Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005

Performance analysis of derivative-free filters

Jindřich Duník, Miroslav Šimandl, Ondřej Straka, Ladislav Král

Abstract— **Nonlinear state estimation by the derivative-free Sigma Point Kalman Filters is treated. Particularly, impact of the derivative-free Kalman filters on estimation quality of the Sigma Point Gaussian Sum Filters is discussed. New relations between the Unscented Kalman Filter and the Divided Difference Filters are derived. The main stress is laid on the covariance matrixes which have crucial role for the behaviour explanation of the Sigma Point Gaussian Sum Filters. The theoretical results are illustrated in some numerical examples.**

I. Introduction

The general solution of the estimation problem is given by the Bayesian Recursive Relations (BRR). Exact solution of the BRR is available only for a few special cases [1], [2], e.g. for linear Gaussian system. In other cases it is necessary to apply some approximative methods. These methods can be divided into local and global methods [3].

The local methods are often based on approximation of the nonlinear functions in the state or measurement equation so that the technique of the Kalman filter can be used for the BRR solution. This approach causes that all conditional pdf's of the state estimate are given by 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 ensure the convergency of the local filter estimates. The resulting estimates of the local filters are suitable mainly for point estimates. On the other hand, the advantage of the local methods can be found in the simplicity of the BRR solution. As a suitable tool for approximation of the nonlinear functions the Taylor expansion or the Stirling's interpolation can be used, which leads to the Extended Kalman Filter (EKF) [1], the Second Order Filter (SOF) [4] or to the Divide Difference Filters (DDF's) [5], [6], respectively. Further, the DDF's can be divided into the Divide Difference Filter 1st Order (DD1) and the Divide Difference Filter 2nd order (DD2) which are based on the Stirling's interpolation 1st and 2nd order, respectively. Instead of substitution of the nonlinear functions in the system description an approximation of the pdf's representing state estimates by a set of the deterministically chosen weighted points (so called σ-points) can be utilized as a base for the local filters. The Unscented Kalman Filter (UKF) [7], [8] or the Gauss-Hermite Filter [9] exemplify this approach. It is very important that for the UKF and the DDF common features can be found although they come from quite different assumptions [5], [10], [11]. Thus these local filters are often mentioned together as the Sigma Point Kalman Filters (SPKF's) or derivative-free Kalman filters.

The global methods are based on approximation of the conditional pdf of the state estimate of some kind to accomplish better state estimates. These methods are more sophisticated but they have higher computational demands than the local methods. There are three main global approaches: analytical approach often based on the model linearization and Gaussian sum approximation of all pdf's [12], [13], which leads to e.g. the Gaussian Sum Filter (GSF), numerical approach based on the numerical solution of integrals in the BRR [14], [15], e.g. the Point-Mass Filter (PMF), and simulation approach using Monte Carlo approximation [16], [17], e.g. the Particle Filter.

Some global methods are based on multiple application of local methods. As an example of these methods the GSF can be mentioned [12]. The advancement of the local filters influenced development in the area of the global filters. In the last five years a few new global approaches have been designed by this way, e.g. the Unscented Particle Filter, the Gaussian Mixture Sigma Point Particle Filter [16] or the Sigma Point Gaussian Sum Filter (SPGSF) [8] which are grounded on the SPKF's.

On the contrary to the local methods, the main result of the global filters should be the state estimate in the form of conditional pdf of the state. Accordingly, the comparison of the global filters' performance should be primarily based on the estimated pdf of the state. Unfortunately, the new global filters, which are based on the SPKF's, have been compared from the viewpoint of the point estimates only and the quality of a posteriori pdf's has not been discussed.

This paper is focused on comparison of the global filters based on various derivative-free local filters. This comparison is based on the estimated pdf's mainly. As it will be shown, behavior of these global filters is influenced by choice of the local filters. Hence, the main stress of this paper is laid on analysis and comparison of the local filters, especially from the viewpoint of the estimated covariance matrixes.

II. Problem statement

Consider the discrete-time nonlinear non-Gaussian stochastic system

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

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

where $\mathbf{x}_k \in \mathbb{R}^{n_x}$ and $\mathbf{z}_k \in \mathbb{R}^{n_z}$ represent the immeasurable state of the system and the measurement at time instant k,

This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic, projects No. 275/2005/G1 and 1M0572.

Authors are with Department of Cybernetics, Faculty of Applied Sciences, University of West Bohemia in Pilsen, Univerzitní 8, 306 14 Plzeň, Czech Republic.
E-mail address: dunikj@kky.zcu.cz (Jindřich Duník).

TABLE I

MEAN SQUARE ERROR OF DERIVATIVE-FREE FILTERS.

	DD1	UKF	SPGSF(DD1)	SPGSF(UKF)	PMF
MSE	0.1955	0.1304	0.0222	0.0133	$3.55 \cdot 10^{-4}$
Time(s)	0.086	0.080	2.520	2.140	3170

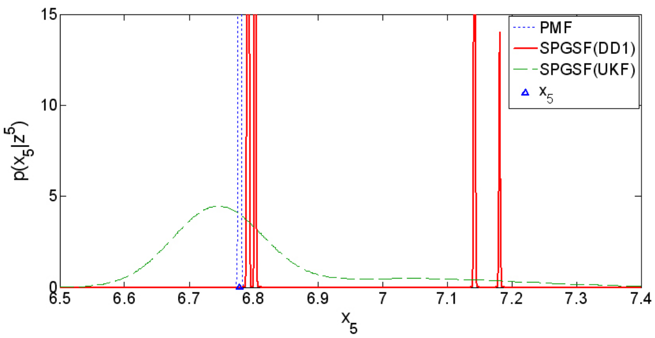

Fig. 1. Estimated filtering pdf's ($\kappa = 2$, $h^2 = 3$).

respectively, and $\mathbf{f}_k : \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$, $\mathbf{h}_k : \mathbb{R}^{n_x} \to \mathbb{R}^{n_z}$ are known vektor functions. The variables $\mathbf{w}_k \in \mathbb{R}^{n_x}$, $\mathbf{v}_k \in \mathbb{R}^{n_z}$ are state and measurement zero mean white noises. The pdf's of both noises $p_{\mathbf{w}_k}(\mathbf{w}_k)$, $p_{\mathbf{v}_k}(\mathbf{v}_k)$ are assumed to be known as well as the pdf of the initial state $p_{\mathbf{x}_0}(\mathbf{x}_0)$. The noises are mutually independent and independent of the initial state.

