7)$$

The aspired extension towards unknown but bounded errors implies that the mean $\hat{\underline{x}}^{\mathrm{e}}_k$ of the state estimate $\underline{\boldsymbol{x}}^{\mathrm{e}}_k$ is not unique anymore, but rather is given by an ellipsoid $\mathcal{X}^{\mathrm{e}}_k = \mathcal{E}(\hat{\underline{c}}^{\mathrm{e}}_k, \mathbf{X}^{\mathrm{e}}_k)$ of "possible" means. Also, the input $\hat{\underline{u}}_k$ can now be additionally affected by an error term $\underline{d}_k \in \mathcal{D}_k = \mathcal{E}(\underline{0}, \mathbf{D}_k)$, i.e.,

$$\underline{u}_k = \hat{\underline{u}}_k + \underline{\boldsymbol{w}}_k + \underline{d}_k \ .$$

The impact of $\underline{d}_k$ on the input can be interpreted as a set

$$\mathcal{U}_k = \{\hat{\underline{u}}_k + \underline{d}_k \mid \underline{d}_k \in \mathcal{D}_k\} = \mathcal{E}(\hat{\underline{u}}, \mathbf{D}_k) \ ,$$

of possible inputs. The set of predicted means is then given by

$$\begin{aligned} \mathcal{X}^{\mathrm{p}}_{k+1} &\overset{\text{eq. (6)}}{=} \mathbf{A}_k \, \mathcal{X}^{\mathrm{e}}_k \oplus \mathbf{B}_k \, \mathcal{U}_k \\ &\overset{\text{eq. (3)}}{=} \mathcal{E}(\mathbf{A}_k \, \hat{\underline{c}}^{\mathrm{e}}_k, \mathbf{A}_k \mathbf{X}^{\mathrm{e}}_k \mathbf{A}^{\mathrm{T}}_k) \oplus \mathcal{E}(\mathbf{B}_k \, \hat{\underline{u}}_k, \mathbf{B}_k \mathbf{D}_k \mathbf{B}^{\mathrm{T}}_k) \\ &\overset{\text{eq. (4)}}{\subseteq} \mathcal{E}(\hat{\underline{c}}^{\mathrm{p}}_{k+1}, \mathbf{X}^{\mathrm{p}}_{k+1}) \ , \end{aligned} \quad (8)$$

where the latter ellipsoid is an outer approximation of $\mathcal{X}^{\mathrm{p}}_{k+1}$ with

$$\mathbf{X}^{\mathrm{p}}_{k+1} = (1 + p^{-1})\mathbf{A}_k \mathbf{X}^{\mathrm{e}}_k \mathbf{A}^{\mathrm{T}}_k + (1 + p)\mathbf{B}_k \mathbf{D}_k \mathbf{B}^{\mathrm{T}}_k \ .$$

Finally, the predicted state $\underline{\boldsymbol{x}}^{\mathrm{p}}_{k+1}$ is characterized by the set $\mathcal{E}(\hat{\underline{c}}^{\mathrm{p}}_{k+1}, \mathbf{X}^{\mathrm{p}}_{k+1})$ of means and the covariance matrix $\mathbf{C}^{\mathrm{p}}_{k+1}$, which together form a set of translated Gaussian densities.

2.2 Filtering

The prior or predicted state estimate, given by the set $\mathcal{E}(\hat{\underline{x}}^{\mathrm{p}}_k, \mathbf{X}^{\mathrm{p}}_k)$ of means and the covariance matrix $\mathbf{C}^{\mathrm{p}}_k$, can then be fused with a specific measurement $\hat{\underline{y}}_k$, which is related to the system state through a linear measurement model

$$\hat{\underline{y}}_k = \underline{y}_k + \underline{\boldsymbol{v}}_k + \underline{e}_k = \mathbf{H}_k \, \underline{x}_k + \underline{\boldsymbol{v}}_k + \underline{e}_k \ .$$

This measurement can be affected by a zero-mean white Gaussian noise $\underline{\boldsymbol{v}}_k \sim \mathcal{N}(\underline{0}, \mathbf{C}^v_k)$ and – in contrast to the standard Kalman filter assumptions – also by an unknown but bounded perturbation $\underline{e}_k \in \mathcal{E}_k$. Again, we characterize this error term by an ellipsoid $\mathcal{E}_k = \mathcal{E}(\underline{0}, \mathbf{E}_k)$ and regard the set

$$\mathcal{Y}_k = \{\hat{\underline{y}}_k - \underline{e}_k \mid \underline{e}_k \in \mathcal{E}_k\} = \mathcal{E}(\hat{\underline{y}}_k, \mathbf{E}_k) \ ,$$

as a set of "possible" observations. Following the train of thought of the previous subsection, the standard Kalman filter equation

$$\hat{\underline{x}}^{\mathrm{e}}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)\hat{\underline{x}}^{\mathrm{p}}_k + \mathbf{K}_k \, \hat{\underline{y}}_k \quad (9)$$

