

Adaptive Choice of Scaling Parameter in Derivative-Free Local Filters

Jindřich Duník, Miroslav Šimandl, and Ondřej Straka

Research Centre Data-Algorithms-Decision Making & Department of Cybernetics,
Faculty of Applied Sciences, University of West Bohemia,
Czech Republic

dunikj@kky.zcu.cz, simandl@kky.zcu.cz, straka30@kky.zcu.cz

Abstract – *The paper deals with adaptive choice of the scaling parameter in derivative-free local filters. In the last decade several novel local derivative-free filtering methods have been proposed. These methods exploiting Stirling’s interpolation and the unscented transformation are, however, conditioned by specification of a scaling parameter significantly influencing the quality of the state estimate. Surprisingly, almost no attention has been devoted to a suitable choice of the parameter. In fact, only a few basic recommendations have been provided, which are rather general and do not respect the particular system description. The choice of the parameter thus remains mainly on a user. The goal of the paper is to provide a technique for adaptive choice of the scaling parameter of the derivative-free local filters.*

Keywords: state estimation, nonlinear filtering, Kalman filtering

1 Introduction

The problem of recursive state estimation of discrete-time stochastic dynamic systems from noisy or incomplete measurement data has been a subject of considerable research interest for the last several decades.

The general solution to the estimation problem is given by the Bayesian recursive relations (BRR’s) for computation of probability density functions (pdf’s) of the state conditioned by the measurements [1]. These pdf’s provide a full description of immeasurable state. The closed form solution to the BRR’s is available only for a few special cases [1], e.g. for linear Gaussian system, which leads to the well-known Kalman filter (KF). In other cases, it is necessary to apply some approximate methods. These methods can be divided into two groups: local and global methods [2, 3].

The local methods are often based on an approximation of the nonlinear functions in the state or measurement equation so that the Kalman filter design technique can be used for the BRR’s solution. This approach leads to description of the conditional pdf’s of the state estimate by only the first two moments, i.e. mean value and covariance matrix. This rough approximation of the a posteriori estimates induces

local validity of the state estimates and consequently impossibility to generally ensure convergence of the local filter estimates. Thus, estimates generated by the local filters are suitable mainly for point estimates. On the other hand, the advantage of the local methods can be found in relative simplicity of the BRR’s solution and acceptable computational demands.

The global methods are based on a certain type of approximation of the conditional pdf of the state estimate to accomplish better state estimates. Unfortunately, they have higher computational demands than the local methods [3]. The global methods are not considered in this paper.

The standard local methods approximate nonlinear functions in the state or the measurement equation by the Taylor expansion up to the first or second order. The BRR’s solution based on the this approximation leads to e.g. the extended Kalman filter or the second order filter [1].

In the last decade, the novel approaches to local filter design based on the polynomial interpolation [4, 5, 6, 7, 3] or on the unscented transformation [8, 6, 9, 10, 11, 7, 3, 12], have been published. The approximation of the nonlinear functions by means of Stirling’s polynomial interpolation leads to the divided difference filters (DDF’s). Further, the DDF’s can be classified into the divided difference filter first order (DD1) and the divided difference filter second order (DD2) which are based on the first and second order Stirling’s interpolation formula, respectively [4]. Instead of the direct substitution of the nonlinear functions in the system description, an approximation of the “already approximated” pdf’s representing state estimates by a set of deterministically chosen weighted points (so called σ -points) can be utilised as a base for the local filters. This transformation is often called the unscented transformation. The unscented Kalman filter (UKF) [8], the Gauss-Hermite filter [6] or the cubature Kalman Filter [12] exemplify this approach. It is very important to mention that for the UKF and the DDF’s (and their variants) common features can be found although they come from quite different basic ideas [4, 13, 3]. Therefore, these local filters are often referred together as the sigma point Kalman filters or the derivative-

free Kalman (local) filters.

Almost all derivative-free filters are characteristic by that their design is conditioned by specification of a scaling parameter significantly influencing the quality of approximation and thus influencing the quality of the state estimates. Although, there exist some basic rules and recommendations on choice of the scaling parameter, the resulting choice of the parameter is independent of the system description or at best they depend on the system order only. The parameter is thus chosen prior to the estimation experiment and it remains constant during the whole experiment.

The goal of the paper is to propose a novel adaptive technique for determination of the scaling parameter for derivative-free local filters.

The paper is organised as follows. System specification and state estimation problem are briefly introduced in Section 2. Description of the derivative-free local filters together with a motivational example and the problem statement is given in Section 3. Section 4 is devoted to the proposal of the novel adaptive technique for scaling parameter determination. An algorithm of the derivative-free local filters with the adaptive technique for scaling parameter determination is summarised in Section 5. Numerical illustrations of the proposed algorithm are provided in Section 6. Concluding remarks are drawn in Section 7.

2 System Specification and State Estimation

Let the discrete-time linear stochastic system with nonlinear measurement equation be considered

$$\mathbf{x}_{k+1} = \mathbf{F}_k \mathbf{x}_k + \mathbf{w}_k, k = 0, 1, 2, \dots, \quad (1)$$

$$\mathbf{z}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{v}_k, k = 0, 1, 2, \dots, \quad (2)$$

where the vectors $\mathbf{x}_k \in \mathbb{R}^{n_x}$ and $\mathbf{z}_k \in \mathbb{R}^{n_z}$ represent the immeasurable state of the system and measurement at time instant k , respectively, $\mathbf{F}_k \in \mathbb{R}^{n_x \times n_x}$ is a known matrix, $\mathbf{h}_k : \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_z}$ is a known vector function, and $\mathbf{w}_k \in \mathbb{R}^{n_x}$, $\mathbf{v}_k \in \mathbb{R}^{n_z}$ are the state and measurement white noises. The pdf's of the noises are supposed to be Gaussian with zero means and known covariance matrices \mathbf{Q}_k and \mathbf{R}_k , i.e. $p_{\mathbf{w}_k}(\mathbf{w}_k) = \mathcal{N}\{\mathbf{w}_k : \mathbf{0}, \mathbf{Q}_k\}$ and $p_{\mathbf{v}_k}(\mathbf{v}_k) = \mathcal{N}\{\mathbf{v}_k : \mathbf{0}, \mathbf{R}_k\}$, respectively. The pdf of the initial state is supposed to be Gaussian and known as well, i.e. $p_{\mathbf{x}_0}(\mathbf{x}_0) = \mathcal{N}\{\mathbf{x}_0 : \bar{\mathbf{x}}_0, \mathbf{P}_0\}$, and independent of the noises.