The aim of the filtering is to find the state estimate in the form of the conditional pdf $p(\mathbf{x}_k|\mathbf{z}^k)$ given by the Bayesian rule [1], where $\mathbf{z}^k = [\mathbf{z}_0, \ldots, \mathbf{z}_k]$.

For clear specification of the motivation for this paper, it is suitable to begin with an example. Suppose the nonlinear non-Gaussian system with one-dimensional state [16]

$$x_{k+1} = \phi_1 x_k + 1 + sin(\omega \pi k) + w_k \tag{3}$$

with the state noise w_k with Gamma pdf $Ga(3, 2)$, $\forall k$, $\phi_1 = 0.5$, $\omega = 0.04$ are scalar parameters and $k = 0, 1, \ldots, 60$. The state is observed by the scalar measurement z_k given as

$$z_k = \begin{cases} \phi_2 x_k^2 + v_k, & k \leq 30, \\ \phi_3 x_k - 2 + v_k, & k > 30. \end{cases} \tag{4}$$

The measurement z_k is influenced by the measurement noise $v_k \sim \mathcal{N}(v_k : 0, R_k)$, $R_k = 10^{-5}$, $\forall k$, and the scalar parameters are chosen as $\phi_2 = 0.2$ and $\phi_3 = 0.5$. The initial condition is given by $p(x_0) = \mathcal{N}(x_0 : 0, 12)$ and $p(x_0|z^{-1}) = p(x_0)$. An approximation of the Gamma pdf in the state equation (3) is considered in the form of the Gaussian mixture $\hat{p}(w_k) = 0.29 \times \mathcal{N}(w_k : 2.14, 0.72) + 0.18 \times \mathcal{N}(w_k : 7.45, 8.05) + 0.53 \times \mathcal{N}(w_k : 4.31, 2.29)$, for the SPGSF's and in the form of the normal distribution $\hat{p}(w_k) = \mathcal{N}(w_k : 4.26, 6.03)$ for the local filters, $\forall k$.

Accuracy of the point estimates in the form of the Mean Square Error (MSE) and the average computational demands (in seconds) of the DD1, the UKF, the SPGSF based on the DD1 (SPGSF(DD1)), the SPGSF based on the UKF (SPGSF(UKF)) and the PMF are given in Table I. Better estimation performance of the UKF and the DD2 over the DD1 is well known [5]. The estimation quality should be improved by utilization of a bank of the local filters in the Gaussian sum framework which is confirmed by results in Table I. From Table I it could be supposed the estimated pdf's of the SPGSF(DD1) and the SPGSF(UKF) are very similar. In fact the estimated pdf's, as a main result of global filters, of the SPGSF's are completely different as it is shown in Fig. 1 in the time instant $k = 5$.

It is important to mention that the nearly same shape of estimated pdf as the SPGSF(UKF) is acquired by the SPGSF based on the DD2 (SPGSF(DD2)) for this case. On the other hand, the similar estimates to the SPGSF(DD1) are obtained by the well-known GSF based on the EKF [4], [13]. The pdf given by the PMF can be considered as a "nearly true" pdf if a sufficient number of grid points is used [15]. For this example 50001 grid points for the PMF were employed.

If the time instant is $k \leq 30$, the difference between pdf's estimated by the SPGSF(DD1) and the SPGSF(UKF,DD2) is significant. For $k > 30$ the estimated pdf's are very similar for all the filters because the system (3), (4) becomes linear and therefore all local filters turned to the KF.

The objective of the previous example was to show the considerable differences between the shape of estimated pdf's by means of the global filters based on the various local filters. For that reason it is not suitable to compare the global filters only from viewpoint of the point estimates as it was done in [8], [16].

Therefore the aim of this paper is to compare the global methods exploiting local filters, namely SPGSF(DD1), SPGSF(DD2) and SPGSF(UKF). This confrontation will be based on the estimated pdf's and also on the point estimates. For the analysis of behavior of these global filters it will be necessary to deal with the algorithms of the SPKF's.

III. SOME ASPECTS OF SIGMA POINT KALMAN FILTERS

The goal of this section is to analyse algorithms of the local filters, namely the UKF, the DD1 and the DD2, and consequently to find the reason which causes the significant differences in the estimated pdf's of the global filters SPGSF(UKF,DD2) and SPGSF(DD1). The main stress is laid on the relations for the covariance matrix computation.

The algorithms of all local filters have the same structure where the filtering and the predictive mean and covariance matrix are recursively computed to obtain the Gaussian approximation of the estimated pdf's [1], [5], [7]. The crucial difference between the particular local filters can be found in the transformation of a random variable through the nonlinear function [5], [7].

A. Transformation of a 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 $\mathbf{y} = \mathbf{g}(\mathbf{x}) = [g_1(\mathbf{x}), \ldots, g_{n_y}(\mathbf{x})]^T$. The variable $\mathbf{x}$ is given by the first two moments, i.e. the mean $\bar{\mathbf{x}}$ and the covariance matrix $\mathbf{P}_x$, and the aim is to calculate the mean $\bar{\mathbf{y}}$ and the covariance matrix $\mathbf{P}_y$ of $\mathbf{y}$ and the cross-covariance matrix $\mathbf{P}_{xy}$.