for the mean now generalizes to the Minkowski sum

$$\begin{aligned} \mathcal{X}^{\mathrm{e}}_k &\overset{\text{eq. (9)}}{=} (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \, \mathcal{E}(\hat{\underline{c}}^{\mathrm{p}}_k, \mathbf{X}^{\mathrm{p}}_k) \oplus \mathbf{K}_k \, \mathcal{E}(\hat{\underline{y}}_k, \mathbf{E}_k) \\ &\overset{\text{eq. (3)}}{=} \mathcal{E}\big((\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \, \hat{\underline{c}}^{\mathrm{p}}_k, (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)\mathbf{X}^{\mathrm{p}}_k(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)^{\mathrm{T}}\big) \\ &\qquad \oplus \mathcal{E}(\hat{\underline{y}}_k, \mathbf{K}_k \mathbf{E}_k \mathbf{K}^{\mathrm{T}}_k) \\ &\overset{\text{eq. (4)}}{\subseteq} \mathcal{E}(\hat{\underline{c}}^{\mathrm{e}}_k, \mathbf{X}^{\mathrm{e}}_k) \ , \end{aligned} \quad (10)$$

where $\mathbf{I}$ is the unit matrix and $\mathbf{K}_k$ is the Kalman gain

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

The shape matrix $\mathbf{X}^{\mathrm{e}}_k$ of the outer approximation (10) is computed from

$$\begin{aligned} \mathbf{X}^{\mathrm{e}}_k = \ & (1 + p^{-1})(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)\mathbf{X}^{\mathrm{p}}_k(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)^{\mathrm{T}} \\ & + (1 + p)\mathbf{K}_k \mathbf{E}_k \mathbf{K}^{\mathrm{T}}_k \ , \end{aligned}$$

where p can be determined by (5). The estimated covariance matrix is obtained from the common Kalman filter equation

$$\mathbf{C}^{\mathrm{e}}_k = \mathbf{C}^{\mathrm{p}}_k - \mathbf{K}_k \mathbf{H}_k \mathbf{C}^{\mathrm{p}}_k \ . \quad (11)$$

Apparently, introducing unknown but bounded errors into the Kalman filtering scheme has turned out to be quite straightforward. Here, only the equations for the predicted and filtered mean have to be generalized to Minkowski sums. At each time step, the state estimate is given by a set of translated Gaussian densities with the same covariance matrix. Contrary to nonlinear models and estimators, the unknown but bounded and the stochastic uncertainties remain distinguishable from each other: The former are represented by the matrix $\mathbf{X}_k$ and the latter are given by the covariance matrix $\mathbf{C}_k$. For a thorough derivation of this concept refer to [14].

3 Sets of Densities for Bounding Linearization Errors

Because they are in general easy to implement and require only low computational resources, Kalman filtering techniques are also widely applied to nonlinear system dynamics and observation models. In order to achieve this, linear approximations of the underlying nonlinearities have to be determined. One should then be prepared for inconsistent and non-optimal estimation results. Some Kalman filter derivatives attempt to overcome this issue by adding an extra linearization uncertainty [5], i.e., by enlarging the covariance matrix. However, the choice of this linearization noise in general involves some ambiguity and, of course, linearization errors cannot simply be described stochastically. Instead, we now derive a concept that provides us with systematic bounds on linearization errors.

3.1 The Idea Behind

The trouble with nonlinear system and measurement functions is that Gaussian random quantities will not be Gaussian anymore. E.g., quadratic measurement models can result in multi-modal densities. In general, the resulting density can be an arbitrary one providing no finite parameterization. So, the reason to revert the true density back to a Gaussian one is its simple parameterization and processing. This implies that the nonlinear part g^{Nonlin} in a mapping

$$\underline{\boldsymbol{z}} = g(\underline{\boldsymbol{x}}) = g^{\mathrm{Lin}}(\underline{\boldsymbol{x}}) + g^{\mathrm{Nonlin}}(\underline{\boldsymbol{x}}) \tag{12}$$

of an uncertain quantity $\underline{\boldsymbol{x}}$ is therefore omitted in order to preserve Gaussianity. This is done implicitly, whenever Kalman filter derivatives are deployed.

As illustrated earlier in Fig. 1, the key idea now consists in replacing $g^{\mathrm{Nonlin}}(\underline{\boldsymbol{x}})$ by a set that bounds the nonlinear parts properly. That means, the unknown but bounded part, i.e., the set, is intended to account for the typically neglected nonlinearities. Of course, linearization errors cannot, in general, be bounded over the whole domain. Hence, we consider these errors only over the most probable region of $\underline{\boldsymbol{x}}$, which means the most interesting "uncertainty region" around the state estimate $\hat{\underline{x}}$. More precisely, we calculate the error bounds over a confidence set in which the true state lies with a predefined probability level P.