Note that such model is often used in e.g. tracking applications or parameter estimation in neural networks.

The aim of the state estimation is to find the state estimate in the form of the conditional pdf $p(\mathbf{x}_k | \mathbf{z}^k)$ in which $\mathbf{z}^k = [\mathbf{z}_0, \mathbf{z}_1, \dots, \mathbf{z}_k]$. In some cases it is sufficient to find the first two conditional moments, i.e. the mean $\hat{\mathbf{x}}_{k|k} = E[\mathbf{x}_k | \mathbf{z}^k]$ and the covariance matrix $\mathbf{P}_{k|k} = \text{cov}[\mathbf{x}_k | \mathbf{z}^k]$, which can be understood as a Gaussian approximation of the conditional pdf, i.e. $p(\mathbf{x}_k | \mathbf{z}^k) \approx \mathcal{N}\{\mathbf{x}_k : \hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}\}$ [12].

3 Derivative-Free Local Filters and Problem Statement

In this section the basic ideas, algorithms, and properties of the derivative-free local filters are presented.

Generally, there are two main classes of derivative-free local estimators; the first class is based on the polynomial interpolation and the second one is based on the unscented transformation. These approximation techniques can be illustrated by an example of transformation of a random variable through a nonlinear function [4, 8, 3].

3.1 Transformation of random variable

Let $\mathbf{x} \in \mathbb{R}^{n_x}$ and $\mathbf{y} \in \mathbb{R}^{n_y}$ be random vector variables related through the known nonlinear function

$$\mathbf{y} = \mathbf{g}(\mathbf{x}) = [g_1(\mathbf{x}), \dots, g_{n_y}(\mathbf{x})]^T. \quad (3)$$

The variable \mathbf{x} is given by the first two moments, i.e. the mean $\hat{\mathbf{x}}$ and the covariance matrix \mathbf{P}_x . The aim is to calculate the mean and the covariance matrix of \mathbf{y} , and the cross-covariance matrix \mathbf{P}_{xy} , i.e.

$$\hat{\mathbf{y}} = E[\mathbf{y}] = E[\mathbf{g}(\mathbf{x})], \quad (4)$$

$$\mathbf{P}_y = \text{cov}[\mathbf{y}] = E[(\mathbf{y} - \hat{\mathbf{y}})(\mathbf{y} - \hat{\mathbf{y}})^T], \quad (5)$$

$$\mathbf{P}_{xy} = E[(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{y} - \hat{\mathbf{y}})^T]. \quad (6)$$

Unscented transformation: One of the possible solutions is based on the unscented transformation (UT) [8] where the random variable \mathbf{x} is approximated by a set of deterministically chosen σ -points $\{\mathcal{X}_i\}$ with corresponding weights $\{\mathcal{W}_i\}$

$$\mathcal{X}_0 = \hat{\mathbf{x}}, \mathcal{W}_0 = \frac{\kappa}{n_x + \kappa}, \quad (7)$$

$$\mathcal{X}_i = \hat{\mathbf{x}} + \left(\sqrt{(n_x + \kappa)\mathbf{P}_x} \right)_i, \mathcal{W}_i = \frac{1}{2(n_x + \kappa)}, \quad (8)$$

$$\mathcal{X}_{n_x+i} = \hat{\mathbf{x}} - \left(\sqrt{(n_x + \kappa)\mathbf{P}_x} \right)_i, \mathcal{W}_{n_x+i} = \mathcal{W}_i, \quad (9)$$

where $i = 1, \dots, n_x$ and the term $\left(\sqrt{(n_x + \kappa)\mathbf{P}_x} \right)_i$ represents i -th column of the matrix $\sqrt{(n_x + \kappa)\mathbf{P}_x}$. Then, each point is transformed via the nonlinear function

$$\mathcal{Y}_i = \mathbf{g}(\mathcal{X}_i), \forall i, \quad (10)$$

and the resulting characteristics are given as

$$\hat{\mathbf{y}}^{\text{UKF}} = \sum_{i=0}^{2n_x} \mathcal{W}_i \mathcal{Y}_i, \quad (11)$$

$$\mathbf{P}_y^{\text{UKF}} = \sum_{i=0}^{2n_x} \mathcal{W}_i (\mathcal{Y}_i - \hat{\mathbf{y}}^{\text{UKF}})(\mathcal{Y}_i - \hat{\mathbf{y}}^{\text{UKF}})^T, \quad (12)$$

$$\mathbf{P}_{xy}^{\text{UKF}} = \sum_{i=0}^{2n_x} \mathcal{W}_i (\mathcal{X}_i - \hat{\mathbf{x}})(\mathcal{Y}_i - \hat{\mathbf{y}}^{\text{UKF}})^T, \quad (13)$$

Note that these results are only an approximation of the true mean and covariance matrices which cannot be generally computed and κ is the scaling parameter influencing accuracy of the approximation.

First order Stirling's interpolation: Another approximation utilises the first order Stirling's interpolation formula (SI1) [4]. To simplify the following steps it is advantageous to introduce a linear transformation of \mathbf{x} , called stochastic decoupling, given by

$$\mathbf{z} = \mathbf{S}_x^{-1} \mathbf{x}, \quad (14)$$

where \mathbf{S}_x is the square-root of the covariance matrix \mathbf{P}_x , i.e. $\mathbf{P}_x = \mathbf{S}_x \mathbf{S}_x^T$. Then, the elements of \mathbf{z} become mutually uncorrelated with unit variance [4], i.e. $E[(\mathbf{z} - \hat{\mathbf{z}})(\mathbf{z} - \hat{\mathbf{z}})^T] = \mathbf{I}$, and the nonlinear function $\mathbf{g}(\cdot)$ can be rewritten to the form

$$\mathbf{y} = \mathbf{g}(\mathbf{x}) = \mathbf{g}(\mathbf{S}_x \mathbf{z}) = \tilde{\mathbf{g}}(\mathbf{z}). \quad (15)$$

