Measures of Non-Gaussianity in Unscented Kalman Filter Framework

Ondřej Straka, Jindřich Duník, and Miroslav Šimandl
European Centre of Excellence - New Technologies for the Information Society,
Department of Cybernetics, Faculty of Applied Sciences,
University of West Bohemia, Univerzitní 8, 306 14 Pilsen, Czech Republic
Email: straka30@kky.zcu.cz, dunikj@kky.zcu.cz, simandl@kky.zcu.cz

***Abstract*—The paper deals with state estimation of nonlinear stochastic dynamic systems. In particular, the stress is laid on the recursive computation of the higher order moments of the state estimate in a local filter. The higher order moments are then used as a cornerstone of non-Gaussianity measures having the ability to indicate a possible decrease of the estimate quality. A technique is proposed for a general algorithm of the local filters, detailed for the unscented Kalman filter, and illustrated in numerical examples.**

Keywords: state estimation, nonlinear filters, Kalman filtering, non-Gaussianity measures, nonlinearity measures

I. Introduction

Recursive state estimation of nonlinear discrete-time stochastic dynamic systems from noisy or incomplete measured data has been a subject of considerable research interest for the last four decades. It plays an essential role in fields such as target tracking, navigation, signal processing, fault detection, and adaptive and optimal control problems.

General solution to the recursive state estimation problem is given by the Bayesian recursive relations (BRRs) computing the probability density functions (PDFs) of the state conditioned by the measurements. These PDFs represent a complete description of the state, which itself cannot be fully measured. The closed-form solution to the BRRs is available only for a few special cases, such as the linear Gaussian system [1]. In other cases, approximate approaches must be used. The filters using the approximate approaches can be divided with respect to the validity of the resulting estimates into global and local filters [2]–[4].

The global filters provide estimates in the form of conditional PDFs valid within almost whole state space. They are capable to estimate the state of a strongly nonlinear or non-Gaussian system but usually at the cost of substantial computational demands. Among these filters, the point-mass filter [5], the Gaussian sum filter [6], and the particle filter [7] have lately attracted considerable attention.

As opposed to the global filters, the local filters provide computationally feasible estimates predominantly in the form of the conditional mean and the covariance matrix of the estimate error[1] but with limited validity (within a close neighbourhood of a point estimate only). These filters offer sufficient estimation performance (with possibly bounded estimation error) mainly for systems with mild nonlinearities, exact initial condition, and small enough disturbing noises [8]. The satisfactory performance of local filters (LFs) is also conditioned by the Gaussian-like shape of the conditional PDFs, i.e., the unimodal and non-heavy-tailed PDF. Indeed, many recently developed LFs are based on the assumption of Gaussian conditional PDF. Among those, the filters based on various deterministic or stochastic integration rules or Fourier-Hermite expansion [3], [9]–[13] can be mentioned. To have an instrument to monitor whether a particular LF assumption is valid, non-linearity measures (NLMs) and non-Gaussianity measures (NGMs) have been recently a focal point of several studies and quite a few local and global measures have been proposed [14]–[17].

The global measures aim for assessment of the overall nonlinearity of the system or the overall impact of the nonlinearity on the PDFs of the state (without regard to estimation method selection). The measures are intended for an off-line analysis (to facilitate to answer the question whether a LF is expected to provide satisfactorily results or a global filter should be used instead). The conclusions stemming from the analysis are, however, strictly tied with a problem set-up (e.g., initial condition, noise properties); any change in the set-up results in a necessity to remake the analysis.

The local measures, on the other hand, are designed for an on-line (self-monitoring) monitoring of local filters. It means that at each time instant the nonlinear transformation in filtering or prediction step of the LF is assessed with respect to its severity or its impact on the properties of the propagated state estimate. The vast majority of the local NGMs have been designed as instant measures, i.e., the measures[2] at two subsequent time instants are independent; even the measure in the filtering step is independent of the measure in the prediction step at the same time. Thus, these measures have a very limited capability in monitoring of a transformation cumulative effect (as was illustrated in [17], a sequence of mild nonlinearities with very low values of nonlinearity measures or NGMs on, for example, an initially Gaussian PDF, might result in a heavy-tailed or multimodal distribution negatively

[1]First two moments usually do not represent a full description of the immeasurable state.

[2]In this paper, the measure form and its value are not distinguished. Both are simply called measure.

impacting the precision of the numerical integration rules). For this reason, the cumulative local measure based on a recursive propagation of the third moment has been studied in [18] but for scalar systems only.

The goal of the paper is to design an UKF with recursive computation of higher order moments. Those moments are subsequently used as a base for non-Gaussianity measures (NGMs) indicating possible deviation of the state conditional PDF from a desired Gaussian-like one (and thus potential degradation of the estimation quality). The proposed algorithm is illustrated in a multi-dimensional tracking example with a constant-turn model.

The rest of the paper is organized as follows. Section II provides systems specification, a generic LF algorithm with focus on the UKF, and discussion of nonlinearity and non-Gaussianity measures. Recursive computation of the higher order moments of the state estimate in the LF framework are discussed in Sections III and IV. Numerical illustration and concluding remarks are given in Section V and VI, respectively.

II. System Description, State Estimation, and Measures

A. System description

Let the discrete-time stochastic system be considered

$$\mathbf{x}_{k+1} = \mathbf{f}_k(\mathbf{x}_k) + \mathbf{w}_k, k = 0, 1, 2, \ldots, \tag{1}$$

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

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, $\mathbf{f}_k : \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_x}$ and $\mathbf{h}_k : \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_z}$ are known vector functions, and $\mathbf{w}_k \in \mathbb{R}^{n_x}$ and $\mathbf{v}_k \in \mathbb{R}^{n_z}$ are the state and measurement white noises. The PDFs of the noises are supposed to be Gaussian with zero means and known covariance matrices $\mathbf{\Sigma}_k^{\mathbf{w}}$ and $\mathbf{\Sigma}_k^{\mathbf{v}}$, i.e., $p_{\mathbf{w}_k}(\mathbf{w}_k) = \mathcal{N}\{\mathbf{w}_k : \mathbf{0}, \mathbf{\Sigma}_k^{\mathbf{w}}\}$ and $p_{\mathbf{v}_k}(\mathbf{v}_k) = \mathcal{N}\{\mathbf{v}_k : \mathbf{0}, \mathbf{\Sigma}_k^{\mathbf{v}}\}$, respectively. The PDF of the initial state is known, Gaussian, i.e., $p_{\mathbf{x}_0}(\mathbf{x}_0) = \mathcal{N}\{\mathbf{x}_0 : \bar{\mathbf{x}}_0, \mathbf{P}_0^{\mathbf{xx}}\}$, and independent of the noises.

B. State estimation by local filters

The aim of a LF is to compute the first two moments of the state conditioned by the measurements, namely, the conditional mean $\hat{\mathbf{x}}_{k|k} = \mathsf{E}[\mathbf{x}_k|\mathbf{z}^k]$ and covariance matrix of the estimate error $\mathbf{P}_{k|k}^{\mathbf{xx}} = \mathsf{cov}[\mathbf{x}_k|\mathbf{z}^k]$ in which $\mathbf{z}^k = [\mathbf{z}_0, \mathbf{z}_1, \ldots, \mathbf{z}_k]$. The moments might be understood as a Gaussian approximation of the conditional PDF, i.e., $p(\mathbf{x}_k|\mathbf{z}^k) \approx \mathcal{N}\{\mathbf{x}_k : \hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}^{\mathbf{xx}}\}$ [3], [12], depending on the type of used approximation.