Remark. *Essentially, we refer to the Bayesian viewpoint of confidence sets (also called credible sets or Bayesian confidence sets). The level P then states the probability with which the confidence set covers the true state.*

For an n-dimensional Gaussian random vector $\underline{\boldsymbol{x}} \sim \mathcal{N}(\hat{\underline{x}}, \mathbf{C})$, the confidence set to a certain probability level P is given by an ellipsoid

$$\left\{ \underline{x} \in \mathbb{R}^n \mid (\underline{x} - \hat{\underline{x}})^{\mathrm{T}} \mathbf{C}^{-1} (\underline{x} - \hat{\underline{x}}) \leq s \right\} .$$

The parameter s can be determined by means of the distribution of the squared Mahalanobis distance, which is that of a chi-squared variate with n degrees of freedom (see [7, pp. 524-525]). So, the scalar s depends on the chosen probability P and the dimensionality n. According to (2), the confidence ellipsoid can then be written as $\mathcal{E}(\hat{\underline{x}}, s \cdot \mathbf{C})$, which corresponds to certain sigma bounds of the error around $\hat{\underline{x}}$. Fig. 2(a) depicts a confidence set for a one-dimensional Gaussian density. Over this set, linearization errors can be bounded by a set

$$g^{\mathrm{Nonlin}}\big(\mathcal{E}(\hat{\underline{x}}, s \cdot \mathbf{C})\big) \subseteq \mathcal{L} .$$

This set $\mathcal{L}$ then acts as an unknown but bounded perturbation, so that (12) becomes

$$g(\underline{\boldsymbol{x}}) \approx g^{\mathrm{Lin}}(\underline{\boldsymbol{x}}) + \mathcal{L} ,$$

which implies that $g(\underline{\boldsymbol{x}})$ is now approximated by a set of translated Gaussian densities. Compliant with Section 2, we will employ ellipsoidal bounds $\mathcal{L} = \mathcal{E}(\underline{0}, \mathbf{X})$ for the linearization errors. With the probability P, the linearization error does not exceed the set $\mathcal{L}$. This implies that, to the same probability level P, the corresponding confidence set of the nonlinearly transformed state estimate (12) is then bounded by a set

$$g^{\mathrm{Lin}}(\mathcal{E}) \oplus g^{\mathrm{Nonlin}}(\mathcal{E}) \subseteq g^{\mathrm{Lin}}(\mathcal{E}) \oplus \mathcal{E}(\underline{0}, \mathbf{X})$$

with $\mathcal{E} = \mathcal{E}(\hat{\underline{x}}, s \cdot \mathbf{C})$. The latter confidence set is the Minkowski sum of the covariance ellipsoid $\mathcal{E}(\tilde{\underline{x}}, s \cdot \tilde{\mathbf{C}})$ and the error bound $\mathcal{E}(\underline{0}, \mathbf{X})$, where $\tilde{\underline{x}}$ and $\tilde{\mathbf{C}}$ are the new mean and covariance matrix after applying g^{Lin} to $\underline{\boldsymbol{x}}$. More precisely, we have obtained a set of Gaussian densities. Thereby, $\mathcal{E}(\tilde{\underline{x}}, \mathbf{X})$ is to be interpreted as a set of means and $s \cdot \tilde{\mathbf{C}}$ describes bounds for the stochastic uncertainties. To the chosen probability P, the matrix $\mathbf{X}$ characterizes the maximal impact of the neglected nonlinearities and $s \cdot \tilde{\mathbf{C}}$ the maximum stochastic error. For the one-dimensional case, Fig. 2(b) shows the overall set bounding the maximum error on the state estimate with the probability P.

In the following, we will show that the error bounds can be propagated through Kalman prediction and filtering steps and that still a certain probability level can

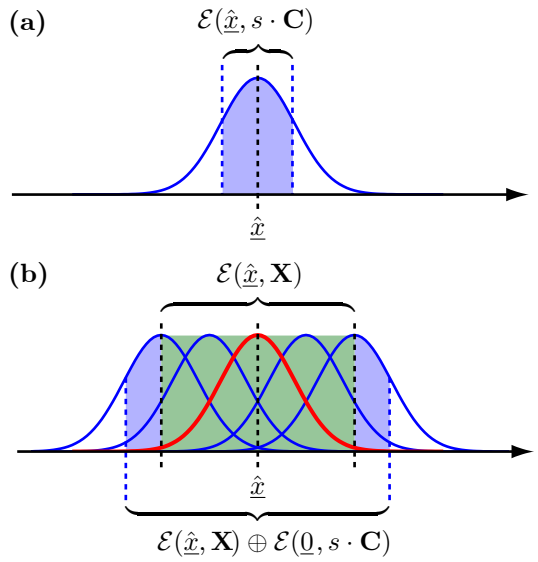

Figure 2: (a) The "uncertainty region" around the estimated mean $\hat{\underline{x}}_k$ is given by a multiple of the covariance ellipsoid. (b) "Uncertainty region" for a set of densities: For every element in the set of densities, a covariance ellipsoid has to be considered, the union of these, written as a Minkowski sum, gives an outer approximation of the true confidence set.

be guaranteed, to which the maximal linearization error lies within these bounds. In particular, only the bounding sets for the nonlinear parts need to be processed through the Minkowski sums (8) and (10). At each time step, the a priori defined probability level P is used to determine the region over which the linearization errors are to be bounded. For the prediction step, this concept will be detailed in Subsection 3.2 and the derivation of error bounds for the observation model will then be elucidated in Subsection 3.3.