The nonlinear function $\tilde{\mathbf{g}}(\cdot)$, approximated by the first order Stirling's interpolation formula, is given by

$$\mathbf{y} = \tilde{\mathbf{g}}(\mathbf{z}) \approx \tilde{\mathbf{g}}(\hat{\mathbf{z}}) + \sum_{i=1}^{n_x} \Delta z_i \frac{\tilde{\mathbf{g}}(\hat{\mathbf{z}} + h\mathbf{e}_i) - \tilde{\mathbf{g}}(\hat{\mathbf{z}} - h\mathbf{e}_i)}{2h}, \quad (16)$$

where \mathbf{e}_i is the i -th column of the identity matrix, Δz_i is the i -th element of $\Delta \mathbf{z} = \mathbf{z} - \hat{\mathbf{z}}$. For the desired characteristics it holds

$$\hat{\mathbf{y}}^{\text{DD1}} = \mathbf{g}(\hat{\mathbf{x}}), \quad (17)$$

$$\mathbf{P}_y^{\text{DD1}} = \frac{1}{4h^2} \sum_{i=1}^{n_x} (\mathbf{g}(\hat{\mathbf{x}} + hs_i) - \mathbf{g}(\hat{\mathbf{x}} - hs_i)) \times (\mathbf{g}(\hat{\mathbf{x}} + hs_i) - \mathbf{g}(\hat{\mathbf{x}} - hs_i))^T, \quad (18)$$

$$\mathbf{P}_{xy}^{\text{DD1}} = \frac{1}{2h} \sum_{i=1}^{n_x} s_i (\mathbf{g}(\hat{\mathbf{x}} + hs_i) - \mathbf{g}(\hat{\mathbf{x}} - hs_i))^T, \quad (19)$$

where s_i is the i -th column of the matrix \mathbf{S}_x and h is the scaling parameter representing a half of the interpolation interval.

Second order Stirling's interpolation: A more exact approximation is feasible by exploiting the second order Stirling's interpolation (SI2) [4, 3] which leads to more accurate characteristics

$$\hat{\mathbf{y}}^{\text{DD2}} = \frac{h^2 - n_x}{h^2} \mathbf{g}(\hat{\mathbf{x}}) + \frac{1}{2h^2} \sum_{i=1}^{n_x} (\mathbf{g}(\hat{\mathbf{x}} + hs_i) + \mathbf{g}(\hat{\mathbf{x}} - hs_i)), \quad (20)$$

$$\mathbf{P}_y^{\text{DD2}} = \mathbf{P}_{y,A}^{\text{DD1}} + \frac{h^2 - 1}{4h^4} \sum_{i=1}^{n_x} (\mathbf{g}(\hat{\mathbf{x}} + hs_i) + \mathbf{g}(\hat{\mathbf{x}} - hs_i) - 2\mathbf{g}(\hat{\mathbf{x}})) (\mathbf{g}(\hat{\mathbf{x}} + hs_i) + \mathbf{g}(\hat{\mathbf{x}} - hs_i) - 2\mathbf{g}(\hat{\mathbf{x}}))^T, \quad (21)$$

$$\mathbf{P}_{xy}^{\text{DD2}} = \frac{1}{2h} \sum_{i=1}^{n_x} s_i (\mathbf{g}(\hat{\mathbf{x}} + hs_i) - \mathbf{g}(\hat{\mathbf{x}} - hs_i))^T. \quad (22)$$

3.2 Algorithm of derivative-free local filters

Algorithms of all local filters have the same structure where the filtering and predictive mean and covariance matrix are recursively computed to obtain the Gaussian approximation of the estimated pdf's [1, 4, 8, 3]. The *structure of the local filter algorithm* for the system described by (1) and (2) can be summarised in the following steps.

Step 1: Set the time instant $k = 0$ and define a priori initial condition by the predictive mean $\hat{\mathbf{x}}_{0|-1} = E[\mathbf{x}_0] = \bar{\mathbf{x}}_0$ and the predictive covariance matrix $\mathbf{P}_{0|-1} = \text{cov}[\mathbf{x}_0] = \mathbf{P}_0$.

Step 2: The state predictive estimate is updated with respect to the last measurement \mathbf{z}_k according to

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_{k|k} (\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1}), \quad (23)$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_{k|k} \mathbf{P}_{z,k|k-1} \mathbf{K}_{k|k}^T, \quad (24)$$

where $\mathbf{K}_{k|k} = \mathbf{P}_{xz,k|k-1} (\mathbf{P}_{z,k|k-1})^{-1}$ is the filter gain and

$$\hat{\mathbf{z}}_{k|k-1} = E[\mathbf{z}_k | \mathbf{z}^{k-1}] = E[\mathbf{h}_k(\mathbf{x}_k) | \mathbf{z}^{k-1}], \quad (25)$$

$$\begin{aligned} \mathbf{P}_{z,k|k-1} &= E[(\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1})(\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1})^T | \mathbf{z}^{k-1}] = \\ &= E[(\mathbf{h}_k(\mathbf{x}_k) - \hat{\mathbf{z}}_{k|k-1}) \times \\ &\quad \times (\mathbf{h}_k(\mathbf{x}_k) - \hat{\mathbf{z}}_{k|k-1})^T | \mathbf{z}^{k-1}] + \mathbf{R}_k, \end{aligned} \quad (26)$$

$$\mathbf{P}_{xz,k|k-1} = E[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})(\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1})^T | \mathbf{z}^{k-1}]. \quad (27)$$

Step 3: The predictive statistics are given by the relations

$$\hat{\mathbf{x}}_{k+1|k} = E[\mathbf{x}_{k+1} | \mathbf{z}^k] = \mathbf{F}_k \hat{\mathbf{x}}_{k|k}, \quad (28)$$

$$\begin{aligned} \mathbf{P}_{k+1|k} &= E[(\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k})(\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k})^T | \mathbf{z}^k] = \\ &= \mathbf{F}_k \mathbf{P}_{k|k} \mathbf{F}_k^T + \mathbf{Q}_k. \end{aligned} \quad (29)$$

Let $k = k + 1$ and algorithm continues by **Step 2**.