Approximation techniques used in the LF design and subsequently also the LFs[3] can be divided into two groups[4]; derivative and derivative-free. *Derivative* techniques approximate nonlinear functions in the system description by derivative-based expansions. As an example, techniques based on the Taylor or Fourier-Hermite series can be mentioned which lead to e.g., the extended Kalman filter (EKF), second order filter, or Fourier-Hermite Kalman filter [13], [19]. *Derivative-free* techniques in the LF design are based on differential-based polynomial interpolations, the unscented transform (UT), or various numerical integration rules. The filters based on the derivative-free techniques are represented by the divided difference filters (DDFs) utilizing the Stirling polynomial interpolation [20], unscented Kalman filter (UKF) using the UT [21], or the quadrature, cubature, and stochastic integration based filters utilizing deterministic and stochastic integration rules [3], [12], [22].

All the LF algorithms follow the structure of the following generic local filter algorithm [3], [4]:

Algorithm 1: Generic Local Filter

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

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

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1}), \tag{3}$$

$$\mathbf{P}_{k|k}^{\mathbf{xx}} = \mathbf{P}_{k|k-1}^{\mathbf{xx}} - \mathbf{K}_k \mathbf{P}_{k|k-1}^{\mathbf{zz}} \mathbf{K}_k^T, \tag{4}$$

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

$$\hat{\mathbf{z}}_{k|k-1} = \mathsf{E}[\mathbf{z}_k|\mathbf{z}^{k-1}] = E[\mathbf{h}_k(\mathbf{x}_k)|\mathbf{z}^{k-1}], \tag{5}$$

$$\begin{aligned}\mathbf{P}_{k|k-1}^{\mathbf{zz}} &= \mathsf{E}[(\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1})(\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1})^T|\mathbf{z}^{k-1}] = \\ &= \mathsf{E}[(\mathbf{h}_k(\mathbf{x}_k) - \hat{\mathbf{z}}_{k|k-1}) \times \\ &\times (\mathbf{h}_k(\mathbf{x}_k) - \hat{\mathbf{z}}_{k|k-1})^T|\mathbf{z}^{k-1}] + \mathbf{\Sigma}_k^{\mathbf{v}},\end{aligned} \tag{6}$$

$$\mathbf{P}_{k|k-1}^{\mathbf{xz}} = \mathsf{E}[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})(\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1})^T|\mathbf{z}^{k-1}]. \tag{7}$$

Step 3: (**prediction**) The predictive statistics are given by the relations

$$\hat{\mathbf{x}}_{k+1|k} = \mathsf{E}[\mathbf{x}_{k+1}|\mathbf{z}^k] = E[\mathbf{f}_k(\mathbf{x}_k)|\mathbf{z}^k], \tag{8}$$

$$\begin{aligned}\mathbf{P}_{k+1|k}^{\mathbf{xx}} &= \mathsf{E}[(\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k})(\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k})^T|\mathbf{z}^k] = \\ &= \mathsf{E}[(\mathbf{f}_k(\mathbf{x}_k) - \hat{\mathbf{x}}_{k|k})(\mathbf{f}_k(\mathbf{x}_k) - \hat{\mathbf{x}}_{k|k})^T|\mathbf{z}^k] + \mathbf{\Sigma}_k^{\mathbf{w}}.\end{aligned} \tag{9}$$

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

The particular LFs differ in approximation applied for the solution to the measurement and state predictive statistics (5)–(7) and (8)–(9). For example, the UKF computes approximate predictive statistics of the state as

$$\hat{\mathbf{x}}_{k+1|k} = \sum_{i=0}^{2n_x} \mathcal{W}_{k|k}^i \mathcal{X}_{k+1|k}^i, \tag{10}$$

$$\mathbf{P}_{k+1|k}^{\mathbf{xx}} = \sum_{i=0}^{2n_x} \mathcal{W}_{k|k}^i (\mathcal{X}_{k+1|k}^i - \hat{\mathbf{x}}_{k+1|k})(\cdot)^T + \mathbf{\Sigma}_k^{\mathbf{w}}, \tag{11}$$

[3] In literature instead of the term 'local filters', terms such as 'Gaussian filters' or 'Kalman filters' can be found.

[4] Note that this is just one of the possible categorisations; another might be from the perspective whether the model or the state estimate description is approximated.

and of the measurement as

$$\hat{\mathbf{z}}_{k|k-1} = \sum_{i=0}^{2n_x} \mathcal{W}^i_{k|k-1} \mathcal{Z}^i_{k|k-1}, \tag{12}$$

$$\mathbf{P}^{\mathbf{zz}}_{k|k-1} = \sum_{i=0}^{2n_x} \mathcal{W}^i_{k|k-1} (\mathcal{Z}^i_{k|k-1} - \hat{\mathbf{z}}_{k|k-1})(\cdot)^T + \mathbf{\Sigma}^{\mathbf{v}}_k, \tag{13}$$

$$\mathbf{P}^{\mathbf{xz}}_{k|k-1} = \sum_{i=0}^{2n_x} \mathcal{W}^i_{k|k-1} (\mathcal{X}^i_{k|k-1} - \hat{\mathbf{x}}_{k|k-1})(\mathcal{Z}^i_{k|k-1} - \hat{\mathbf{z}}_{k|k-1})^T, \tag{14}$$

where the filtering and predictive σ-points and respective weights are given as

$$\mathcal{X}^{0:2n_x}_{k|k} = \hat{\mathbf{x}}_{k|k} \mathbf{1}_{1\times b} + c\left[\mathbf{0}_{n_x\times 1}, \sqrt{\mathbf{P}^{\mathbf{xx}}_{k|k}}, -\sqrt{\mathbf{P}^{\mathbf{xx}}_{k|k}}\right], \tag{15}$$

$$\mathcal{X}^{0:2n_x}_{k|k-1} = \hat{\mathbf{x}}_{k|k-1} \mathbf{1}_{1\times b} + c\left[\mathbf{0}_{n_x\times 1}, \sqrt{\mathbf{P}^{\mathbf{xx}}_{k|k-1}}, -\sqrt{\mathbf{P}^{\mathbf{xx}}_{k|k-1}}\right], \tag{16}$$

$$\mathcal{W}^{0:2n_x}_{k|k} = \mathcal{W}^{0:2n_x}_{k|k-1} = \tfrac{1}{n_x+\kappa}[\kappa, \tfrac{1}{2}, \dots, \tfrac{1}{2}], \tag{17}$$

and

$$\mathcal{X}^i_{k+1|k} = \mathbf{f}_k(\mathcal{X}^i_{k|k}), \tag{18}$$

$$\mathcal{Z}^i_{k|k-1} = \mathbf{h}_k(\mathcal{X}^i_{k|k-1}), i = 0, 1, \dots, (2n_x + 1). \tag{19}$$

The used notation is $\mathcal{X}^{a:b} = [\mathcal{X}^a, \mathcal{X}^{a+1}, \dots, \mathcal{X}^b]$ and $\mathbf{1}_{a\times b}$ and $\mathbf{0}_{a\times b}$ are matrices of ones and zeros, respectively. The constant c is equal to $\sqrt{n_x + \kappa}$, where κ is the scaling parameter affecting the estimation performance of the UKF. The parameter can be set either constant or time-varying [23].