3.2 Prediction

Given that linearization errors from previous processing steps are incorporated in the current state estimate $\underline{\boldsymbol{x}}_k^{\mathrm{e}}$ with respect to a certain probability level P, we aspire to maintain this level, when approximating the system function

$$\underline{\boldsymbol{x}}_{k+1}^{\mathrm{p}} = \underline{a}_k(\underline{\boldsymbol{x}}_k^{\mathrm{e}}, \underline{\boldsymbol{u}}_k)$$

linearly and predicting the system state. The current estimate $\underline{\boldsymbol{x}}_k^{\mathrm{e}}$ is hereby characterized by a set of means $\mathcal{E}(\hat{\underline{c}}_k^{\mathrm{e}}, \mathbf{X}_k^{\mathrm{e}})$, which accounts for the linearization errors made so far, and the covariance matrix $\mathbf{C}_k^{\mathrm{e}}$ determined through the linearized mappings. The "uncertainty region" of $\underline{\boldsymbol{x}}_k^{\mathrm{e}}$ then consists of the Minkowski sum

$$\mathcal{C}_k = \mathcal{E}(\hat{\underline{x}}_k^{\mathrm{e}}, \mathbf{X}_k^{\mathrm{e}}) \oplus \mathcal{E}(\underline{0}, s \cdot \mathbf{C}_k^{\mathrm{e}}) \ ,$$

where $\mathcal{E}(\underline{0}, s \cdot \mathbf{C}_k^{\mathrm{e}})$ is the P confidence region for the stochastic deviations. For the uncertain input

$$\underline{\boldsymbol{u}}_k = \hat{\underline{u}}_k + \underline{\boldsymbol{w}}_k \ ,$$

perturbations can be bounded with respect to the same probability level P by an ellipsoid $\mathcal{U}_k := \mathcal{E}(\hat{\underline{u}}_k, \tilde{s} \cdot \mathbf{C}_k^w)$. The linearization errors of an approximation

$$\underline{a}_k(\underline{x}_k, \underline{u}_k) \approx \mathbf{A}_k \underline{x}_k + \mathbf{B}_k \underline{u}_k + \underline{b}_k$$

over the considered regions yield a set

$$\left\{ \underline{a}_k(\underline{x}_k, \underline{u}_k) - (\mathbf{A}_k \underline{x}_k + \mathbf{B}_k \underline{u}_k + \underline{b}_k) \,\middle|\, \underline{x}_k \in \mathcal{C}_k, \underline{u}_k \in \mathcal{U}_k \right\}.$$

A bound for this set can, for example, be determined by taking the elementwise supremum

$$\underline{R}_k^u := \sup_{\substack{\underline{x}_k \in \mathcal{C}_k \\ \underline{u}_k \in \mathcal{U}_k}} \left[\underline{a}_k(\underline{x}_k, \underline{u}_k) - \left(\mathbf{A}_k \underline{x}_k + \mathbf{B}_k \underline{u}_k + \underline{b}_k \right) \right]$$

and infimum $\underline{R}_k^l$, respectively. This yields an n-dimensional rectangle, which itself can be interpreted as the Minkowski sum of n one-dimensional ellipsoids. So, an enclosing ellipsoid $\mathcal{E}(\hat{\underline{r}}_k^a, \mathbf{R}_k^a)$ is obtained through an outer approximation (4). A simpler but more conservative bound is a ball $\mathcal{E}(\underline{0}, r \cdot \mathbf{I})$ with radius

$$r = \sup_{\substack{\underline{x}_k \in \mathcal{C}_k \\ \underline{u}_k \in \mathcal{U}_k}} \left\| \underline{a}_k(\underline{x}_k, \underline{u}_k) - \left(\mathbf{A}_k \underline{x}_k + \mathbf{B}_k \underline{u}_k + \underline{b}_k \right) \right\|_2 \ ,$$

where the maximal possible error is used as bound in every direction.

The result of the Kalman prediction step now becomes the set

$$\begin{aligned} \mathcal{X}_{k+1}^{\mathrm{p}} &= \left(\mathbf{A}_k \mathcal{E}(\hat{\underline{x}}_k^{\mathrm{e}}, \mathbf{X}_k^{\mathrm{e}}) + \mathbf{B}_k \hat{\underline{u}}_k + \underline{b}_k \right) \oplus \mathcal{E}(\hat{\underline{r}}_k^a, \mathbf{R}_k^a) \\ &= \mathcal{E}(\hat{\underline{x}}_{k+1}^{\mathrm{p}}, \mathbf{X}_{k+1}^{\mathrm{p}}) \oplus \mathcal{E}(\hat{\underline{r}}_k^a, \mathbf{R}_k^a) \end{aligned}$$

of means containing the linearization errors and the covariance matrix $\mathbf{C}_{k+1}^{\mathrm{p}}$. $\mathcal{X}_{k+1}^{\mathrm{p}}$ is computed from (8), where we here, for the sake of simplicity, assumed that $\underline{\hat{u}}$ is not additionally affected by an unknown but bounded perturbation. $\mathbf{C}_{k+1}^{\mathrm{p}}$ is given by the Kalman filter equation (7).