As is argued in [12], it is reasonable to consider the filtering pdf and likelihood pdf to be Gaussian with particular means and covariance matrices, i.e. $p(\mathbf{x}_k | \mathbf{z}^k) \approx \mathcal{N}\{\mathbf{x}_k : \hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}\}$ and $p(\mathbf{z}_k | \mathbf{z}^{k-1}) \approx \mathcal{N}\{\mathbf{z}_k : \hat{\mathbf{z}}_{k|k-1}, \mathbf{P}_{z,k|k-1}\}$. With respect to linear state equation (1) the predictive pdf is supposed to be Gaussian as well.

The crucial difference between particular local filters can be found in the approximation applied to the computation of the measurement predictive statistics (25)–(27). The DD1 is based on the approximation of the nonlinear function $\mathbf{h}_k(\cdot)$ by the SI1. This means, that the measurement predictive statistics (25)–(27) can be computed according to (17)–(19), considering $\hat{\mathbf{x}}$ as $\hat{\mathbf{x}}_{k|k-1}$, \mathbf{P}_x as $\mathbf{P}_{k|k-1}$, $\hat{\mathbf{y}}^{\text{DD1}}$ as $\hat{\mathbf{z}}_{k|k-1}$, $\mathbf{P}_y^{\text{DD1}}$ as $\mathbf{P}_{z,k|k-1}$, $\mathbf{P}_{xy}^{\text{DD1}}$ as $\mathbf{P}_{xz,k|k-1}$, and $\mathbf{g}(\cdot)$ as $\mathbf{h}_k(\cdot)$. Analogously, the divided difference filter second order utilises the SI2 (20)–(22) and the unscented Kalman filter utilises the UT (11)–(13). As the measurement predictive statistics are computed using the derivative-free approximations, they depend on the scaling parameter (κ for the UKF and h^2 for the DDF's).

Note that for the linear dynamics (1), the predictive part of all local filters, given by (28) and (29), is the same as

predictive part of the KF and thus independent of the scaling parameter. For nonlinear dynamics, the state predictive statistics (28) and (29), also depend on the scaling parameter [3]. Further discussion concerning nonlinear dynamics can be found in Section 6.

3.3 Standard choice of scaling parameter

The recommended settings for both scaling parameters, namely κ and h , were determined by the analysis of the unscented transformation [8] and Stirling's interpolation [4], respectively. The analysis was based on a term-by-term comparison of the Taylor series of the true mean $\hat{\mathbf{y}}$ and covariance matrix \mathbf{P}_y (stemming from (4) and (5)) with the Taylor series expansion of the approximated statistics, namely with the Taylor expansion of relations (11), (12) for the UT, (17), (18) for the SI1, and (20), (21) for the SI2.

In the case of the UT, the recommended setting of the scaling parameter κ is $\kappa = 3 - n_x$. Note that the introduced UT (11), (12) represents a basic algorithm which suffers from one main weakness, namely a possible loss of positive definiteness of $\mathbf{P}_y^{\text{UKF}}$ (12) for multidimensional variable \mathbf{x} due to negative κ (if $n_x > 3$, then $\kappa < 0$). Thus several improved algorithms have been proposed which differ mainly in the σ -point computation, e.g. scaled, reduced or high order UT or transformation based on the Gauss-Hermite quadrature [11, 6]. Another possibility is to choose $\kappa = 0$ for any dimension n_x which leads to the cubature Kalman filter [12]. The recommended choice of the interval length for approximation using Stirling's interpolation is $h^2 = 3$ [4].

It is necessary to mention that the analysis of the scaling parameter were done provided that the pdf of the random variable \mathbf{x} in (3) is Gaussian. This assumption is not obviously true when the derivative-free approximation techniques are used in the framework of the local filters. Nevertheless, the recommendations concerning parameter setting were adopted for the local filters as well. It should also be mentioned that within the term-by-term comparison it is not possible to evaluate an effect of high-order terms.

To illustrate the impact of the scaling parameter on filter performance the following motivational example is given.

3.4 Motivational example

Let the model describing the bearings only tracking [14] be considered

$$\mathbf{x}_{k+1} = \begin{bmatrix} 0.9 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{x}_k + \mathbf{w}_k, \quad (30)$$

$$z_k = \tan^{-1} \left(\frac{x_{2,k} - \sin(k)}{x_{1,k} - \cos(k)} \right) + v_k, \quad (31)$$

where $k = 0, 1, \dots, N$, $N = 500$, $p(\mathbf{x}_0) = \mathcal{N}\{\mathbf{x}_0 : [20, 5]^T, 0.1\mathbf{I}\}$, $\mathbf{Q}_k = \begin{bmatrix} 0.1 & 0.01 \\ 0.01 & 0.1 \end{bmatrix}$, $R_k = 0.025$, $\forall k$, and \mathbf{I} is the identity matrix of appropriate size.

The state is estimated using five UKF's with scaling parameter equal to $\kappa = 0, 1, 2, 3, 4$. The filter performance

Table 1: MSE of the UKF's for different choices of κ .

	$\kappa = 0$	$\kappa = 1$	$\kappa = 2$	$\kappa = 3$	$\kappa = 4$
MSE	23.66	14.35	9.09	6.20	4.79

is measured using the mean square error (MSE) using $M = 10^3$ Monte Carlo simulations, i.e.

$$\text{MSE} = \frac{\sum_{m=1}^M \sum_{k=0}^N \sum_{i=1}^{n_x} (x_{i,k}^{(m)} - \hat{x}_{i,k|k}^{(m)})^2}{M(N+1)n_x}, \quad (32)$$

where $x_{i,k}^{(m)}$ is the i -th component of the true state in the m -th MC simulation at time k and $\hat{x}_{i,k|k}^{(m)}$ its filtering estimate.

In Table 1 the resulting MSE's are summarised. It can be seen that although the recommended choice of the scaling parameter is $\kappa = 3 - n_x = 1$ according to [8] and $\kappa = 0$ according to [12], a better estimation quality can be obtained by setting different value of the scaling parameter. Therefore, the scaling parameter is still rather a user-defined parameter.

Table 1 demonstrates that the proposed rules and recommendations concerning parameter choice need not lead to the best results. One of the reason of these results is that the scaling parameter is chosen a priori, without respecting particular system description. At best, its choice depends on the system order. The question which may arise is: How to choose the scaling parameter with respect to particular system description?