C. Nonlinearity and non-Gaussianity measures

The LFs provide reliable results if the point at which the functions or around which PDFs are approximated is close to the working point, i.e., the actual state. This is usually true if the nonlinearities in system description are mild, PDFs of the state and measurement noises are Gaussian and the filter and system initial conditions are Gaussian and close to each other. Then, the true conditional PDF of the state estimate is expected to be close to the Gaussian PDF [8] (unimodal and not heavy-tailed) and computationally efficient LFs can be used instead of much more demanding global filters.

Assessment of the system nonlinearity or the resulting state PDF properties is, therefore, a vital part of any filter design procedure and several nonlinearity or non-Gaussian measures[5] have been proposed so far.

The global measures asses the overall nonlinearity or non-Gaussianity of the given system without any reference to the usage of a specific filter. Basically, the measure gives an answer whether a LF is expected to perform well or a global filter needs to be used. The global nonlinearity measure proposed in [14] measures the mean-square distance (in a functional space) between the given nonlinear system and a set of linear systems. The global NGM assesses the impact of the nonlinear transformation of the resulting PDF and subsequently the closeness of the (unconditional) PDF of the state to an ideal Gaussian one [17]. As an example, the NGMs based on the third or fourth moments of the transformed PDF can be mentioned.

The local nonlinearity and non-Gaussianity measures provide information about

- severity of the nonlinear functions $\mathbf{f}_k(\cdot)$ and $\mathbf{h}_k(\cdot)$ at a given approximation point specified by the predictive state estimate and the filtering estimate and
- impact of the nonlinear transformation on the distribution of the predictive state and measurement estimates,

respectively, at a single time instant assuming the previous (to-be-transformed) filtering and predictive PDFs being Gaussian. From the functional point of view, the particular measures are analogous to the global ones, i.e., NLMs assess for example closeness of the true nonlinear function to its best linear approximation and NGMs compare the higher order moment of the transformed PDF with the expect ones computed on the basis of the Gaussian distribution. The local measures provide information predominantly valid at a given time only without any regard to previous time instants; thus, the measure might be denoted as instant. Such a measure might therefore fail in detection of a sequence of mild nonlinearities resulting in the end in e.g., heavy-tailed PDF, as illustrated in [17].

The paper extends the instant local NGMs to the cumulative ones with the ability to monitor time behaviour of the measure taking into account the impact of a multidimensional nonlinear transformation in the filtering and prediction steps. That means the measure at time k depends on the measure from the previous time $k-1$. The NGMs considered in this paper are based on the third and fourth-order moments.

III. Recursive Computation of Higher-Order Moments

LFs recursively compute the conditional mean and covariance matrix of the state. To compute also the higher-order conditional moments, the respective recursive formulas for filtering and predictive steps have to be proposed. In this section, general recursive relations are provided and then detailed for the UKF.

A. Higher-order moment parametrization

In this paper, the following parametrization of the n-th moment of a vector random variable $\mathbf{x}$, denoted as $\mathrm{M}^{\overbrace{\mathbf{xx}\dots\mathbf{x}}^{n \text{ terms}}}$, is adopted

$$\mathrm{M}^{\overbrace{\mathbf{xx}\dots\mathbf{x}}^{n \text{ terms}}} \triangleq \mathsf{E}[(\mathbf{x}-\hat{\mathbf{x}})(\mathbf{x}-\hat{\mathbf{x}})^T \underbrace{\otimes(\mathbf{x}-\hat{\mathbf{x}})^T \otimes (\mathbf{x}-\hat{\mathbf{x}})^T \dots \otimes (\mathbf{x}-\hat{\mathbf{x}})^T}_{(n-2) \text{ terms}}], \tag{20}$$

where $\hat{\mathbf{x}} = \mathsf{E}[\mathbf{x}]$ and the symbol $\otimes$ stands for the Kronecker product [24]. Alternative parametrizations can be found e.g., in [25].

[5]Nonlinearity can be viewed as a reason for non-Gaussian PDFs of the state estimate. Therefore, the NLMs assess the cause of the non-Gaussianity and the NGMs assess effect of the nonlinearity.

B. Recursive computation of third-order moment

Following the notation (20), the filtering third-order moment is defined as

$$\mathbf{M}_{k|k}^{\mathbf{xxx}} \triangleq \mathsf{E}[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k})(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k})^T \otimes (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k})^T]. \quad (21)$$

Substituting (3) into (21) and defining the error variables $\tilde{\mathbf{x}}_{k|k-1} = \mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}$ and $\tilde{\boldsymbol{\zeta}} = \mathbf{K}_k \tilde{\mathbf{z}}_{k|k-1} = \mathbf{K}_k(\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1})$, equation (21) can be further treated as (omitting time indices):

$$\begin{aligned}\mathbf{M}_{k|k}^{\mathbf{xxx}} &= \mathsf{E}[(\tilde{\mathbf{x}} - \tilde{\boldsymbol{\zeta}})(\tilde{\mathbf{x}} - \tilde{\boldsymbol{\zeta}})^T \otimes (\tilde{\mathbf{x}} - \tilde{\boldsymbol{\zeta}})^T] \quad (22)\\ &= \mathsf{E}[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T \otimes \tilde{\mathbf{x}}^T - \tilde{\boldsymbol{\zeta}}\tilde{\mathbf{x}}^T \otimes \tilde{\mathbf{x}}^T - \tilde{\mathbf{x}}\tilde{\boldsymbol{\zeta}}^T \otimes \tilde{\mathbf{x}}^T + \tilde{\boldsymbol{\zeta}}\tilde{\boldsymbol{\zeta}}^T \otimes \tilde{\mathbf{x}}^T\\ &\quad - \tilde{\mathbf{x}}\tilde{\mathbf{x}}^T \otimes \tilde{\boldsymbol{\zeta}}^T + \tilde{\boldsymbol{\zeta}}\tilde{\mathbf{x}}^T \otimes \tilde{\boldsymbol{\zeta}}^T + \tilde{\mathbf{x}}\tilde{\boldsymbol{\zeta}}^T \otimes \tilde{\boldsymbol{\zeta}}^T - \tilde{\boldsymbol{\zeta}}\tilde{\boldsymbol{\zeta}}^T \otimes \tilde{\boldsymbol{\zeta}}^T].\end{aligned}$$

Respecting the notation for a third-order moment of variables $\tilde{\mathbf{x}}_{k|k-1}$ and $\tilde{\boldsymbol{\zeta}}_{k|k-1}$ of the form

$$\mathbf{M}_{k|k-1}^{\boldsymbol{\zeta}\mathbf{xx}} \triangleq \mathsf{E}[\boldsymbol{\zeta}_{k|k-1}\tilde{\mathbf{x}}_{k|k-1}^T \otimes \tilde{\mathbf{x}}_{k|k-1}^T], \quad (23)$$

the final form of the *filtering third moment* (22) is

$$\begin{aligned}\mathbf{M}_{k|k}^{\mathbf{xxx}} &= \mathbf{M}_{k|k-1}^{\mathbf{xxx}} - \mathbf{M}_{k|k-1}^{\boldsymbol{\zeta}\mathbf{xx}} - \mathbf{M}_{k|k-1}^{\mathbf{x}\boldsymbol{\zeta}\mathbf{x}} + \mathbf{M}_{k|k-1}^{\boldsymbol{\zeta\zeta}\mathbf{x}} - \mathbf{M}_{k|k-1}^{\mathbf{xx}\boldsymbol{\zeta}}\\ &\quad + \mathbf{M}_{k|k-1}^{\boldsymbol{\zeta}\mathbf{x}\boldsymbol{\zeta}} + \mathbf{M}_{k|k-1}^{\mathbf{x}\boldsymbol{\zeta\zeta}} - \mathbf{M}_{k|k-1}^{\boldsymbol{\zeta\zeta\zeta}}. \quad (24)\end{aligned}$$