3.3 Filtering

Similarly, linearization errors for the measurement mapping will be taken into account by the filtered state estimate $\underline{\boldsymbol{x}}_k^{\mathrm{e}}$. The set of estimated means is thereby given by the elementwise sum

$$\underline{x}_k^{\mathrm{e}} = \underline{x}_k^{\mathrm{p}} + \mathbf{K}_k \left[\underline{\hat{y}}_k - \left(\mathbf{H}_k \underline{x}_k^{\mathrm{p}} + \underline{b}_k + \underline{r}_k^h \right) \right]$$

for every $\underline{x}_k^{\mathrm{p}}$ in the set $\mathcal{E}(\hat{\underline{x}}_k^{\mathrm{p}}, \mathbf{X}_k^{\mathrm{p}})$ of predicted or prior means and for every $\underline{r}_k^h$ in $\mathcal{E}(\hat{\underline{r}}_k^h, \mathbf{R}_k^h)$, where the former set bounds the previously made linearization errors and the latter set characterizes the errors arising from the approximation

$$\underline{\hat{y}}_k = \underline{h}_k(\underline{\boldsymbol{x}}_k) + \underline{\boldsymbol{v}}_k \approx \mathbf{H}_k \underline{\boldsymbol{x}}_k + \underline{b}_k + \underline{\boldsymbol{v}}_k$$

of the nonlinear observation model $\underline{h}_k$. Both, $\mathcal{E}(\hat{\underline{x}}_k^{\mathrm{p}}, \mathbf{X}_k^{\mathrm{p}})$ and $\mathcal{E}(\hat{\underline{r}}_k^h, \mathbf{R}_k^h)$ refer to the same probability level P. In

order to determine $\mathcal{E}(\hat{\underline{r}}_k^h, \mathbf{R}_k^h)$, we consider the P confidence region of the measurement result $\hat{\underline{y}}_k$, i.e., $\mathcal{V}_k := \mathcal{E}(0, \tilde{s} \cdot \mathbf{C}_k^v)$. So, with probability P, there exists an $\underline{v}_k \in \mathcal{V}_k$ with

$$\hat{\underline{y}}_k = \underline{h}_k(\underline{x}_k) + \underline{v}_k \ .$$

This implies the inclusion

$$\hat{\underline{y}}_k - \underline{h}_k(\underline{x}_k) \in \mathcal{V}_k \ . \tag{13}$$

By $\mathcal{X}_k^h$, we denote the set of all $\underline{x}_k$ fulfilling this inclusion. The maximal linearization error can then, for example, be bounded by the elementwise supremum

$$\underline{R}_k^u := \sup_{\underline{x}_k^h \in \mathcal{X}_k^h} \left[\underline{h}_k(\underline{x}_k^h) - \left(\mathbf{H}_k \underline{x}_k^h + \underline{b}_k\right)\right]$$

and infimum $\underline{R}_k^l$, respectively. Again, this rectangle can be enclosed by an ellipsoid $\mathcal{E}(\hat{\underline{r}}_k^h, \mathbf{R}_k^h)$.

A way to determine the set $\mathcal{X}_k^h$ for the inclusion (13) is to define auxiliary linear equations, which are especially used in the context of set-membership state estimation [8, 9]. Such an auxiliary equation $\underline{y}_k = \mathbf{H}_k^a \underline{x}_k + \underline{b}^a$ fulfills

$$\left\{\underline{h}_k(\underline{x}_k) + \underline{v}_k \mid \underline{v}_k \in \mathcal{V}_k\right\} \subseteq \mathcal{E}(\mathbf{H}_k^a \underline{x}_k + \underline{b}^a, \mathbf{L})$$

for every $\underline{x}_k \in \mathbb{R}^n$ and a nonnegative definite $\mathbf{L} \in \mathbb{R}^{n \times n}$. Reversely, for a concrete measurement $\hat{\underline{y}}_k$, this gives the implication

$$\hat{\underline{y}}_k - \underline{h}_k(\underline{x}_k) \in \mathcal{V}_k \Longrightarrow \hat{\underline{y}}_k - (\mathbf{H}_k^a \, \underline{x}_k + \underline{b}^a) \in \mathcal{E}(0, \mathbf{L}_k) \ .$$

Thus, the inclusion

$$\begin{aligned} &\left\{\underline{x}_k \in \mathbb{R}^n \mid \hat{\underline{y}}_k - \underline{h}_k(\underline{x}_k) \in \mathcal{V}_k\right\} \\ &\subseteq \left\{\underline{x}_k \in \mathbb{R}^n \mid \hat{\underline{y}}_k - (\mathbf{H}_k^a \, \underline{x}_k + \underline{b}^a) \in \mathcal{E}(0, \mathbf{L}_k)\right\} =: \mathcal{X}_k^h \end{aligned}$$

holds and yields the set $\mathcal{X}_k^h$. The auxiliary function can also be used to compute a bound $\mathcal{E}(\hat{\underline{r}}_k^h, \mathbf{R}_k^h)$ for the maximum linearization error by considering the differences