In this example with five considered choices of κ given in Table 1, the best choice according to the MSE is $\kappa = 4$. However, it should be noted that another value of κ can be suitable for a different nonlinear model. The parameter κ depends on both the function in the measurement equation and the properties of the noises. Because of the fact the considered model is nonlinear, the scaling parameter might also depend on the particular operating point (on the part of the state space in which the filter is operating), i.e. on the estimated means and covariance matrices. Therefore, it seems to be reasonable to suppose that the scaling parameter is time variant and to find it an adaptive algorithm should be considered.

Unfortunately, no algorithm for adaptive setting of the scaling parameter has been proposed yet.

3.5 Problem statement

The aim of this paper is thus to propose an adaptive technique which allows to determine the scaling parameters with respect to particular system description and estimated state, i.e. estimated mean and covariance matrix. The novel adaptive technique should be applicable to all derivative-free local filters with one or more scaling parameters. The technique is proposed in the following section.

4 Adaptive Choice of Scaling Parameter

In this section the novel adaptive technique for determination of the scaling parameters of the derivative-free filters is proposed. The idea of the adaptive technique is that the likelihood function is maximised with respect to the scaling parameter in the filtering step at each time instant.

Note that for the sake of presentation clarity any scaling parameter of the derivative-free local filter will be denoted in this section as $\boldsymbol{\theta} \in \mathbb{R}^{n_\theta}$, i.e. in the case of the UKF $\theta = \kappa$ and in the case of the DDF's $\theta = h$. Moreover, $\boldsymbol{\theta}$ can represent a vector of scaling parameters because e.g. the modified version of the unscented transformation is based on two scaling parameters [11].

4.1 Likelihood function

The algorithm of the derivative-free local filter is given by relations (23)–(29). It can be seen that the only statistics directly depending on the scaling parameter $\boldsymbol{\theta}$, with respect to the system description (1), (2), are the predictive mean of the measurement $\hat{\mathbf{z}}_{k|k-1} = \hat{\mathbf{z}}_{k|k-1}(\boldsymbol{\theta})$ (25), its covariance matrix $\mathbf{P}_{z,k|k-1} = \mathbf{P}_{z,k|k-1}(\boldsymbol{\theta})$ (26), and the predictive cross-covariance matrix of state and measurement $\mathbf{P}_{xz,k|k-1} = \mathbf{P}_{xz,k|k-1}(\boldsymbol{\theta})$ (27). Thus, the approximate likelihood function

$$\hat{p}(\mathbf{z}_k | \mathbf{z}^{k-1}, \boldsymbol{\theta}) = \mathcal{N}\{\mathbf{z}_k : \hat{\mathbf{z}}_{k|k-1}(\boldsymbol{\theta}), \mathbf{P}_{z,k|k-1}(\boldsymbol{\theta})\} \quad (33)$$

also depends on the scaling parameter.

Because the measurement \mathbf{z}_k is available at time instant k , it is possible to compute the maximal likelihood estimate of the scaling parameter $\boldsymbol{\theta}$ at time k , which is of the form

$$\hat{\boldsymbol{\theta}}_k = \arg \max_{\boldsymbol{\theta}} \hat{p}(\mathbf{z}_k | \mathbf{z}^{k-1}, \boldsymbol{\theta}). \quad (34)$$

The particular value of the likelihood function for the estimate is then

$$\zeta_k = \hat{p}(\mathbf{z}_k | \mathbf{z}^{k-1}, \hat{\boldsymbol{\theta}}_k). \quad (35)$$

4.2 Maximisation of likelihood function

Unfortunately, in most cases it is not possible to find a closed-form solution to relation (34) with respect to relations (25), (26) computed using the UT, the SI1 or the SI2 and it is necessary to use a numerical technique for the maximisation.

There exist many general optimisation methods for finding an extreme of a nonlinear function. As an example the gradient method, Newton-type methods or the random search maximisation method can be mentioned [15]. The likelihood function (33) is not, however, usually strongly concave and thus there are more (local) maxima. Therefore, the gradient method or the Newton-type methods are not suitable because they often converge to a local maximum. Moreover, these methods require computation of the derivative of the criterion function which can be tedious or

even impossible. On the other hand, the random search method belonging into global optimisation methods has significantly higher ability to find a global maximum [15]. As other advantages of the method the following can be mentioned: very simple implementation, acceptable computational demands, facile incorporation of a scaling parameters domain (e.g. the scaling parameter κ in the UT cannot be negative otherwise the covariance matrix $\mathbf{P}_{z,k|k-1}$ (26) may lost the positive-semidefiniteness [8]), applicability to non-differentiable or discontinuous functions. The *adaptive random search maximisation method* [15] considered in this paper can be outlined as follows:

Step 1: Specify the feasible domain¹ for the scaling parameter vector $\boldsymbol{\theta}_k$ at time k , i.e.

$$\mathbb{D} = \{\boldsymbol{\theta}_k | \theta_{i,k,\min} \leq \theta_{i,k} \leq \theta_{i,k,\max}, i = 1, 2, \dots, n_\theta\}, \quad (36)$$

where $\theta_{i,k}$ is the i -th element of $\boldsymbol{\theta}_k$ and $\theta_{i,k,\min}$ and $\theta_{i,k,\max}$ its minimum and maximum allowed value, respectively.

Step 2: Choose the initial condition for recursion, i.e. $\hat{\boldsymbol{\theta}}_k^0$, according to

$$\hat{\theta}_{i,k}^0 = \frac{\theta_{i,k,\min} + \theta_{i,k,\max}}{2}, i = 1, 2, \dots, n_\theta, \quad (37)$$

and set $j = 0$.

Step 3: Compute a trial vector $\hat{\boldsymbol{\theta}}_k^{j+}$ as

$$\hat{\boldsymbol{\theta}}_k^{j+} = \hat{\boldsymbol{\theta}}_k^j + \mathbf{e}^j, \quad (38)$$

where \mathbf{e}^j is a realisation of a random vector generated according to $\mathcal{N}\{\mathbf{e} : \mathbf{0}, \mathbf{P}_e^j\}$ and the covariance matrix \mathbf{P}_e^j is suitably (often adaptively) chosen.

Step 4: If $\hat{p}(\mathbf{z}_k | \mathbf{z}^{k-1}, \hat{\boldsymbol{\theta}}_k^{j+}) > \hat{p}(\mathbf{z}_k | \mathbf{z}^{k-1}, \hat{\boldsymbol{\theta}}_k^j)$ then $\hat{\boldsymbol{\theta}}_k^{j+1} = \hat{\boldsymbol{\theta}}_k^{j+}$, otherwise $\hat{\boldsymbol{\theta}}_k^{j+1} = \hat{\boldsymbol{\theta}}_k^j$.