One of the possible solution is based on the unscented transformation (UT) where the random variable $\mathbf{x}$ is approximated by a set of deterministically chosen weighted σ-points

$$\mathcal{X}_0 = \bar{\mathbf{x}}, \mathcal{W}_0 = \frac{\kappa}{n_x+\kappa},$$
$$\mathcal{X}_i = \bar{\mathbf{x}} + \left(\sqrt{(n_x+\kappa)\mathbf{P}_x}\right)_i,$$
$$\mathcal{X}_j = \bar{\mathbf{x}} - \left(\sqrt{(n_x+\kappa)\mathbf{P}_x}\right)_{j-n_x},$$

where $i = 1,\ldots,n_x$, $j = n_x+1,\ldots,2n_x$, $\mathcal{W}_i = \mathcal{W}_j = 1/\big(2(n_x+\kappa)\big)$, $\forall i, j$ 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

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

And the resulting characteristics are given as

$$\bar{\mathbf{y}}_A^{UKF} = \sum_{i=0}^{2n_x} \mathcal{W}_i \mathcal{Y}_i, \tag{5}$$

$$\mathbf{P}_{y,A}^{UKF} = \sum_{i=0}^{2n_x} \mathcal{W}_i (\mathcal{Y}_i - \bar{\mathbf{y}}_A^{UKF})(\mathcal{Y}_i - \bar{\mathbf{y}}_A^{UKF})^T, \tag{6}$$

$$\mathbf{P}_{xy,A}^{UKF} = \sum_{i=0}^{2n_x} \mathcal{W}_i (\mathcal{X}_i - \bar{\mathbf{x}})(\mathcal{Y}_i - \bar{\mathbf{y}}_A^{UKF})^T, \tag{7}$$

where the subscript A highlights that these results are only approximations of the true mean and the covariance matrixes which can not be generally computed. This approach leads to the UKF. The recommended settings of the scaling parameter κ is $\kappa = 3 - n_x$ for the Gaussian distribution of $\mathbf{x}$ [7].

The next possible approximation utilizes the Stirling's interpolation formula first order [5].

$$\mathbf{y} \approx \mathbf{g}(\bar{\mathbf{x}}) + \frac{1}{h}\left(\sum_{i=1}^{n_x} \Delta x_i \mu_i \delta_i\right)\mathbf{g}(\bar{\mathbf{x}}),$$

where $\Delta x_i = x_i - \bar{x}_i$, h is half of the interpolation interval, $\mu_i\delta_i \mathbf{g}(\bar{\mathbf{x}}) = \frac{\mathbf{g}(\bar{\mathbf{x}}+h\mathbf{s}_i)-\mathbf{g}(\bar{\mathbf{x}}-h\mathbf{s}_i)}{2}$ and $\mathbf{s}_i$ is the i-th column of the matrix $\mathbf{S}_x$ which is the square root of the covariance matrix $\mathbf{P}_x = \mathbf{S}_x\mathbf{S}_x^T$. For the desired characteristics it holds

$$\bar{\mathbf{y}}_A^{DD1} = \mathbf{g}(\bar{\mathbf{x}}), \tag{8}$$

$$\mathbf{P}_{y,A}^{DD1} = \frac{1}{4h^2}\sum_{i=1}^{n_x}\big(\mathbf{g}(\bar{\mathbf{x}}+h\mathbf{s}_i) - \mathbf{g}(\bar{\mathbf{x}}-h\mathbf{s}_i)\big)\times \times \big(\mathbf{g}(\bar{\mathbf{x}}+h\mathbf{s}_i) - \mathbf{g}(\bar{\mathbf{x}}-h\mathbf{s}_i)\big)^T = \mathbf{S}_{y,A}^{DD1}(\mathbf{S}_{y,A}^{DD1})^T, \tag{9}$$

$$\mathbf{P}_{xy,A}^{DD1} = \frac{1}{2h}\sum_{i=1}^{n_x}\mathbf{s}_i\big(\mathbf{g}(\bar{\mathbf{x}}+h\mathbf{s}_i) - \mathbf{g}(\bar{\mathbf{x}}-h\mathbf{s}_i)\big)^T = \mathbf{S}_x(\mathbf{S}_{y,A}^{DD1})^T, \tag{10}$$

$\mathbf{S}_{y,A}^{DD1} = \{S_{y,A}^{DD1}(i,j)\} = \{\frac{1}{2h}\big(g_i(\bar{\mathbf{x}}+h\mathbf{s}_j) - g_i(\bar{\mathbf{x}}-h\mathbf{s}_j)\big)\}$.

The more exact approximation is feasible by exploiting of the Stirling's interpolation second order [5]

$$\mathbf{y} \approx \mathbf{g}(\bar{\mathbf{x}}) + \frac{1}{h}\left(\sum_{i=1}^{n_x}\Delta x_i\mu_i\delta_i\right)\mathbf{g}(\bar{\mathbf{x}}) + \frac{1}{2h^2}\Bigg(\sum_{i=1}^{n_x}(\Delta x_i)^2 + \sum_{i=1}^{n_x}\sum_{j=1, i\neq j}^{n_x}\Delta x_i \Delta x_j(\mu_i\delta_i)(\mu_j\delta_j)\Bigg)\mathbf{g}(\bar{\mathbf{x}}),$$

where $\delta_i^2\mathbf{g}(\bar{\mathbf{x}}) = \mathbf{g}(\bar{\mathbf{x}}+h\mathbf{s}_i) + \mathbf{g}(\bar{\mathbf{x}}-h\mathbf{s}_i)$ which lead to the more accurate characteristics

$$\bar{\mathbf{y}}_A^{DD2} = \frac{h^2-n_x}{h^2}\mathbf{g}(\bar{\mathbf{x}}) + \frac{1}{2h^2}\sum_{i=1}^{n_x}\big(\mathbf{g}(\mathbf{x}+h\mathbf{s}_i) + \mathbf{g}(\mathbf{x}-h\mathbf{s}_i)\big), \tag{11}$$