$$\left[\left(\mathbf{H}_k^a \, \underline{x}_k + \underline{b}^a\right) - \left(\mathbf{H}_k \, \underline{x}_k + \underline{b}_k\right)\right]$$

over $\mathcal{X}_k^h$, instead of $\left[\underline{h}_k(\underline{x}_k) - \left(\mathbf{H}_k \, \underline{x}_k + \underline{b}_k\right)\right]$. With the determined ellipsoidal error bound, a new set of estimated states can then be calculated by

$$(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathcal{X}_k^{\mathrm{p}} + \mathbf{K}_k \mathcal{Y}_k$$

with $\mathcal{X}_k^{\mathrm{p}} = \mathcal{E}(\hat{\underline{x}}_k^{\mathrm{p}}, \mathbf{X}_k^{\mathrm{p}})$ and $\mathcal{Y}_k = \mathcal{E}\left(\hat{\underline{y}}_k - \underline{b}_k - \hat{\underline{r}}_k^h, \mathbf{R}_k\right)$. The covariance of the set of estimated Gaussian densities is still given by the standard equation (11). Because of the linearization, this matrix still can be underestimated, but the set of means, i.e., the bounds for linearization errors, enables us to compute a guaranteed confidence set with respect to the a priori defined probability level P.

4 A Brief Guide for Specific Kalman Filter Derivatives

This section provides a short survey on specific Kalman filter derivatives in order to unveil where linearization errors are hidden.

4.1 Extended Kalman Filter

The extended Kalman filter [2] appears to be the most apparent method for applying Kalman filtering techniques to nonlinear state estimation. For that purpose, differentiable system and measurement functions are approximated by first-order Taylor series expansions. Such an expansion of a nonlinear mapping $\underline{z} = \underline{g}_k(\underline{x})$ is evaluated at the current state estimate $\hat{\underline{x}}$, i.e.,

$$\begin{aligned} \underline{z} = \underline{g}(\underline{x}) = \underline{g}_k(\hat{\underline{x}}) + \underbrace{\left(\left.\frac{\partial \underline{g}}{\partial \underline{x}}\right|_{\underline{x}=\hat{\underline{x}}}\right)}_{=:\mathbf{A}} \Delta \underline{x} \\ + \left(\left.\frac{\partial^2 \underline{g}}{\partial^2 \underline{x}}\right|_{\underline{x}=\hat{\underline{x}}}\right) \frac{(\Delta \underline{x})^2}{2!} + \ldots \end{aligned}$$

with $\Delta \underline{x}_k = (\underline{x} - \hat{\underline{x}}_k)$. The linear approximation of $\underline{g}$ then yields

$$\underline{z} \approx \mathbf{A} \underline{x} + \underbrace{\underline{g}(\hat{\underline{x}}) - \mathbf{A}_k \hat{\underline{x}}}_{=:\underline{b}} \ .$$

As a result, a system or measurement mapping is approximated by its Jacobian matrix at $\hat{\underline{x}}_k^{\mathrm{e}}$ or $\hat{\underline{x}}_k^{\mathrm{p}}$, respectively, and all higher-order terms are neglected. The influence of the higher-order terms can be described, for example, by the Cauchy or Lagrange form of remainder, with which the nonlinear part can be bounded over the "uncertainty region" $\mathcal{X}$ around the state estimate. Such a bound can, for example, be calculated by means of Hessian matrices from

$$R_k = \sup_{\underline{x} \in \mathcal{X}} \frac{1}{2} T \left\| \underline{x} - \hat{\underline{x}} \right\|_\infty \ ,$$

where T is given by

$$T = \sum_{i,j=1}^{N} \sup_{\underline{x} \in \mathcal{X}} \left\| \frac{\partial^2 \underline{g}}{\partial \underline{x}_i \partial \underline{x}_j}(\underline{x}) \right\|$$

and $\underline{x}_i$ is the ith component of $\underline{x}$. With the concept derived in Section 3, the bounds for the remainder can now be taken into account and be processed through prediction and filtering steps.

4.2 Linear Regression Kalman Filter

A well-known example for linear regression Kalman filters [5] is the unscented Kalman filter [3]. These filters determine statistical linearizations of the nonlinear system and measurement functions by means of a

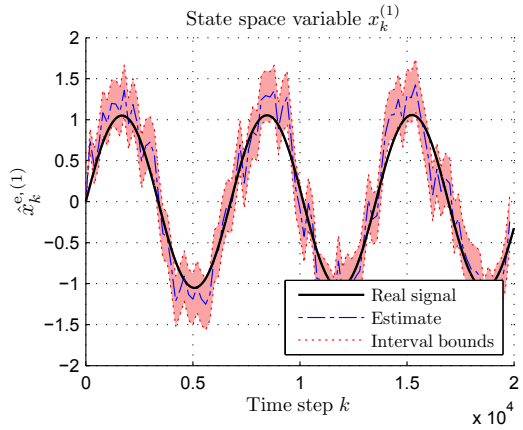

Figure 3: First component of $\hat{\underline{x}}_k^{\mathrm{e}}$. In blue, the center of the set of means is depicted. The bounds are drawn red.