Step 5: Increment j by one and go to **Step 3**.

Note that in the section devoted to the numerical illustrations also a numerical “grid method” will be utilised for finding the global maximum of (34) within the domain. It means that the domain is covered by an equally spaced grid of points and the likelihood function is evaluated at these points. Then, the grid point with the maximal value of the likelihood function, i.e. with maximal ζ_k (35), is chosen and considered to be the resulting estimate $\hat{\boldsymbol{\theta}}_k$ (34).

5 Derivative-free filter with adaptive choice of scaling parameter

The algorithm of the derivative-free filter with adaptive choice of scaling parameter can be summarised as follows.

Step 1: Set the time instant $k = 0$ and define a priori initial condition as $\hat{\mathbf{x}}_{0|-1} = \bar{\mathbf{x}}_0$ and $\mathbf{P}_{0|-1} = \mathbf{P}_0$.

¹The feasible domain for κ in the UKF is $\kappa \in [0, \infty)$ and for h in the DDF's is $h \in (0, \infty)$.

Table 2: MSE, average filtering covariance, and computational costs of the UKF's for different choices of κ .

	$\kappa = 0$	$\kappa = 4$	$\kappa \in \{0:0.1:4\}$	$\kappa \in \{0:1:4\}$	$\kappa \in \{0:4:4\}$
MSE	23.66	4.79	2.69	2.75	2.76
V_c	0.19	0.26	0.31	0.31	0.32
time	0.0016		0.0330	0.0060	0.0030

Step 2: The vector of scaling parameters $\hat{\boldsymbol{\theta}}_k$ is computed so that the likelihood function (33) is maximised. Maximisation can be performed e.g. by the adaptive random search maximisation method.

Step 3: The state predictive estimate is updated with respect to the last measurement \mathbf{z}_k according to

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_{k|k}(\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1}), \quad (39)$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_{k|k}\mathbf{P}_{z,k|k-1}\mathbf{K}_{k|k}^T, \quad (40)$$

where $\mathbf{K}_{k|k} = \mathbf{P}_{xz,k|k-1}(\mathbf{P}_{z,k|k-1})^{-1}$ is the gain and $\hat{\mathbf{z}}_{k|k-1}$, $\mathbf{P}_{z,k|k-1}$, $\mathbf{P}_{xz,k|k-1}$ are computed by means of relations (25)–(27), chosen approximation (e.g. the UT or the SI), and the scaling parameter $\hat{\boldsymbol{\theta}}_k$.

Step 4: The predictive statistics are given by the relations

$$\hat{\mathbf{x}}_{k+1|k} = E[\mathbf{x}_{k+1}|\mathbf{z}^k] = \mathbf{F}_k\hat{\mathbf{x}}_{k|k}, \quad (41)$$

$$\begin{aligned} \mathbf{P}_{k+1|k} &= E[(\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k})(\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k})^T|\mathbf{z}^k] = \\ &= \mathbf{F}_k\mathbf{P}_{k|k}\mathbf{F}_k^T + \mathbf{Q}_k. \end{aligned} \quad (42)$$

Let $k = k + 1$ and algorithm continues by **Step 2**.

This algorithm has the same structure as the algorithm described by relations (23)–(29). The only difference is computation of the vector of scaling parameters $\hat{\boldsymbol{\theta}}_k$ maximising likelihood function at each time instant.

6 Numerical Illustrations and Software Implementation of Algorithm

The novel adaptive technique for scaling parameter selection is illustrated and discussed in this section using two examples. Also its software implementation is treated.

6.1 Bearings only tracking

Let the bearings only tracking problem defined in Section 3.4 by relations (30), (31) be considered as the first example.

In this example the UKF will be mainly considered as a representative of the derivative-free local filters. The performance of the UKF's with respect to different choices of the scaling parameter κ will be measured using the mean square error criterion (32) and the criterion characterising the average filtering covariance matrix $\mathbf{P}_{k|k}$ (24) given by

$$V_c = \frac{\sum_{m=1}^M \sum_{k=0}^N |\mathbf{P}_{k|k}^{(m)}|}{M(N+1)}, \quad (43)$$

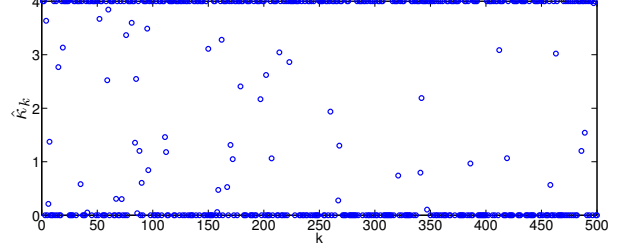


Figure 1: Values of the scaling parameter κ with maximal likelihood function.

where $|\mathbf{P}_{k|k}^{(m)}|$ is the determinant of the filtering covariance matrix in the m -th MC simulation.

Table 2 summarises the results for two fixed values of κ and three adaptive choices of κ with respect to (34), where $\hat{\boldsymbol{\theta}}_k = \hat{\kappa}_k$. The used notation $\kappa \in \{\kappa_{min} : \kappa_{step} : \kappa_{max}\}$ means that the likelihood function (33) is maximised at each time instant k by means of the grid method on interval κ_{min} and κ_{max} with the increment κ_{step} . From the table it can be seen that the adaptive choice of the scaling parameter with respect to maximisation of the likelihood function has a significant impact on the improvement of the estimation quality measured by the MSE. In this case, the impact of the chosen increment κ_{step} is rather insignificant. The reason is the likelihood function $\hat{p}(\mathbf{z}_k|\mathbf{z}^{k-1}, \kappa_k)$ is at most time instants monotonous at given interval $\kappa_{min} = 0$ and $\kappa_{max} = 4$ and the optimum scaling parameter estimate $\hat{\kappa}_k$ usually lies on the bounds. This can be illustrated by Figure 1 where the maximal likelihood estimate of $\hat{\kappa}_k$ is shown. An example of behaviour of the particular likelihood function can be found in Figure 2. The last row in Table 2 illustrates average computational costs in seconds necessary for computation of filtering and prediction step at one time instant. It can be seen that better results can be obtained by the UKF with the adaptive choice of the scaling parameter with a slight increase of the computational demands (as in case of the UKF($\kappa = \{0 : 4 : 4\}$)). For completeness, a typical example of the true and estimated state behaviour is shown in Figure 3. It should be noted that contrary to the UKF($\kappa \in \{0 : 0.1 : 4\}$) with adaptive choice of the scaling parameter, the UKF($\kappa = 2$) with fixed κ often diverges for a few time instants.