The *predictive third moment* of the state is defined and is given as

$$\begin{aligned}\mathbf{M}_{k+1|k}^{\mathbf{xxx}} &\triangleq \mathsf{E}[(\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k})(\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k})^T\\ &\quad \otimes (\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k})^T]. \quad (25)\end{aligned}$$

The relation between the predictive $\mathbf{M}_{k+1|k}^{\mathbf{xxx}}$ and filtering $\mathbf{M}_{k|k}^{\mathbf{xxx}}$ cannot be further specified for a general LF as it depends on the particular approximation.

Following the UKF concept, approximation of a third order moment can be computed according to the relation

$$\mathbf{M}_{k|k-1}^{\boldsymbol{\zeta}\mathbf{xx}} = \sum_{i=0}^{2n_x} \mathcal{W}_{k|k-1}^i \tilde{\phi}_{k|k-1}^i (\tilde{\mathcal{X}}_{k|k-1}^i)^T \otimes (\tilde{\mathcal{X}}_{k|k-1}^i)^T, \quad (26)$$

where

$$\begin{aligned}\tilde{\phi}_{k|k-1}^i &= \mathbf{K}_k \tilde{\mathcal{Z}}_{k|k-1}^i = \mathbf{K}_k(\mathcal{Z}_{k|k-1}^i - \hat{\mathbf{z}}_{k|k-1}), \quad (27)\\ \tilde{\mathcal{X}}_{k|k-1}^i &= \mathcal{X}_{k|k-1}^i - \hat{\mathbf{x}}_{k|k-1}. \quad (28)\end{aligned}$$

Here, it must be noted that the third (and also higher) order moment depends on the actual measurement indirectly only, i.e., via the linearisation or approximation point given by the estimated state mean (similarly to the covariance matrix), and thus, it is rather a function of the system model.

C. Recursive computation of fourth-order moment

The *filtering fourth moment* is defined analogously to (21) as

$$\begin{aligned}\mathbf{M}_{k|k}^{\mathbf{xxxx}} &= \mathbf{M}_{k|k-1}^{\mathbf{xxxx}} - \mathbf{M}_{k|k-1}^{\mathbf{x}\boldsymbol{\zeta}\mathbf{xx}} - \mathbf{M}_{k|k-1}^{\boldsymbol{\zeta}\mathbf{xxx}} + \mathbf{M}_{k|k-1}^{\boldsymbol{\zeta\zeta}\mathbf{xx}}\\ &\quad - \mathbf{M}_{k|k-1}^{\mathbf{xx}\boldsymbol{\zeta}\mathbf{x}} + \mathbf{M}_{k|k-1}^{\mathbf{x}\boldsymbol{\zeta\zeta}\mathbf{x}} + \mathbf{M}_{k|k-1}^{\boldsymbol{\zeta}\mathbf{x}\boldsymbol{\zeta}\mathbf{x}} - \mathbf{M}_{k|k-1}^{\boldsymbol{\zeta\zeta\zeta}\mathbf{x}}\\ &\quad - \mathbf{M}_{k|k-1}^{\mathbf{xxx}\boldsymbol{\zeta}} + \mathbf{M}_{k|k-1}^{\mathbf{x}\boldsymbol{\zeta}\mathbf{x}\boldsymbol{\zeta}} + \mathbf{M}_{k|k-1}^{\boldsymbol{\zeta}\mathbf{xx}\boldsymbol{\zeta}} - \mathbf{M}_{k|k-1}^{\boldsymbol{\zeta\zeta}\mathbf{x}\boldsymbol{\zeta}}\\ &\quad + \mathbf{M}_{k|k-1}^{\mathbf{xx}\boldsymbol{\zeta\zeta}} - \mathbf{M}_{k|k-1}^{\mathbf{x}\boldsymbol{\zeta\zeta\zeta}} - \mathbf{M}_{k|k-1}^{\boldsymbol{\zeta}\mathbf{x}\boldsymbol{\zeta\zeta}} + \mathbf{M}_{k|k-1}^{\boldsymbol{\zeta\zeta\zeta\zeta}}. \quad (29)\end{aligned}$$

The *predictive fourth moment* of the state is

$$\begin{aligned}\mathbf{M}_{k+1|k}^{\mathbf{xxxx}} &\triangleq \mathsf{E}[(\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k})(\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k})^T\\ &\quad \otimes (\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k})^T \otimes (\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k})^T]. \quad (30)\end{aligned}$$

The UT-based approximation of a fourth order moment computation, considering (27) and (28), is computed as

$$\begin{aligned}\mathbf{M}_{k|k-1}^{\boldsymbol{\zeta}\mathbf{xx}\boldsymbol{\zeta}} &= \sum_{i=0}^{2n_x} \mathcal{W}_{k|k-1}^i \tilde{\phi}_{k|k-1}^i (\tilde{\mathcal{X}}_{k|k-1}^i)^T\\ &\quad \otimes (\tilde{\mathcal{X}}_{k|k-1}^i)^T \otimes (\tilde{\phi}_{k|k-1}^i)^T. \quad (31)\end{aligned}$$

IV. Non-Gaussianity Measures in Local Filters

Non-Gaussianity measures are based on a comparison of the higher-order moments computed according to (24), (25), (29), and (30) with the expected moments computed on the basis of the corresponding covariance matrices (4) and (9) assuming a Gaussian distribution[6] [17].

A. Non-Gaussianity measure and threshold

Essentially, the NGMs are based on a moment-based comparison of two PDFs, the estimated one and its Gaussian approximation.

In particular, the measures described in this section compare only some elements of the true moment of the state vector and its expected value based on the Gaussian approximation [7] to reduce the computational complexity.

As the third moment of a Gaussian distribution is zero, the NGMs related to the third moment of n-th element of the state is given by

$$J_3(x_n) = \mathbf{M}_{k|k}^{\mathbf{xxx}}\left(n, (n-1)n_x + n\right), n = 1, 2, \ldots, n_x, \quad (32)$$

where $\mathbf{M}_{k|k}^{\mathbf{xxx}}(i,j)$ is the element of the matrix $\mathbf{M}_{k|k}^{\mathbf{xxx}}$ in i-th row and j-th column.

The NGMs related to the fourth moment of n-th state element is equal to

$$\begin{aligned}J_4(x_n) &= \mathbf{M}_{k|k}^{\mathbf{xxxx}}\left(n, (n-1)(n_x^2 + n_x) + n\right)\\ &\quad - 3\left(\mathbf{P}_{k|k}^{\mathbf{xx}}(n,n)\right)^2, n = 1, 2, \ldots, n_x, \quad (33)\end{aligned}$$

where $\mathbf{M}_{k|k}^{\mathbf{xxxx}}(i,j)$ is the element of the matrix $\mathbf{M}_{k|k}^{\mathbf{xxxx}}$ and $\mathbf{P}_{k|k}^{\mathbf{xx}}(n,n)$ is the filtering covariance matrix element in n-th row and column.