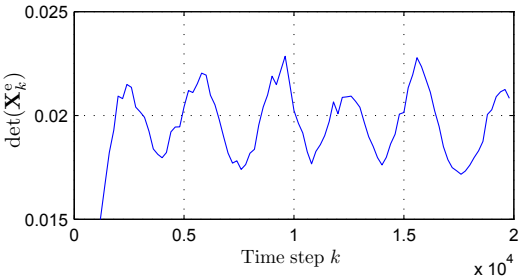

Figure 4: Determinant of $\mathbf{X}_k^{\mathrm{e}}$.

certain number L of regression points $\{\underline{x}_i\}$, which are usually taken around the mean of the actual state estimate, i.e., inside the "uncertainty region". For these points, the function values of the considered nonlinearity $\underline{z}_i = \underline{g}(\underline{x}_i)$ are calculated in order to obtain the set $\{\omega_i, \underline{x}_i, \underline{z}_i\}$ of regression points with weights ω_i. Essentially, mean and covariance matrix of the new state estimate are then computed from the regression points. The underlying linear mapping in general remains hidden to the user. As explicated in [5], this linear mapping is obtained as the solution of

$$\{\mathbf{A}, \underline{b}\} = \arg\min_{\mathbf{A},\underline{b}} \sum_{i=1}^{L} \omega_i \cdot \underline{e}_i^{\mathrm{T}} \cdot \underline{e}_i \ ,$$

where the weighted sum of squared errors $\underline{e}_i = \underline{y}_i - (\mathbf{A}\underline{x}_i + \underline{b})$ is minimized. This least-squares problem is solved by

$$\mathbf{A} = \mathbf{C}_{xz}^{\mathrm{T}} \mathbf{C}_{xx}^{-1} \ \text{ and } \ \underline{b} = \hat{\underline{z}} - \mathbf{A} \cdot \hat{\underline{x}}$$

with $\hat{\underline{x}} = \sum_{i=1}^{L} \omega_i \cdot \underline{x}_i$, $\hat{\underline{z}} = \sum_{i=1}^{L} \omega_i \cdot \underline{z}_i$, $\mathbf{C}_{xx} = \sum_{i=1}^{L} \omega_i \cdot (\underline{x}_i - \hat{\underline{x}}) \cdot (\underline{x}_i - \hat{\underline{x}})^{\mathrm{T}}$, and $\mathbf{C}_{xz} = \sum_{i=1}^{L} \omega_i \cdot (\underline{x}_i - \hat{\underline{x}}) \cdot (\underline{z}_i - \hat{\underline{z}})^{\mathrm{T}}$. Through the obtained linear mapping, the mean and the covariance matrix of the tranformed state estimate are evaluated. In order to achieve more consistent estimates, an additional zero-mean linearization noise is generally added [5], whose covariance matrix is

$$\mathbf{C}_k^* = \sum_{i=1}^{L} \omega_i \cdot \underline{e}_i \cdot \underline{e}_i^{\mathrm{T}} \ .$$

Instead of using this stochastical description of linearization errors, which only bases on a few sample points, one can employ systematic bounds: The results from Section 3 can be applied in order to attain guaranteed bounds for these errors.

5 Example

In this section, we will discuss the presented idea by means of a two-dimensional example. More precisely, we investigate the soft symmetric spring equation

$$\ddot{x} + x - \frac{1}{6}x^3 = 0 \ .$$

By time-discretization, we obtain the model

$$\underline{x}_{k+1} = \begin{bmatrix} x_{k+1}^{(1)} \\ x_{k+1}^{(2)} \end{bmatrix} = \Delta t \begin{bmatrix} x_k^{(2)} \\ \frac{1}{6}\left(x_k^{(1)}\right)^3 - x_k^{(1)} \end{bmatrix} + \begin{bmatrix} x_k^{(1)} \\ x_k^{(2)} \end{bmatrix}$$

of the process dynamics with step-size Δt. We assume that this discrete-time system is perturbed by a (zero-mean) white noise process $\underline{\boldsymbol{w}}_k \sim \mathcal{N}(0, \mathbf{C}_k^w)$ with covariance matrix $\mathbf{C}_k^w$. The measurements are modeled by

$$\underline{\boldsymbol{y}}_k = \underline{\boldsymbol{x}}_k + \underline{\boldsymbol{v}}_k$$

affected by the additive noise $\underline{\boldsymbol{v}}_k \sim \mathcal{N}(0, \mathbf{C}_k^v)$ with covariance matrix $\mathbf{C}_k^v = \left[\begin{smallmatrix} 0.1 & 0 \\ 0 & 0.1 \end{smallmatrix}\right]$. The extended Kalman filter then uses a linearized model

$$\underline{x}_{k+1} = f(\underline{x}_k, \Delta t) \approx \mathbf{A}_k \underline{x}_k + \underline{b}_k \ ,$$

where the matrix $\mathbf{A}_k$ is the Jacobian

$$\mathbf{A}_k := \mathbf{J}_f = \begin{bmatrix} 1 & \Delta t \\ \frac{1}{2}\Delta t\left(x_k^{(1)}\right)^2 - 1 & 1 \end{bmatrix}$$

of f. In the simulation, linearization errors have been bounded over 3σ confidence regions. The results are presented in Fig. 3, where only the first component of the state is shown. The interval bounds are the Minkoski sum (compare to Fig. 2(b)) of the set of means and the 3σ confidence region at each time step. The parameters of this simulation run are listed in Table 1. In Fig. 4, the determinant of the shape matrix $\mathbf{X}_k^{\mathrm{e}}$ of the ellipsoid of means is plotted, which is proportional to the volume of the ellipsoid. It can be seen that the bounds for the linearization errors become small, when the system behaves linearly, and increase, when nonlinear changes in the state are more dominant.