$$\mathbf{P}_{y,A}^{DD2} = \mathbf{P}_{y,A}^{DD1} + \frac{h^2-1}{4h^4}\sum_{i=1}^{n_x}\big(\mathbf{g}(\bar{\mathbf{x}}+h\mathbf{s}_i) + \mathbf{g}(\bar{\mathbf{x}}-h\mathbf{s}_i) - 2\mathbf{g}(\bar{\mathbf{x}})\big)\times\big(\mathbf{g}(\bar{\mathbf{x}}+h\mathbf{s}_i) + \mathbf{g}(\bar{\mathbf{x}}-h\mathbf{s}_i) - 2\mathbf{g}(\bar{\mathbf{x}})\big)^T = \mathbf{P}_{y,A}^{DD1} + \mathbf{P}_{y,e}^{DD2} \tag{12}$$

$$\mathbf{P}_{xy,A}^{DD2} = \frac{1}{2h}\sum_{i=1}^{n_x}\mathbf{s}_i\big(\mathbf{g}(\bar{\mathbf{x}}+h\mathbf{s}_i) - \mathbf{g}(\bar{\mathbf{x}}-h\mathbf{s}_i)\big)^T. \tag{13}$$

The optimal choice of the interval length is $h^2 = 3$ for the Gaussian distribution of $\mathbf{x}$ [5]. Note that if length of the interval is chosen $h^2 = 1$ then the relation $\mathbf{P}_{y,A}^{DD2}$ (12) will have formally the same form as $\mathbf{P}_{y,A}^{DD1}$ (9). The DDF's are based on this approximation.

The resulting estimates of the local filters are assumed to be Gaussian but the true pdf's are unknown. Therefore, the choice of the scaling parameters κ and h^2 can be different from the recommended settings.

B. Common features of unscented transformation and Stirling's interpolation

It was shown that the relation for the mean computation $\bar{\mathbf{y}}_A^{UKF}$ and $\bar{\mathbf{y}}_A^{DD2}$ are the same and the covariance matrix $\mathbf{P}_{y,A}^{UKF}$ is less exact than $\mathbf{P}_{y,A}^{DD2}$ [5]. The aim of this section is to show that the relation for $\mathbf{P}_{y,A}^{UKF}$ can be written as

$$\mathbf{P}_{y,A}^{UKF} = \mathbf{P}_{y,A}^{DD1} + \mathbf{P}_{y,e}^{UKF} \tag{14}$$

similarly to $\mathbf{P}_{y,A}^{DD2}$ (12).

Firstly, to confirm the relation (14), it is advantageous to express the unscented transformation in the form [10]

$$\mathbf{y} \approx \mathbf{b} + \mathbf{A}\mathbf{x},$$

where $\mathbf{A} = (\mathbf{P}_{xy,A}^{UKF})^T\mathbf{P}_x^{-1}$, $\mathbf{b} = \bar{\mathbf{y}}_A^{UKF} - \mathbf{A}\bar{\mathbf{x}}$. Then, the covariance matrix can be rewritten as [10]

$$\mathbf{P}_{y,A}^{UKF} = \mathbf{A}\mathbf{P}_x\mathbf{A}^T + \mathbf{P}_{y,e}^{UKF}. \tag{15}$$

Secondly, it is necessary to show the equality $\mathbf{P}_{xy,A}^{UKF} = \mathbf{P}_{xy,A}^{DD1,2}$. Proceeding from the relation (7), where $n_x+\kappa = h^2$ and $\left(\sqrt{(n_x+\kappa)\mathbf{P}_x}\right)_i = h\mathbf{s}_i$, it holds that

$$\mathbf{P}_{xy,A}^{UKF} = \frac{\kappa}{n_x+\kappa}(\mathcal{X}_0 - \bar{\mathbf{x}})(\mathcal{Y}_0 - \bar{\mathbf{y}}_A^{UKF})^T + \frac{1}{2(n_x+\kappa)}\sum_{i=1}^{2n_x}(\mathcal{X}_i - \bar{\mathbf{x}})(\mathcal{Y}_i - \bar{\mathbf{y}}_A^{UKF})^T$$
$$= \frac{1}{2h^2}\sum_{i=1}^{n_x}\Big[h\mathbf{s}_i\big(\mathbf{g}(\bar{\mathbf{x}}+h\mathbf{s}_i) - \mathbf{g}(\bar{\mathbf{x}})\big)^T + (-h\mathbf{s}_i)\big(\mathbf{g}(\bar{\mathbf{x}}-h\mathbf{s}_i) - \mathbf{g}(\bar{\mathbf{x}})\big)^T\Big] =$$

$$= \frac{1}{2h}\sum_{i=1}^{n_x} \mathbf{s}_i\big(\mathbf{g}(\bar{\mathbf{x}}+h\mathbf{s}_i)-\mathbf{g}(\bar{\mathbf{x}}-h\mathbf{s}_i)\big)^T = \mathbf{P}_{xy,A}^{DD1,2}$$

Finally, return to the relation (15) and recall (10)

$$\begin{aligned}\mathbf{P}_{y,A}^{UKF} &= (\mathbf{P}_{xy,A}^{UKF})^T\mathbf{P}_x^{-1}\mathbf{P}_{xy,A}^{UKF} + \mathbf{P}_{y,e}^{UKF} \\ &= \mathbf{S}_{y,A}^{DD1}\mathbf{S}_x^T(\mathbf{S}_x\mathbf{S}_x^T)^{-1}\mathbf{S}_x(\mathbf{S}_{y,A}^{DD1})^T + \mathbf{P}_{y,e}^{UKF} \\ &= \mathbf{S}_{y,A}^{DD1}\mathbf{I}_{n_x}(\mathbf{S}_{y,A}^{DD1})^T + \mathbf{P}_{y,e}^{UKF} = \mathbf{P}_{y,A}^{DD1} + \mathbf{P}_{y,e}^{UKF}.\end{aligned} \tag{16}$$

That means the covariance matrixes $\mathbf{P}_{y,A}^{DD2}$ (12) and $\mathbf{P}_{y,A}^{UKF}$ (6) are greater or equal to $\mathbf{P}_{y,A}^{DD1}$ (9) by preservation of the condition $n_x + \kappa = h^2$. Unfortunately, the relation between $\mathbf{P}_{y,e}^{DD2}$ in (12) and $\mathbf{P}_{y,e}^{UKF}$ is generally unknown.