Both measures might have any real value, but, from the standpoint of the closeness to a Gaussian PDF, the closer to zero the better. If only one measure is preferred, then the

[6] All odd higher moments of a (scalar) Gaussian random variable are zero and all even moments are function of its variance. In fact $2k$-th moment of a zero-mean random variable x with a variance P_x is equal to $M^{\overbrace{xx\ldots x}^{2k \text{ terms}}} = E[x^{2k}] = (P^{xx})^k(2k-1)!!$ where the symbol "!!" stands for a double factorial [26].

[7] Considering the given parametrization of higher moments, it can be seen that the n-th moment is described by a matrix of the dimension $n_x \times (n_x)^{n-1}$; thus, the number of elements of the true and expected moments, which can be compared, grows exponentially. To reduce the costs of moments computation and comparison for higher dimensions n_x, the mixed elements may be discarded and only n_x elements are compared.

absolute values of e.g., $J_3(x_n)$ might be summed over n forming an overall measure.

The value of the measures is then compared with thresholds $J_{3,thr}$ or $J_{4,thr}$ and inequalities $J_3(\cdot) \geq J_{3,thr}$ or $J_4(\cdot) \geq J_{4,thr}$ indicate that the corresponding estimated PDF is far from being Gaussian. Thus, in this case the estimate quality is possibly low. The threshold values need to be specified on the basis of an off-line analysis or a simulation study and might be different for each state component.

B. LF algorithm with NGMs

Having stated the relations necessary for the recursive state higher-order moment computation and the NGMs, the algorithm of the LF with self-assessment can be summarized.

Algorithm 2: Local Filter with NGM-based Self-Assessment

Step 1: (**initialization**) as **Step 1** of Algorithm 1, supplemented with specification of an initial higher moments estimate.
Step 2: (**filtering**) as **Step 2** of Algorithm 1, supplemented with computation of $\mathrm{M}^{\mathbf{xxx}}_{k|k}$ according to (24) and $\mathrm{M}^{\mathbf{xxxx}}_{k|k}$ according to (29).
Step 3: (**self-assessment**) If any measure $J_3(x_n), n = 1, 2, \ldots, n_x$, of the filtering estimate is greater than or equal to the respective threshold $J_{3,thr}$, then there is a possibility of a lower estimate quality. The same holds for $J_4(\cdot)$.
Step 4: (**prediction**) as **Step 3** of Algorithm 1, supplemented with computation of $\mathrm{M}^{\mathbf{xxx}}_{k+1|k}$ according to (25) and $\mathrm{M}^{\mathbf{xxxx}}_{k+1|k}$ according to (30).
Step 5: (**self-assessment**) If any measure $J_3(x_n), n = 1, 2, \ldots, n_x$, of the predictive estimate is greater than or equal to the respective threshold $J_{3,thr}$, then there is a possibility of a lower estimate quality. The same holds for $J_4(\cdot)$.
Let $k = k + 1$ and algorithm continues by **Step 2**.

C. Notes and discussion

Note 1: The filtering higher order moments, i.e., (24) and (29), explicitly depend on the predictive moments, i.e., on (25) and (30). Conversely, the predictive moments can be virtually computed independently of the filtering moments at the previous time instant. The only way how to transfer the filtering moments to the predictive moments is to create the filtering σ-point set (given by (15) and (17)) which in addition to the usual first two moments (3) and (4) is able to capture arbitrary higher-order moments (24) and (29).

Note 2: In literature a rather minor attention has been paid to the creation of a σ-point set having the same moments (up to a certain order greater than two) as the approximated random variable. As an example the higher order unscented transform (HOUT) [27] or the conjugate unscented transform [28] can be mentioned which are based on an extended σ-point set. However, the HOUT was designed to capture the even higher order moments only (symmetric distribution assumed). In [29], the reduced σ-point set was proposed (with $n_x + 1$ σ-points) being able to capture also the third order moment. Nevertheless, for a multidimensional state the third moment cannot be chosen arbitrarily as a design parameter; it rather comes out nonzero as an effect of the sigma points specification. The motivation behind providing a reduced σ-point set was reduction of the computational costs. Thus, in the appendix two techniques are developed that might be used for computation of the σ-point set with the given mean, covariance, and higher order moments.

Note 3: In the EKF framework, the predictive moments explicitly depend on the filtering ones [18]. This is because of the different nature of the EKF (the EKF approximates the nonlinear functions, whereas the UKF approximates the distribution of the state estimate).

V. Numerical Illustration

The UKF with the recursive computation of the higher order moments is illustrated by a tracking example with the (discretized) constant turn motion model in the horizontal plane [30]

$$\begin{aligned}\mathbf{x}_{k+1} &= \left[x_{k+1}, y_{k+1}, v_{x,k+1}, v_{y,k+1}, \omega_{k+1}\right]^T \\ &= \mathbf{f}_k(\mathbf{x}_k) + \mathbf{G}_k \mathbf{w}_k,\end{aligned} \tag{34}$$

where the state vector is composed of the position and velocity in x- and y-directions, i.e., x, y, v_x, and v_y, and the turning rate (or angular velocity) ω. The function is defined as

$$\mathbf{f}_k(\mathbf{x}_k) = \begin{bmatrix} 1 & 0 & (\sin(\omega_k)T))/\omega_k & -(1-\cos(\omega_k T))/\omega_k & 0 \\ 0 & 1 & (1-\cos(\omega_k T))/\omega_k & (\sin(\omega_k T))/\omega_k & 0 \\ 0 & 0 & \cos(\omega_k T) & -\sin(\omega_k T) & 0 \\ 0 & 0 & \sin(\omega_k T) & \cos(\omega_k T) & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{35}$$

with the sampling period $T = 0.1$. The state noise is zero mean with the covariance matrix

$$\mathbf{\Sigma}^{\mathbf{w}}_k = \mathrm{diag}([1, 1, 0.2]), \tag{36}$$

where the function $\mathrm{diag}(\mathbf{y})$ forms a diagonal matrix with a vector $\mathbf{y}$ on diagonal, and the noise gain matrix is given by

$$\mathbf{G}_k = \begin{bmatrix} T^2/3 & 0 & 0 \\ 0 & T^2/3 & 0 \\ T & 0 & 0 \\ 0 & T & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{37}$$

The position is observed indirectly by means of the range and bearing measurements

$$\mathbf{z}_k = \left[\sqrt{x_k^2 + y_k^2}, \mathrm{atan}\left(\frac{y_k}{y_k}\right)\right]^T + \mathbf{v}_k, \tag{38}$$

where the measurement noise $\mathbf{v}_k$ is zero mean with the covariance matrix

$$\mathbf{\Sigma}^{\mathbf{v}}_k = \mathrm{diag}([1, 10^{-6}]) \tag{39}$$

The system and filter initial conditions are considered to be

$$\hat{\mathbf{x}}_0 = \hat{\mathbf{x}}_{0|-1} = [3000, 2000, -200, 0, 0.01]^T, \tag{40}$$
$$\mathbf{P}^{\mathbf{xx}}_0 = \mathbf{P}^{\mathbf{xx}}_{0|-1} = \mathrm{diag}([1, 1, 10^{-2}, 10^{-2}, 10^{-4}]) \tag{41}$$

and the scenario is simulated for $k = 0, 1, \ldots, K$ with $K = 80$. The true turning rate ω_k is assumed to be $\omega_k = 0.01, \forall k$, except of the time intervals $k \in \mathcal{K}_1 = \langle 10, 15\rangle$ with $\omega_k = 1$, $k \in \mathcal{K}_2 = \langle 16, 20\rangle$ with $\omega_k = -1$, and $k \in \mathcal{K}_3 = \langle 60, 62\rangle$ with $\omega_k = -1$.