Property	Notation	Value
Discretization step size	Δt	0,001
Filter step size	Δt_m	0,2
Number of steps	–	20000
Initial estimate	$\hat{\underline{x}}_1^{\mathrm{e}}$	$[0\ 1]^T$
Initial covariance	$\mathbf{C}_1^{\mathrm{e}}$	diag[1 1]
Initial state ellipsoid	$\mathbf{X}_1^{\mathrm{e}}$	$\underline{0}$
Covariance matrix	$\mathbf{C}_k^w$	diag[0.05 0.05]

Table 1: Simulation parameters - soft spring equation.

6 Conclusions and Future Work

In order to apply the easy-to-implement and computationally feasible Kalman filter to arbitrary systems, the underlying process and observation models are linearized. Instead of neglecting the nonlinear parts, we have proposed a method, which enables us to bound these "inconvenient" parts by ellipsoidal sets. Since linearization errors cannot be considered over the entire domain, they are only enclosed over the most probable regions, which correspond to a certain probability level. This implies that linearization errors are interpreted as unknown but bounded perturbations, which can easily be incorporated into the Kalman filtering scheme as stated in Section 2. The resulting state estimate of this generalized Kalman filter is then given by a set of translated Gaussian densities, where the corresponding set of means accounts for the unknown but bounded uncertainties and the covariance matrix characterizes the linearly processed stochastic uncertainties. This Kalman filter for ellipsoidal sets provides the advantage of being easy to implement, since only one additional matrix calculation is required at each time step.

Of course, a set of densities cannot be regarded as a precise state estimate, since every element in the set is an equally acceptable candidate to model the stochastic uncertainty surrounding the true state. So, we do not obtain better estimation results, but we attain more reliable results and we gain insight into the robustness and sensitivity towards linearization errors, i.e., this concept allows for a robust Bayesian analysis and sensitivity analysis. The processed error bounds can assist in assessing the estimation quality on-line and figuring out when approximations break down.

Prospective research will in particular focus on discretization errors, when discrete-time models are generated from partial differential equations. Accordingly, for sampling-based approximations of densities, this paper may also provide a basis for incorporating discretization errors. Furthermore, more efficient ways to calculate the bounds in Subsection 3.2 and 3.3 will be evaluated. For example, the use of sparse grids seems promising, especially when the functions are of bounded variation, so that the effect of nonlinearities can be assessed by means of a few sample points.

7 Acknowledgments

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

References

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

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

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

[4] D. Bizup and D. Brown, "The Over-Extended Kalman Filter - Don't Use It!" in *Proceedings of the Sixth International Conference on Information Fusion. (Fusion 2003)*, 2003, pp. 40–46.

[5] T. Lefebvre, H. Bruyninckx, and J. D. Schutte, "Kalman Filters for Non-linear Systems: A Comparison of Performance," *International Journal of Control*, vol. 77, no. 7, pp. 639–653, May 2004.

[6] L. Jaulin, M. Kieffer, O. Didrit, and É. Walter, *Applied Interval Analysis: With Examples in Parameter and State Estimation, Robust Control and Robotics*. London: Springer, 2001.

[7] F. C. Schweppe, *Uncertain Dynamic Systems*. Prentice-Hall, 1973.

[8] F. L. Chernousko, *State Estimation for Dynamic Systems*. CRC Press, 1994.

[9] A. Kurzhanski and I. Vályi, *Ellipsoidal Calculus for Estimation and Control*. Birkhäuser, 1997.

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

[11] A. Benavoli, M. Zaffalon, and E. Miranda, "Reliable Hidden Markov Model Filtering through Coherent Lower Previsions," in *Proceedings of the 12th International Conference on Information Fusion (Fusion 2009)*, Jul. 2009, pp. 1743–1750.

[12] B. Noack, V. Klumpp, D. Brunn, and U. D. Hanebeck, "Nonlinear Bayesian Estimation with Convex Sets of Probability Densities," in *Proceedings of the 11th International Conference on Information Fusion (Fusion 2008)*, Cologne, Germany, Jul. 2008, pp. 1–8.

[13] D. R. Morrell and W. C. Stirling, "Set-Valued Filtering and Smoothing," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 21, no. 1, pp. 184–193, Jan. 1991.

[14] B. Noack, V. Klumpp, and U. D. Hanebeck, "State Estimation with Sets of Densities considering Stochastic and Systematic Errors," in *Proceedings of the 12th International Conference on Information Fusion (Fusion 2009)*, Seattle, Washington, July 2009.