Transformation of a scalar variable. Nevertheless, if the scalar variables are considered in the system description and the condition $n_x+\kappa = h^2$ is fulfilled, the variances P_y^{DD2} and P_y^{UKF} are the same. To prove this, it is suitable to proceed from the square root version of the UKF [8] and return to the transformation of the random variable via the nonlinear function. Relations (5) and (6) can be rewritten in the form

$$\begin{aligned}\bar{y}_A^{UKF} =& \frac{h^2-1}{h^2}g(\bar{x}) \\ &+\frac{1}{2h^2}\big(g(\bar{x}+hS_x)+g(\bar{x}-hS_x)\big),\end{aligned} \tag{17}$$

$$\begin{aligned}S_{y,A}^{UKF} =& \sqrt{\frac{1}{2h^2}}\big[\xi(g(\bar{x})-\bar{y}_A^{UKF}), g(\bar{x}+hS_x) \\ &- \bar{y}_A^{UKF}, g(\bar{x}-hS_x)-\bar{y}_A^{UKF}\big],\end{aligned} \tag{18}$$

where $\xi = \sqrt{2(h^2-1)}$ and $P_{y,A}^{UKF} = S_{y,A}^{UKF}(S_{y,A}^{UKF})^T$. Then, substitution of $\bar{y}_A^{UKF}$ (17) in (18) and some necessary adjustments yield

$$\begin{aligned}S_{y,A}^{UKF} =& \frac{1}{2h}\big[g(\bar{x}+hS_x)-g(\bar{x}-hS_x), \\ &\frac{\sqrt{h^2-1}}{h}\big(g(\bar{x}+hS_x)+g(\bar{x}-hS_x)+2g(\bar{x})\big)\big].\end{aligned} \tag{19}$$

Considering the relations (12), (18) and (19) it is clear

$$P_{y,A}^{UKF} = S_{y,A}^{UKF}(S_{y,A}^{UKF})^T = P_{y,A}^{DD2}. \tag{20}$$

From (16), (20) the equality $P_{y,e}^{UKF} = P_{y,e}^{DD2}$ follows as well.

Therefore for the scalar systems (1), (2) the algorithms of the square root version of the UKF [8] and the DD2 are the same. Because the estimation performance of the square root UKF and the original UKF is naturally the same [8] the estimation performance of the UKF and the DD2 (for scalar systems) are the same as well.

C. Impact approximation on filtering covariance matrix

The relations for computation of the filtering mean $\hat{\mathbf{x}}_k = E[\mathbf{x}_k|\mathbf{z}^k]$ and the covariance matrix $\mathbf{P}_k = cov[\mathbf{x}_k|\mathbf{z}^k]$ for the local filters represents the Gaussian approximation of the filtering pdf, i.e. $p(\mathbf{x}_k|\mathbf{z}^k) \approx \mathcal{N}\{\mathbf{x}_k : \hat{\mathbf{x}}_k, \mathbf{P}_k\}$, where

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^{'} + \mathbf{P}_{xz,k}^{'}\mathbf{P}_{z,k}^{'-1}(\mathbf{z}_k - \hat{\mathbf{z}}_k^{'}), \tag{21}$$

$$\mathbf{P}_k = \mathbf{P}_k^{'} - \mathbf{P}_{xz,k}^{'}\mathbf{P}_{z,k}^{'-1}\mathbf{P}_{xz,k}^{'T}, \tag{22}$$

and $\hat{\mathbf{x}}_k^{'}$, $\mathbf{P}_k^{'}$, $\hat{\mathbf{z}}_k^{'}$, $\mathbf{P}_{z,k}^{'}$, $\mathbf{P}_{xz,k}^{'}$ represent the predictive mean and covariance matrix of the state estimate, the predictive mean and covariance matrix of the measurement estimate and the predictive cross-covariance matrix at time instant k, respectively. It should be mentioned that the covariance matrix of the measurement noise $\mathbf{R}_k$ is included in the predictive covariance matrix of the measurement $\mathbf{P}_{z,k}^{'}$. Relations (21), (22) are formally the same for all the local filters. Substantial difference lies in the transformation of the predictive state estimate given by $\hat{\mathbf{x}}_k^{'}$, $\mathbf{P}_k^{'}$ through the measurement function $\mathbf{h}_k(\cdot)$ with respect to the above mentioned approximation which yields $\hat{\mathbf{z}}_k^{'}$, $\mathbf{P}_{z,k}^{'}$ and $\mathbf{P}_{xz,k}^{'}$ [5], [7], [8].

