On Directional Splitting of Gaussian Density in Nonlinear Random Variable Transformation

Jindřich Duník[1] ✉, Ondřej Straka[1], Benjamin Noack[2], Jannik Steinbring[2], Uwe. D. Hanebeck[2]

[1] *Faculty of Applied Sciences, University of West Bohemia, Univerzitní 8, Pilsen 306 14, Czech Republic*
[2] *Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology, Adenauerring 2, 76131 Karlsruhe, Germany*
✉ *E-mail: dunikj@kky.zcu.cz*

Abstract: Transformation of a random variable is a common need in a design of many algorithms in signal processing, automatic control, and fault detection. Typically, the design is tied to an assumption on a probability density function of the random variable, often in the form of the Gaussian distribution. The assumption may be, however, difficult to be met in algorithms involving nonlinear transformation of the random variable. This paper focuses on techniques capable to ensure validity of the Gaussian assumption of the nonlinearly transformed Gaussian variable by approximating the to-be-transformed random variable distribution by a Gaussian mixture distribution. The stress is laid on an analysis and selection of design parameters of the approximate Gaussian mixture distribution to minimise the error imposed by the nonlinear transformation such as the location and number of the Gaussian mixture terms. A special attention is devoted to the definition of the novel Gaussian mixture splitting directions based on the measures of non-Gaussianity. The proposed splitting directions are analysed and illustrated in numerical simulations.

1 Introduction

Nonlinear transformation of one or more statistically independent Gaussian random variables is a cornerstone of *recursive* algorithms for state estimation (Gaussian or Gaussian mixture filters [1–4]), system identification (nonlinear least-squares method [5]), automatic control (dual control [6, 7]), fault diagnosis (residual generation methods [8]), or, generally, of recursive algorithms, where an uncertainty is propagated through a nonlinear dynamic system [9, 10]. Because of algorithms' recursive nature, the probability density function (PDF) of the transformed random variable is usually *assumed* to be Gaussian to make the next step of the algorithm feasible. Such an assumption is typically considered to be valid if the nonlinearity in the considered region is mild [1, 11]. Often, the region is determined by the covariance matrices of the to-be-transformed random variables.

Unfortunately, the validity of the Gaussianity assumption of the transformed variable cannot be typically determined analytically and indirect measures of nonlinearity or non-Gaussianity are used instead [4, 12–16]. The *measures of nonlinearity* (MoNLs) assess similarity of the nonlinear transformation (or function) and its, in some sense optimal, linear approximation. The *measures of non-Gaussianity* (MoNGs), assess, on the other hand, similarity of the transformed variable PDF and its, in some sense optimal, Gaussian approximation. A large value of a measure then indicates a large error of the linear approximation in the region and, subsequently, possible severe *violation* of the Gaussianity assumption of the transformed random variable. The assumption violation may, if not treated correctly, adversely impact the performance of the recursive algorithms.

A feasible approach to prevent the violation of the Gaussianity assumption of the transformed variable is an approximation of the Gaussian PDF of the to-be-transformed variables by a mixture of Gaussian PDFs, each with a smaller covariance matrix than the original Gaussian PDF [1, 2, 4, 10, 17–22]. Decreasing the covariance matrix (and thus narrowing the region where the nonlinear function is assessed) inherently results in an asymptotic reduction of the MoNL corresponding to each term of the Gaussian mixture (GM). As a consequence, the nonlinearly transformed variable PDF is in the form of a mixture distribution, where each term can be sufficiently

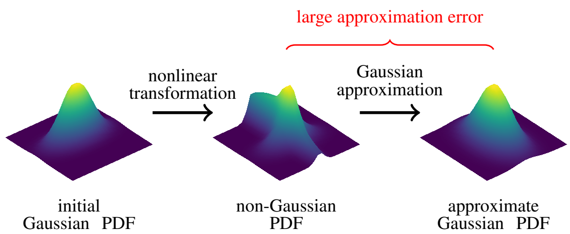

Fig. 1: Illustration of *rough* approximation of PDF of nonlinearly transformed random variable by Gaussian PDF.