Better results can be obtained using the UKF with adaptive choice of the scaling parameter κ computed within a larger domain, e.g. $\kappa_{min} = 0$ and $\kappa_{max} = 9$. The results are given in Table 3. Extending the domain of the scaling parameter naturally increases the computational cost. However, the computational costs in case of the grid method for an one-dimensional scaling parameter increase almost linearly with respect to the domain enlargement.

The above mentioned results of the UKF with adaptive choice of κ were obtained using the grid method (GM). However, to find a maximum of the likelihood function the random search maximisation algorithm (RS) outlined in Section 4.2 can also be exploited. The MSE and computational demands of the UKF's with adaptive choice of κ

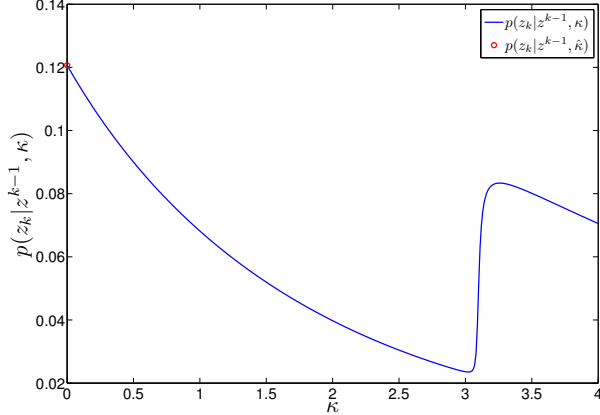


Figure 2: Example of dependency of the likelihood function $\hat{p}(z_k | z^{k-1}, \kappa)$ on the scaling parameter κ .

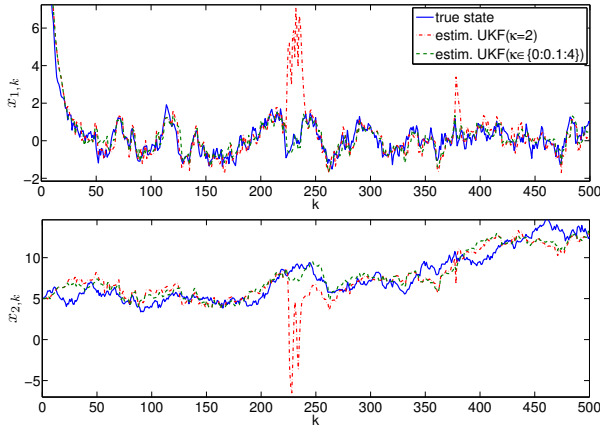


Figure 3: Typical example of true and estimated state.

performed using the GM and the RS are shown in Table 4. The used notation $\kappa \in \langle \kappa_{min}, \kappa_{max} \rangle$ means that the maximum likelihood estimate $\hat{\kappa}$ were computed by the RS within the domain given by κ_{min} and κ_{max} . It can be seen that both maximisation techniques gives almost the same results (for sufficiently small κ_{step} in the GM). The RS has generally higher computational costs. However, the RS seems to be more efficient for derivative-free filters with more than one scaling parameter (for example the UKF mentioned in [11]) or for optimisation within a larger domain. The maximal likelihood estimates of $\hat{\kappa}_k$ for the UKF(GM, $\kappa \in \{0 : 0.1 : 9\}$) and the UKF(RS, $\kappa \in \langle 0, 9 \rangle$) are shown in Figure 4.

For completeness, the estimation quality of the DD1 and DD2 for different values of the scaling parameter h (either constant or adaptively chosen) is compared in Table 5 and Table 6, respectively.

Table 3: MSE, average filtering covariance, and computational costs of the UKF's for different choices of κ .

	$\kappa = 5$	$\kappa = 8$	$\kappa = 9$	$\kappa \in \{0:0.1:9\}$	$\kappa \in \{0:1:9\}$
MSE	4.22	3.69	3.66	1.68	1.68
V_c	0.26	0.22	0.21	0.32	0.32
time	0.0016			0.0720	0.0100

Table 4: MSE and computational costs of the UKF's with respect to different maximisation technique.

	GM, $\kappa \in \{0:0.1:9\}$	GM, $\kappa \in \{0:0.01:9\}$	RS, $\kappa \in \langle 0, 9 \rangle$
MSE	1.68	1.68	1.72
time	0.072	0.670	0.340

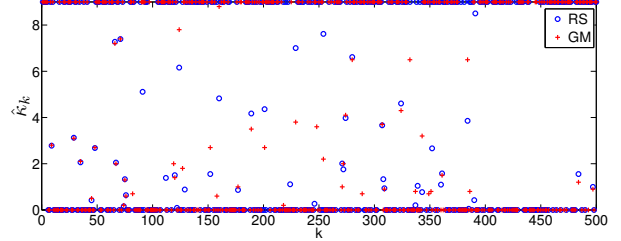


Figure 4: Values of the scaling parameter κ with maximal likelihood function.

6.2 Nonlinear dynamics

As the second example, the following nonlinear system was chosen [16]

$$x_{k+1} = (1 - 0.05\Delta T)x_k + 0.04\Delta T x_k^2 + w_k, \quad (44)$$

$$z_k = x_k^2 + x_k^3 + v_k, \quad (45)$$

where $\Delta T = 0.01$, $k = 0, 1, \dots, N$, $N = 150$, $p(x_0) = \mathcal{N}\{x_0 : 2.3, 0.01\}$, $Q_k = 0.5$, and $R_k = 0.09, \forall k$.

For the nonlinear system, the derivative-free local filters still have the same structure as is given by relations (23)–(29). The only difference is in relations (28), (29) for predictive statistics computation which cannot be computed exactly and some approximation (e.g. the UT, the SI1 or the SI2) has to be used. More details can be found in [3].