Therefore the reason of quite different filtering covariance matrixes $\mathbf{P}_k$ for particular SPKF's can be found in the fact that $\mathbf{P}_{y,A}^{DD2,UKF} \geqq \mathbf{P}_{y,A}^{DD1}$ and in the relation (22). The cross-covariance matrixes $\mathbf{P}_{xz,k}^{'}$ are formally the same for all the filters, i.e. for the UKF, the DD1 and the DD2. However, the covariance matrixes $\mathbf{P}_{z,k}$ of the DD2 or the UKF are greater than (or equal to) the DD1 one. In equation (22) the inversion of $\mathbf{P}_{z,k}^{'}$ is required and if $\mathbf{P}_{z,e,k}^{'DD2,UKF}$ is significant towards $\mathbf{P}_{z,k}^{'DD1}$ and $\mathbf{P}_{z,e,k}^{'DD2,UKF} \gg \mathbf{R}_k$ then $(\mathbf{P}_{z,k}^{'DD2,UKF})^{-1} \ll (\mathbf{P}_{z,k}^{'DD1})^{-1}$ which can cause $\mathbf{P}_k^{DD2,UKF} \gg \mathbf{P}_k^{DD1}$. The last inequality can be found as an objective that induces the differences in the estimated pdf's (see Fig. 1).

In other words, the difference between the filtering covariance matrixes $\mathbf{P}_k^{DD2,UKF}$ and $\mathbf{P}_k^{DD1}$ was shown. This difference becomes significant especially in the situations where the covariance matrix of the measurement noise $\mathbf{v}_k$ in (2) is significantly less than the covariance matrix of the state noise $\mathbf{w}_k$ in (1) and the function $\mathbf{h}_k(\cdot)$ in (2) is highly nonlinear.

IV. Sigma Point Gaussian Sum Filter Performance

This section is devoted to the description of influence of the derivative-free local filters in the Gaussian sum approach. The main idea of this approach is based on the approximation of an arbitrary pdf by a Gaussian mixture [8], [12], [13]. To apply this idea for the system (1), (2), it is necessary to assume the prior pdf, the state and the measurement noise pdf's in the form of Gaussian mixtures. To obtain a close-loop solution of the BRR, the multipoint approximation of the nonlinear functions $\mathbf{f}_k(\cdot)$, $\mathbf{h}_k(\cdot)$ has to be done. Then, the SPGSF or the GSF can be understood as a bank of concurrently running SPKF's or EKF's, respectively, generating filtering pdf's in the form of Gaussian mixtures [8], [13]. In the progress of estimation each local filter and its estimate is assessed by the weight corresponding to the estimate accuracy towards the measurement.

For the analysis of the SPKF's impact on the Gaussian sum properties it is advantageous to proceed from the relations for the filtering mean and the covariance matrix computation of Gaussian mixture of N Gaussians [13]

$$\hat{\mathbf{x}}_k = \sum_{j=1}^{N}\alpha_{j,k}\hat{\mathbf{x}}_{j,k}, \tag{23}$$

TABLE II

NONLINEAR TRANSFORMATION OF RANDOM VARIABLE.

	$h^2=3$		$\kappa=2$	$h^2=1$		$\kappa=0$	
	DD1	DD2	UKF	DD1	DD2	UKF	MC
$\bar{y}$	2.80	4.20	4.20	2.80	4.20	4.20	4.20
P_y	4.48	8.40	8.40	4.48	4.48	4.48	8.40
$P_{y,e}$	–	3.92	3.92	–	0	0	–
P_{xy}	5.60	5.60	5.60	5.60	5.60	5.60	5.60

$$\mathbf{P}_k = \sum_{j=1}^{N} \alpha_{j,k} \left(\mathbf{P}_{j,k} + (\hat{\mathbf{x}}_k - \hat{\mathbf{x}}_{j,k})(\hat{\mathbf{x}}_k - \hat{\mathbf{x}}_{j,k})^T \right), \qquad (24)$$

where $\hat{\mathbf{x}}_{j,k}$ and $\mathbf{P}_{j,k}$ are the mean and the covariance matrix obtained by the j-th SPKF at the time instant k. The weight of the j-th local filter is given by $\alpha_{j,k}$. While the filtering mean $\hat{\mathbf{x}}_k$ (23) is the weighted sum of particular means $\hat{\mathbf{x}}_{j,k}$ (21) the covariance matrix $\mathbf{P}_k$ (24) is not only the sum of the particular covariance matrixes $\mathbf{P}_{j,k}$ (21) but it is also influenced by the placement of the mean values $\hat{\mathbf{x}}_{j,k}$ (21). Then, the impact of the particular local filter covariance matrixes $\mathbf{P}_{j,k}$ on the global $\mathbf{P}_k$ can be reduced especially if the means $\hat{\mathbf{x}}_{j,k}$ are widely spread in the state space. Therefore the global covariance matrixes $\mathbf{P}_k$ of the SPGSF(DD1), the SPGSF(DD2) and the SPGSF(UKF) can be mutually very similar although the covariance matrixes of the DD1, the DD2 and the UKF are considerably different.

V. NUMERICAL ILLUSTRATIONS

This section is devoted to verification of the theoretical results given in the previous parts.

First of all an example of nonlinear transformation of random variable is given. Consider a scalar normal random variable x with known mean $\bar{x} = 2$ and variance $P_x = 7$ and the nonlinear function $y = g(x) = 0.2x^2 + 2$. The resulting characteristics of the random variable y obtained by means of the UT and the Stirling's interpolation 1st and 2nd order are shown in Table II where impact of choice of the parameters κ and h^2 on the variances is illustrated. Therefore in some cases, especially for strong nonlinearities, the increase of covariance for P_y^{UKF} and P_y^{DD2} with respect to P_y^{DD1} is significant. However, if $\kappa = 0$ and $h^2 = 1$ the variance $P_y^{DD2,UKF}$ is equal to P_y^{DD1}. The true values were obtained by the Monte Carlo (MC) simulation with 10^7 samples.

Now, the nonlinear system (3), (4) defined above, which is quite unusual due to the fundamental difference in size of the state and the measurement noise variance is considered. In Fig. 1 the different estimated pdf's of the SPGSF(DD1) and the SPGSF(UKF) are depicted. Note that common choice of the parameters (i.e. $\kappa = 2$ and $h^2 = 3$) were used.