In Figure 1 the estimated characteristics of the UKF with higher order moments recursive computation averaged over 10^4 Monte-Carlo (MC) simulations are shown. Namely, the following characteristics are plotted; the mean square errors (MSE), averaged variances of estimate error calculated by the filter, and averaged third order moments of the position and velocity in one direction and the turn-rate. For completeness, the time intervals with significant change of the true turn-rate are in colour, i.e., blue, red and green colours illustrate the first interval $\mathcal{K}_1$, the second interval $\mathcal{K}_2$, and the third one $\mathcal{K}_3$, respectively.

It can be seen that the turn-rate change has a significant impact on the estimation quality characterized by the MSE. Indeed, once the true turn-rate is changed then the filter requires a certain time period to estimate the new turn rate. In this period, the estimate quality of the turn-rate and subsequently of the velocity and position is worsened. The turn-rate changes affect (with a short delay given by the structure of the model) also the variances and the third order moments in the respective time intervals. In both these quantities the change of the turn-rate are somehow apparent. However, as can be seen, the turn-rate change impacts much severely the third-order moment (especially the moment corresponding to the velocity). In the time behaviour of the moment the intervals with the turn-rate change are clearly visible and can be detected e.g., by setting of the appropriate threshold. Note that the fourth order moment (or the respective NGM (32)) is not plotted in Figure 1 as it is not as illustrative as the third order moment in this case.

In Figure 2, the estimated variance and third order moment for position and velocity for a single run are plotted. Each quantity (e.g., third moment of the position) is represented by two curves; blue, showing the actual time behaviour which is just scaled to be easily visible in the plot, red, showing accumulated absolute value of the quantity. For example, the accumulated third order moment of the position is given by

$$M_k^{\mathbf{xxx},acc} = \sum_{i=0}^{k} |M_{i|i}^{\mathbf{xxx}}|. \quad (42)$$

The figure demonstrates that also in a single realisation the third order moment is clearly much more affected by the change of the turn-rate than the corresponding variance. Thus, the time behaviour of the third order moment (especially its accumulated version) is much more suitable for an analysis of the changes in the underlying system or for an on-line monitoring of the filter performance.

VI. Concluding Remarks

The paper dealt with local filters in state estimation of stochastic nonlinear systems with the stress on the unscented Kalman filter. The filter was supplemented with the recursive computation of the third and fourth order moments. These moments can be, in turn, used for computation of an actual non-Gaussianity measure related to the state estimate (as a consequence of nonlinear transformations). As was illustrated in the numerical example, the time behaviour of the higher order moments is tightly related to the time behaviour of the estimate error. Therefore, the higher order moments might be used as an on-line monitor of the filter indicating expected lower quality of the estimate which is not necessarily reflected by the estimated covariance matrix. Additionally, the information on higher order moments can be used to compute an approximation of the credible interval that is less conservative than that based on the Chebyshev inequality. Furthermore, in the paper a techniques for computation of the σ-point set capturing not only the first two moments of a scalar random variable but also the higher ones were discussed and proposed.

Acknowledgement

This work was supported by the Czech Science Foundation, project GACR P103-13-07058J.

References

[1] F. L. Lewis, *Optimal Estimation*. New York: John Wiley & Sons, 1986.

[2] H. W. Sorenson, "On the development of practical nonlinear filters," *Information Sciences*, vol. 7, pp. 230–270, 1974.

[3] I. Arasaratnam and S. Haykin, "Cubature Kalman filters," *IEEE Transactions on Automatic Control*, vol. 54, no. 6, pp. 1254–1269, 2009.

[4] M. Simandl and J. Dunik, "Derivative-free estimation methods: New results and performance analysis," *Automatica*, vol. 45, no. 7, pp. 1749–1757, 2009.

[5] M. Simandl, J. Kralovec, and T. Soderstrom, "Anticipative grid design in point-mass approach to nonlinear state estimation," *IEEE Transactions on Automatic Control*, vol. 47, no. 4, 2002.

[6] H. W. Sorenson and D. L. Alspach, "Recursive Bayesian estimation using gaussian sums," *Automatica*, vol. 7, pp. 465–479, 1971.

[7] A. Doucet, N. De Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*. Springer, 2001, (Ed. Doucet A., de Freitas N., and Gordon N.).

[8] K. Reif, S. Günther, E. Yaz, and R. Unbehauen, "Stochastic stability of the discrete-time extended Kalman filter," *IEEE Transactions on Automatic Control*, vol. 44, no. 4, pp. 714–728, 1999.

[9] B. Jia, M. Xin, and Y. Cheng, "Sparse-grid quadrature nonlinear filtering," *Automatica*, vol. 48, no. 2, pp. 327–341, 2012.

[10] ——, "High-degree cubature Kalman filter," *Automatica*, vol. 49, no. 2, pp. 510–518, 2013.

[11] P. X. K. Song, "Monte Carlo Kalman filter and smoothing for multivariate discrete state space models," *Canadian Journal of Statistics*, vol. 28, no. 3, pp. 641–652, 2000.

[12] J. Duník, O. Straka, and M. Šimandl, "Stochastic integration filter," *IEEE Transactions on Automatic Control*, vol. 58, no. 6, pp. 1561–1566, 2013.

[13] J. Sarmavuori and S. Sarkka, "Fourier-Hermite Kalman filter," *IEEE Transactions on Automatic Control*, vol. 57, no. 6, pp. 1511–1515, 2012.

[14] X. R. Li, "Measure of nonlinearity for stochastic systems," in *Proceedings of the 15th International Conference on Information Fusion*, Singapore, 2012.

[15] O. Straka, J. Duník, and M. Šimandl, "Scaling parameter in unscented transformation: Analysis and specification," in *Proceedings of the 2012 American Control Conference*, Montreal, Canada, June 2012.

[16] E. Jones, M. Scalzo, A. Bubalo, M. Alford, and B. Arthur, "Measures of nonlinearity for single target tracking problem," in *Proceedings of the SPIE, Signal Processing, Sensor Fusion, and Target Recognition*, Orlando, Florida, USA, 2011.

[17] J. Dunik, O. Straka, and M. Simandl, "Nonlinearity and non-Gaussianity measures for stochastic dynamic systems," in *Proceedings of the 16th International Conference on Information Fusion*, Istanbul, 2013.