Therefore, the predictive statistics of the state also depend on the scaling parameter, i.e. $\hat{\mathbf{x}}_{k+1|k} = \hat{\mathbf{x}}_{k+1|k}(\boldsymbol{\theta})$ and $\mathbf{P}_{k+1|k} = \mathbf{P}_{k+1|k}(\boldsymbol{\theta})$, and the maximal likelihood estimate of the scaling parameter $\hat{\theta}_k$ from filtering step is used in their computation as well.

The estimation quality of the UKF's with fixed and adapted scaling parameter is compared using two criteria (32) and (43) and the results are given in Table 7. Again, the UKF with the adaptive choice of the scaling parameter κ provides the better estimation quality with respect to the UKF's with fixed scaling parameters.

6.3 Implementation of algorithm

Both numerical illustrations were performed with the aid of the Nonlinear Estimation Framework (NEF,

Table 5: MSE and average filtering covariance of the DD1's for different choices of h .

	$h = 1$	$h = 3$	$h = 4$	$h \in \{1:0.1:4\}$	$h \in \{1:1:10\}$
MSE	37.67	26.22	21.89	2.47	3.11
V_c	0.13	0.08	0.07	0.14	0.09

Table 6: MSE and average filtering covariance of the DD2's for different choices of h .

	$h = 1$	$h = 3$	$h = 4$	$h \in \{1:0.1:4\}$	$h \in \{1:1:10\}$
MSE	36.52	3.78	6.79	1.58	1.88
V_c	0.14	0.24	0.15	0.28	0.16

Table 7: MSE and average filtering covariance of the UKF's for different choices of κ for nonlinear system.

	$\kappa = 0$	$\kappa = 1$	$\kappa = 3$	$\kappa = 4$	$\kappa \in \{0:0.1:4\}$
MSE	0.77	0.22	0.11	0.12	0.08
V_c	0.02	0.05	0.08	0.09	0.04

<http://nft.kky.zcu.cz/nef>) [17] which is a collection of MATLAB functions for modelling system behaviour, state estimation (supporting prediction, filtering, and smoothing), and evaluation of the results. Among other estimators, the UKF and the DDF's with possibility to adaptively choose the scaling parameters are implemented in the package NEF.

7 Concluding Remarks

The paper dealt with the choice of the scaling parameters in the derivative-free local filters. The standard and widely-used recommendations for the scaling parameter choice were discussed and their weaknesses were highlighted. The novel adaptive technique for the scaling parameter setting was proposed. The technique is based on maximisation of the approximate likelihood function with respect to the scaling parameter at each time instant. Because the maximisation is exactly unsolvable in general, numerical solutions were discussed and presented. The derivative-free local filters, namely the unscented Kalman filter and the divided difference filters, with technique for adaptive setting of the scaling parameter were illustrated using two numerical examples. It was shown that they provide estimates with significantly higher quality measured by the mean square error than the filters with fixed parameters.

Acknowledgement: The work was supported by the Ministry of Education, Youth and Sports of the Czech Republic, project no. 1M0572, and by the Czech Science Foundation, project no. GACR 102/08/0442.

References

- [1] B. D. O. Anderson and S. B. Moore, *Optimal Filtering*. New Jersey: Prentice Hall Ins.: Englewood Cliffs, 1979.
- [2] H. W. Sorenson, "On the development of practical nonlinear filters," *Inf. Sci.*, vol. 7, pp. 230–270, 1974.
- [3] M. Šimandl and J. Duník, "Derivative-free estimation methods: New results and performance analysis," *Automatica*, vol. 45, no. 7, pp. 1749–1757, 2009.
- [4] M. Nørgaard, N. K. Poulsen, and O. Ravn, "New developments in state estimation for nonlinear systems," *Automatica*, vol. 36, no. 11, pp. 1627–1638, 2000.
- [5] T. S. Schei, "A finite-difference method for linearization in nonlinear estimation algorithms," *Automatica*, vol. 33, no. 11, pp. 2053–2058, 1997.
- [6] K. Ito and K. Xiong, "Gaussian filters for nonlinear filtering problems," *IEEE Transactions on Automatic Control*, vol. 45, no. 5, pp. 910–927, 2000.
- [7] J. Duník, M. Šimandl, O. Straka, and L. Král, "Performance analysis of derivative-free filters," in *Proceedings of the 44th IEEE Conference on Decision and Control, and European Control Conference ECC'05*, pp. 1941–1946, Seville, Spain, December 2005.
- [8] S. J. Julier, J. K. Uhlmann, and H. F. Durrant-White, "A new method for the nonlinear transformation of means and covariances in filters and estimators," *IEEE Transactions on Automatic Control*, vol. 45, no. 3, pp. 477–482, 2000.
- [9] H. Kushner and A. Budhiraja, "A nonlinear filtering algorithm based on an approximation of the conditional distribution," *IEEE Transactions on Automatic Control*, vol. 45, no. 3, pp. 580–585, 2000.
- [10] B. Quine, "A derivative-free implementation of the extended Kalman filter," *Automatica*, vol. 42, no. 11, pp. 1927–1934, 2006.
- [11] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–421, 2004.
- [12] I. Arasaratnam and S. Haykin, "Cubature Kalman filters," *IEEE Transactions on Automatic Control*, vol. 54, no. 6, pp. 1254–1269, 2009.
- [13] T. Lefebvre, H. Bruyninckx, and J. De Schuller, "Comment on "A new method for the nonlinear transformation of means and covariances in filters and estimators","" *IEEE Transactions on Automatic Control*, vol. 47, no. 8, pp. 1406–1409, 2002.
- [14] M. Šimandl, J. Královec, and T. Söderström, "Anticipative grid design in point-mass approach to nonlinear state estimation," *IEEE Transactions on Automatic Control*, vol. 47, no. 4, pp. 699–702, 2002.
- [15] É. Walter and L. Pronzato, *Identification of Parametric Models from Experimental Data*. Springer, 1997.
- [16] J. V. Candy, *Model-Based Signal Processing*. Hoboken, New Jersey: John Wiley & Sons, 2006.
- [17] O. Straka, M. Flídr, J. Duník, and M. Šimandl, "A software framework and tool for nonlinear state estimation," in *Proceedings of the 15th IFAC Symposium on System Identification*, pp. 510–515, Saint-Malo, France, July 2009.