However, in the case of the scaling parameters $\kappa = 0$ or $h^2 = 1$ the relation for the covariance matrixes of the UKF and the DD2 are formally the same as the DD1 ones but the relations for the means remain formally without any change as it was shown above. This fact may be the reason of similar shape of the filtering pdf's of the SPGSF(DD1) and the SPGSF(UKF,DD2) (see Fig. 2) which are moved

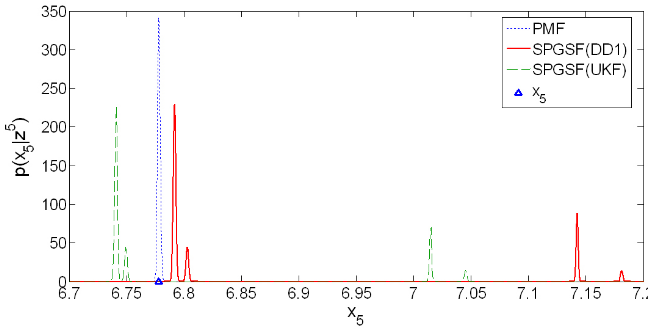

Fig. 2. Estimated filtering pdf's ($\kappa = 0$, $h^2 = 1$).

TABLE III

DEPENDENCE OF MSE ON NOISE VARIANCE.

R_k	10^{-5}	10^{-3}	10^{-2}	10^{-1}	1
$MSE_{h^2=3}^{SPGSF(DD1)}$	0.0222	0.0277	0.0425	0.2130	1.2400
$MSE_{\kappa=2}^{SPGSF(UKF)}$	0.0133	0.0163	0.0327	0.1654	1.2367
$MSE_{\kappa=0}^{SPGSF(UKF)}$	0.0129	0.0173	0.0289	0.2037	1.2454
MSE^{PMF}	$3.55 \cdot 10^{-4}$	0.0021	0.0145	0.1625	1.3997
$MSE_{h^2=3}^{DD1}$	0.1955	0.2123	0.2156	0.4018	1.4064
$MSE_{\kappa=2}^{UKF}$	0.1304	0.1424	0.1508	0.3320	1.3567
$MSE_{\kappa=0}^{UKF}$	0.1690	0.1884	0.1874	0.3694	1.3867

due to the more exact calculation of the mean in the UKF or the DD2 towards the DD1. It is necessary to note that the choice of the parameters $\kappa = 0$ or $h^2 = 1$ causes the accuracy of estimated pdf's to be less exact due to the less exact covariance matrix computation (the true state in most cases is not in the area of nonzero probability) even if the shape is more similar to the "true" pdf obtained by the PMF.

One of the aims of this paper was to demonstrate that the global filter comparison based on a point estimates only is insufficient. To confirm this, two tables are given (Table III includes $MSE = \left(\sum_{i=1}^{100}\sum_{k=1}^{60}(x_k^i - \hat{x}_k^i)^2\right)/6000$ and Table IV includes average values of variance at $k = 5$, i.e. $V_c = \left(\sum_{i=1}^{100} P_5^i)\right)/100)$ with respect to the different choice of the measurement noise variance R_k and different choice of the scaling parameters (specified in the subscript). The experiment was repeated 100 times in order to calculate MC performance estimates for each filter (the superscript i determines the order of repetition). For completeness MSE and V_c for the DD1 and the UKF are shown as well.

TABLE IV

DEPENDENCE OF FILTERING VARIANCE ON NOISE VARIANCE.

R_k	10^{-5}	10^{-3}	10^{-2}	10^{-1}	1
$V_{c,h^2=3}^{SPGSF(DD1)}$	0.0239	0.0268	0.0271	0.0336	0.0881
$V_{c,\kappa=2}^{SPGSF(UKF)}$	0.0809	0.0662	0.0644	0.0728	0.1245
$V_{c,\kappa=0}^{SPGSF(UKF)}$	0.0223	0.0217	0.0189	0.0278	0.0788
V_c^{PMF}	$1.36 \cdot 10^{-6}$	$7.71 \cdot 10^{-5}$	0.0007	0.0076	0.0738
$V_{c,h^2=3}^{DD1}$	$5.93 \cdot 10^{-7}$	$5.73 \cdot 10^{-5}$	0.0005	0.0057	0.0562
$V_{c,\kappa=2}^{UKF}$	0.1753	0.1696	0.1720	0.1751	0.2231
$V_{c,\kappa=0}^{UKF}$	$6.14 \cdot 10^{-7}$	$5.93 \cdot 10^{-5}$	0.0006	0.0059	0.0788

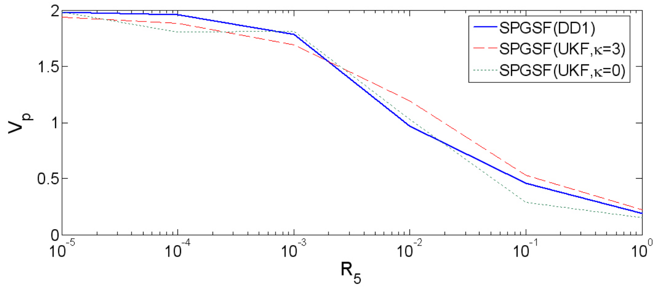

Fig. 3. Dependence of criterion V_p on noise variance.

The average variance V_c (Table IV) is quite different for the UKF($\kappa = 2$) and the DD1 for "small" values of R_k. If $\kappa = 0$ the influence of $P_{y,e}^{UKF}$ is cancelled and the filtering variance P_k of the UKF is comparable with the DD1's one. With increasing of variance R_k the estimated filtering variances of the DD1 and the UKF (for any choice of κ) become the same. That is that impact of $P_{y,e}^{UKF}$ is reduced by R_k (i.e. $R_k \approx P_{y,e}^{UKF}$). The choice of the scaling parameter h^2 for the estimation quality of the DD1 is negligible.