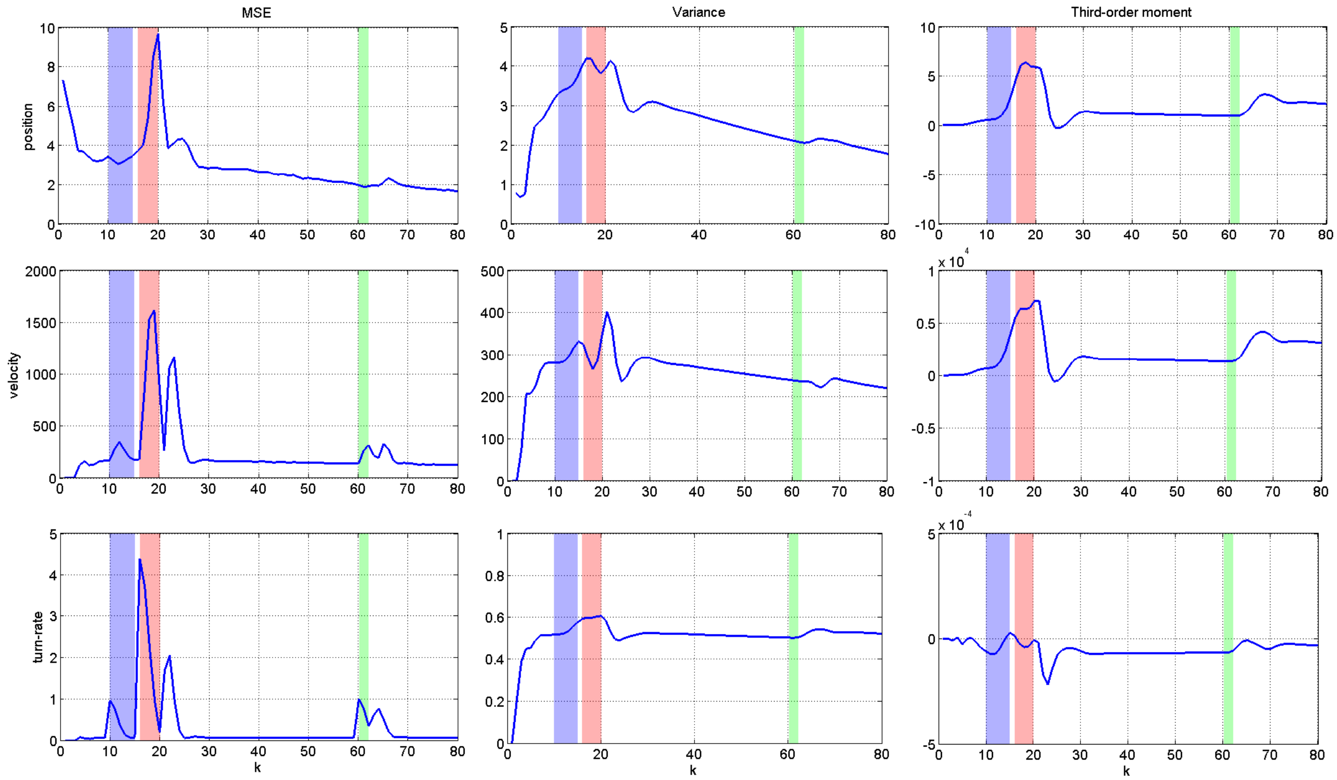

Figure 1. Averaged statistics over a set of MC simulation with highlighted time intervals with changed turn-rate.

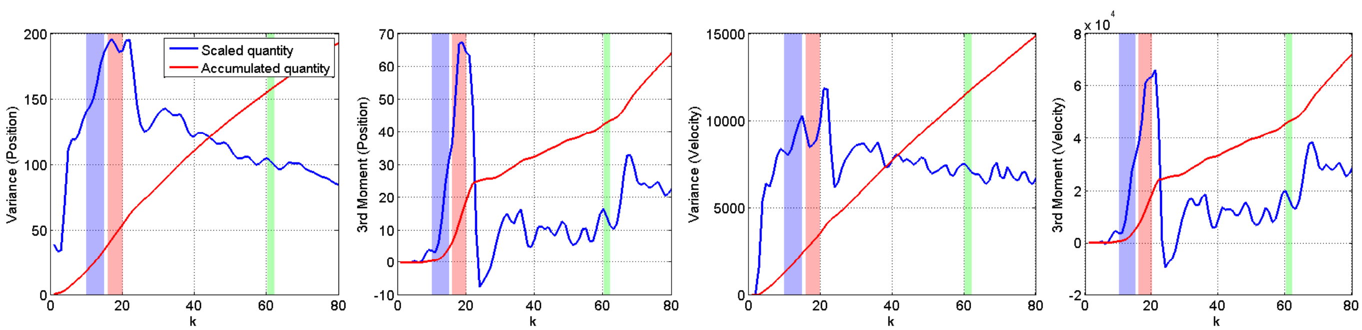

Figure 2. Statistics per one simulation with highlighted time intervals with changed turn-rate.

[18] ——, "Self-assessment of local filters by non-Gaussianity measures," in *accepted to American Control Conference*, 2014.

[19] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. New Jersey: Prentice Hall Ins.: Englewood Cliffs, 1979.

[20] M. Norgaard, N. K. Poulsen, and O. Ravn, "New developments in state estimation for nonlinear systems," *Automatica*, vol. 36, no. 11, pp. 1627–1638, 2000.

[21] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *IEEE Review*, vol. 92, no. 3, pp. 401–421, 2004.

[22] K. Ito and K. Xiong, "Gaussian filters for nonlinear filtering problems," *IEEE Transactions on Automatic Control*, vol. 45, no. 5, pp. 910–927, 2000.

[23] J. Duník, M. Šimandl, and O. Straka, "Unscented Kalman filter: Aspects and adaptive setting of scaling parameter," *IEEE Transactions on Automatic Control*, vol. 57, no. 9, pp. 2411–2416, 2012.

[24] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Transactions on Circuits and Systems*, vol. 25, no. 9, pp. 772–781, 1978.

[25] T. Kollo and D. Von Rosen, *Advanced multivariate statistics with matrices, volume 579 of Mathematics and Its Applications*. Springer, 2005.

[26] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, fourth edition ed. Mc Graw Hill, 2002.

[27] D. Tenne and T. Singh, "The higher order unscented filter," in *Proceedings of the American Control Conference, 2003*, vol. 3. IEEE, 2003, pp. 2441–2446.

[28] N. Adhurthi, P. Singla and T. Singh, "The Conjugate Unscented Transform - An Approach to Evaluate Multi-Dimensional Expectation Integrals," in *Proceedings of the American Control Conference, 2012* Montral, Canada, 2012, pp. 5556 – 5561.

[29] S. J. Julier and J. K. Uhlmann, "Reduced sigma point filters for the propagation of means and covariances through nonlinear transformations," in *Proceedings of the IEEE American Control Conference*. Anchorage AK, USA: IEEE, 2002, pp. 887–892.

[30] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Artech House, 1999.

Appendix

As was reasoned in Section IV-C, to preserve the higher order moments in the run of the UKF, the set of σ-points generated in prediction step should capture not only the mean and covariance matrix of the filtering estimated but also the higher order moments which are recursively estimated. In this appendix the attention is laid on computation of such σ-point sets and two approaches are considered here; direct and indirect approach.