From Tables III, IV and Fig. 1, 2 it is clear that estimated pdf's are specified neither by the mean nor by the variance. For instance, the values of the MSE and the criterion V_c for the SPGSF(DD1) and the SPGSF(UKF,$\kappa = 2$) are quite similar but the shapes of pdf's are completely different. Likewise to the local filters, the estimated pdf's of the global filters become mutually similar if the variance R_k is increasing. If a scalar system is considered and the condition $n_x + \kappa = h^2$ is kept the estimation performance of the algorithms SPGSF(UKF) and the SPGSF(DD2) becomes the same. Moreover, if the square root UKF is considered, the algorithms of the SPGSF(UKF) and the SPGSF(DD2) are the same for scalar systems. In the case of the multidimensional systems, their estimated pdf's are comparable. Further, the estimates of the GSF resemble to the SPGSF(DD1) ones.

The SPGSF's comparison from the viewpoint of the estimated pdf's is done by means of the criterion $V_p = \left(\sum_{i=1}^{100} \int |p^i(x_5|z^5) - \hat{p}^i(x_5|z^5)|\right)/100$ where $p^i(x_5|z^5)$ is the "true" pdf obtained by the PMF and $\hat{p}^i(x_5|z^5)$ is the pdf of the particular SPGSF. The results are depicted in Fig. 3.

Moreover, it should be mentioned the estimated pdf's of the Unscented Particle Filter and the Gaussian Mixture Sigma Point Particle Filter [16], which are a combination of the Particle Filter and the UKF, are close to the SPGSF(UKF) ones for arbitrary chosen scaling parameter κ.

The simulation results support the theoretical conclusions from previous section. The significant difference between the covariance matrixes $P_{y,A}^{DD1}$ and $P_{y,A}^{DD2,UKF}$ is confirmed by Table II. Impact of particular approximations to the derivative-free local and global filters is given in Tables III, IV where influence of the scaling parameter κ is shown as well. It means that the estimated pdf's of the SPGSF's do not depend on the selected local filter (i.e. on the selected approximation) only but they also strongly depend on the choice of the scaling parameters as well. Although the first two moments of the estimated pdf's of the SPGSF's given in Table III, IV seem to be very similar, the estimated pdf's are substantially different. That is the reason for the SPGSF's comparison based on the pdf's given by Fig. 1, 2 and 3.

VI. Conclusion

The derivative-free Kalman filters were introduced under the Gaussian sum framework. The impact of the choice of the particular Sigma Point Kalman Filter on the resulting estimated pdf's of the Sigma Point Gaussian Sum Filter was shown. To explain the quite different shapes of the estimated pdf's of the particular Sigma Point Gaussian Sum Filters, the analysis of the Sigma Point Kalman Filters was done mainly from the viewpoint of their covariance matrixes. The analysis brought the new relations between the Divided Difference Filters and the Unscented Kalman Filter which allow the performance analysis of the derivative-free filters. The theoretical results were verified by the numerical illustrations.

References

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

[2] M. Šimandl, "State Estimation for Non-Gaussian Models", In *Preprints of the 13Th World Congress*, vol. 2, pp 463–468, San Francisco, USA, IFAC, Elsevier Science, 1996.

[3] H.W. Sorenson, On Development of Practical Nonlinear Filters, *Inf. Sci.*, vol. 7, pp 230–270, 1974.

[4] M. Athans, R.P. Wisher and A. Bertolini, Suboptimal State Estimation for Continuous-Time Nonlinear Systems with Discrete Noisy Measurements, *IEEE Transactions On Automatic Control*, (5):504–514, 1968.

[5] M. Nørgaard, N.K. Poulsen and O. Ravn, New Developments in State Estimation for Nonlinear Systems, *Automatica*, 36(11):1627–1638, 2000.

[6] T.S. Schei, A Finite-Difference Method for Linearization in Nonlinear Estimation Algorithms, *Automatica*, 33(11):2053–2058, 1997.

[7] 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 AC*, 45(3):477–482, 2000.

[8] M. Šimandl and J. Duník, "Sigma Point Gaussian Sum Filter Design Using Square Root Unscented Filters", In *Preprints of the 16th IFAC World Congress in Prague*, Prague, Czech Republic, 2005.

[9] K. Ito and K. Xiong, Gaussian Filters for Nonlinear Filtering Problems, *Automatica*, 45(5):910–927, 2000.

[10] 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*, 47(8):1406–1409, 2002.

[11] T. Lefebvre, H. Bruyninckx and J. De Schutter, Kalman Filters for Nonlinear Systems: A Comparsion of Performance, *International Journal of Control*, 77(7):639–653, 2004.

[12] H.W. Sorenson and D.L. Alspach, Recursive Bayesian Estimation Using Gaussian Sums, *Automatica*, 7:465–479, 1971.

[13] M. Šimandl and J. Královec, "Filtering, Prediction and Smoothing with Gaussian Sum Representation", In *Proceedings of the 12th IFAC Symposium on System Identification*, Santa Barbara, California, USA, 2000.

[14] R.S. Bucy and K.D. Senne, Digital Synthesis of Non-linear Filters, *Automatica*, 7:287–298, 1971.

[15] M. Šimandl, J. Královec, and T. Söderström, Anticipative Grid Design in Point-Mass Approach to Nonlinear State Estimation, *IEEE Transactions On Automatic Control*, 47(4), 2002.

[16] R. van der Merwe and E.A. Wan, "Gaussian Mixture Sigma-Point Particle Filters for Sequential Probabilistic Inference in Dynamic State-Space Models", In *Proceedings of the International Conf. On Acoustics, Speech, and Signal Proc. (ICASSP)*, Hong Kong, 2003.

[17] M. Šimandl and O. Straka, "Sampling Density Design for Particle Filters", In *Preprints of the 13Th IFAC Symposium On System Identification*, Rotterdam, 2003.