A. Higher order σ-point set: Direct approach

The direct approach follows the idea used in [27], [29] and the computation of the σ-point set is based on the solution to a system of equations. Considering a scalar filtering estimate in the form of the mean $\hat{x}_{k|k}$, variance $P^{\mathbf{xx}}_{k|k}$, and the third order moment $M^{\mathbf{xxx}}_{k|k}$, the weighted σ-point set $\{\mathcal{W}_i, \mathcal{X}_i\}_{i=0}^2$ having the same moments can be computed by the solution to the following system of equations (omitting the time indices)

$$\mathcal{W}_0 + \mathcal{W}_1 + \mathcal{W}_2 = 0, \tag{43}$$

$$\mathcal{W}_0\mathcal{X}_0 + \mathcal{W}_1\mathcal{X}_1 + \mathcal{W}_2\mathcal{X}_2 = \hat{x}, \tag{44}$$

$$\mathcal{W}_0\mathcal{X}_0^2 + \mathcal{W}_1\mathcal{X}_1^2 + \mathcal{W}_2\mathcal{X}_2^2 = C^{\mathbf{xx}} = P^{\mathbf{xx}} + \hat{x}^2, \tag{45}$$

$$\mathcal{W}_0\mathcal{X}_0^3 + \mathcal{W}_1\mathcal{X}_1^3 + \mathcal{W}_2\mathcal{X}_2^3 = C^{\mathbf{xxx}}, \tag{46}$$

where the third order non-central moment is $C^{\mathbf{xxx}} = M^{\mathbf{xxx}} + 3\hat{x}C^{\mathbf{xx}} + 2\hat{x}^3$ [26]. Imposing two additional conditions to reduce the number of independent variables, e.g., $\mathcal{X}_1 = -\mathcal{X}_2$ and $\mathcal{W}_1 = \mathcal{W}_2$, the solution to (43)–(46) is

$$\mathcal{X}_0 = \sqrt{C^{\mathbf{xxx}}/\hat{x}},\ \mathcal{W}_0 = \hat{x}/\mathcal{X}_0, \tag{47}$$

$$\mathcal{X}_1 = -\mathcal{X}_2 = \sqrt{\frac{P^{\mathbf{xx}} - 2C^{\mathbf{xx}}\mathcal{W}_1}{2(1-2\mathcal{W}_1)\mathcal{W}_1}}, \tag{48}$$

$$\mathcal{W}_1 = \mathcal{W}_2 = (1-\mathcal{W}_0)/2, \tag{49}$$

which forms the σ-point set with required first three moments.

B. Higher order σ-point set: Indirect approach

The second approach proposed is based on the following idea: having first few moments it is possible to find a PDF in the form of the Gaussian mixture (GM) with the same moments which can be afterwards used as a base for standard σ-point set generation (the σ-points are generated independently for each term of the mixture).

Let the filtering estimate defined in the previous part be considered and a weighted GM with two terms, i.e.,

$$p_{GM} = \alpha_1\mathcal{N}\{x : \hat{x}_1, P_1^{\mathbf{xx}}\} + \alpha_2\mathcal{N}\{x : \hat{x}_2, P_2^{\mathbf{xx}}\}, \tag{50}$$

be assumed. The parameters of the GM can be found by solution to the following system of equations

$$\alpha_1 + \alpha_2 = 1, \tag{51}$$

$$\alpha_1\hat{x}_1 + \alpha_2\hat{x}_1 = \hat{x}, \tag{52}$$

$$\sum\nolimits_{i=1}^2 \alpha_i\left((\hat{x}_i - \hat{x})^2 + P_i^{\mathbf{xx}}\right) = P^{\mathbf{xx}}, \tag{53}$$

$$\sum\nolimits_{i=1}^2 \alpha_1\left((\hat{x}_i - \hat{x})^3 + 3(\hat{x}_i - \hat{x})P_i^{\mathbf{xx}} + M_i^{\mathbf{xxx}}\right) = M^{\mathbf{xxx}}. \tag{54}$$

As the Gaussian terms have zero third order moment, it is necessary to find 6 unknown parameters from four equation. To reduce this ambiguity, two additional constraints are set, namely $\hat{x}_2 = \hat{x}_1 + c$ and $\alpha_2 = r\alpha_1$, where c, r are a priori specified parameters. Then, the particular weights and the means are

$$\alpha_1 = 1/(1+r),\ \alpha_2 = r/(1+r), \tag{55}$$

$$\hat{x}_1 = \frac{\hat{x} - r\alpha_1 c}{\alpha_1(1+r)},\ \hat{x}_2 = \frac{\hat{x} + \alpha_1 c}{\alpha_1(1+r)}. \tag{56}$$

The variances of the particular terms are computed by solving the following system of equations for $P_1^{\mathbf{xx}}$ and $P_2^{\mathbf{xx}}$

$$\begin{bmatrix} 1 & r \\ 3(\hat{x}_1 - \hat{x}) & 3r(\hat{x}_1 - \hat{x} + c) \end{bmatrix}\begin{bmatrix} P_1 \\ P_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \tag{57}$$

with $b_1 = \left(P^{\mathbf{xx}} - \sum_{i=1}^2 \alpha_i(\hat{x}_i - \hat{x})^2\right)/\alpha_1$ and $b_2 = \left(M^{\mathbf{xxx}} - \sum_{i=1}^2 \alpha_i(\hat{x}_i - \hat{x})^3\right)/\alpha_1$.

Having all the parameters of the GM with the required global moments, the σ-point set with the same moments is composed of two independent sets; each computed for a particular Gaussian term respecting its particular weight. It means, the first set ${}^{(1)}\mathcal{X}^{0:2}$ is computed according to (15) with $\hat{x}_1$ and $P_1^{\mathbf{xx}}$ and with weights ${}^{(1)}\mathcal{W}^{0:2} = \alpha_1\mathcal{W}^{0:2}$, where $\mathcal{W}^{0:2}$ is given by (17). The second weighted set, i.e., ${}^{(2)}\mathcal{X}^{0:2}$, ${}^{(2)}\mathcal{W}^{0:2}$, is computed analogously. Both sets then form the final σ-point set and, as it is not difficult to show, the final set has the required overall mixture moments.

C. Higher order σ-point set: Discussion

As can be seen, both approaches offer a possibility to design σ-point sets with required moments (not only the first two). In fact, infinitely many different sets can be constructed with the same moments.

The direct approach takes an advantage of lower number of σ-points. On the other hand, its disadvantage can be found in a quite tedious solution to the system of equations which might not lead to a closed-form relations in the end. Also, the final relation might not be solvable for certain combinations of the moments and may lead to complex σ-points.

The second indirect approach works with extended σ-point set of which cardinality depends on the number of terms in the GM. The number of terms and the parameters c and r are user-defined. The number of the terms might be the lowest number that enables to fully capture the required moments. The parameter c affects the placement of the GM terms in the state space, therefore, it should be selected with respect to the overall mixture variance $P^{\mathbf{xx}}_{k|k}$. The parameter r impacts the weights of the particular terms and, thus, mainly the skewness of the GM.

Following the standard approach to σ-point set specification, which consists of making several σ-point set for a normalized distribution (based on the set for scalar variable) and combining their linear transformation, is not a viable option in this case. The reason is that the value of a third order moment cannot be determined by a simple multiplication or addition of σ-points. Therefore, for higher dimensions, the above mentioned approaches can be, in principle, used as well. In this case, each equation in the system of equations (either (43)–(46) or (51)–(54)) is in the vector or matrix form.