well approximated by a Gaussian PDF. With the transformed random variable PDF approximated by a GM PDF, the recursive algorithm may continue for each particular GM term under valid assumptions. the idea of such solution is illustrated in Figs. 1 and 2. In Fig. 1, the to-be-transformed random variable with (initial) Gaussian PDF is directly transformed via a nonlinear transformation. Then, the PDF of the nonlinearly transformed random variable is approximated by a Gaussian PDF. The approximation is, however, rough and results in a large approximation error. In this case, the Gaussian assumption of the transformed variable is *not* valid. In Fig. 2, the (initial) Gaussian PDF is approximated by a GM PDF. Then, the random variable with the GM PDF is nonlinearly transformed which results in a variable with a mixture PDF. The mixture PDF can, subsequently, be accurately approximated by the Gaussian mixture PDF. In this case, the Gaussian assumption of the transformed variable *is locally* valid, i.e., valid for each component of the mixture. Note that the introduced approach and the underlying idea have been widely used and developed in the area of the state estimation in a design of the Gaussian sum filters and the Rao-Blackwellised particle or point-mass filters [1, 23–25]. Gaussian mixture approximation is also conveniently utilised in the uncertainty propagation for the space surveillance or space situational awareness [9, 10, 26].

Approximation of a Gaussian PDF by a suitable GM PDF is, however, an intricate problem requiring specification of many design parameters of the GM PDF. These include the number, locations, and splitting direction of the GM PDF terms. The *number* of the

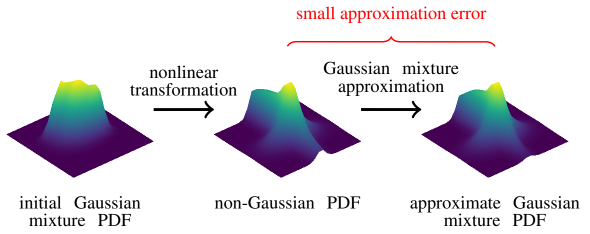

Fig. 2: Illustration of *accurate* approximation of PDF of nonlinearly transformed random variable by Gaussian mixture PDF.

terms is typically a user-defined design parameter; the more terms is used, the better approximation ability can be expected at the expense of increased computational complexity. The *locations* of the terms include distances between mean values, covariance matrices, and the weights of the particular GM PDF terms. The locations are tied with an approach or method used for the GM PDF construction and specified number of terms. The GM PDF can be computed using, e.g., a splitting library [10, 27], an optimisation algorithm with a predefined structure of the GM [2, 28], or a GM moment matching based computation [3, 4, 29]. The *splitting direction* defines the axis along which the GM PDF terms are split, i.e., the line(s) along which the terms of the GM PDF are located. In the literature, three principal splitting directions can be identified; the *direction of (largest) uncertainty* (DoU) [4], the *direction of (largest) nonlinearity* (DoNL) [21, 29–32], and the recent *direction of (largest) non-Gaussianity* (DoNG) [22].

Although the proposed algorithms for splitting direction computation are powerful tools successfully used in various tasks and set-ups, they have certain limitations. The DoU and DoNLs may not be optimal directions in terms of minimisation of the non-Gaussianity of the transformed variable PDF. Two factors affect the PDF of the transformed random variable (i.e., its non-Gaussianity): the PDF of the to-be-transformed variable and the nonlinear transformation. However, the DoU takes into account the PDF of the to-be-transformed variable only and the DoNLs are mainly affected by the properties of the nonlinear transformation. The DoNG is based on *joint* moments of the to-be-transformed and transformed variables. The third-order moment of the transformed variable itself is, however, *not* considered, although it significantly characterises non-Gaussianity of the transformed random variable.

The goal of the paper is to design and analyse a novel methodology for splitting direction computation on the basis of a largest non-Gaussianity addressing the above mentioned limitations and to illustrate its performance in the transformation of polar to Cartesian coordinate systems.

The paper is organised as follows. In Section 2, transformation of a Gaussian random variable through a nonlinear function is reviewed, a general algorithm for the Gaussian PDF splitting into the GM PDF is provided, and the splitting directions are introduced and illustrated. The novel splitting directions based on the largest non-Gaussianity are proposed and discussed in Section 3 and illustrated in Section 4. In Section 5, concluding remarks are given.

2 Problem Formulation, Splitting Directions Definition, and Goal of the Paper

The problem of nonlinear transformation of two independent random variables can be formulated as follows.

2.1 Gaussian Approximation of Nonlinearly Transformed Random Variable Density

Let two statistically independent random variables $\mathbf{x} \in \mathbb{R}^{n_x}$ and $\mathbf{v} \in \mathbb{R}^{n_y}$ with the *known* Gaussian* PDFs $p(\mathbf{x}) = \mathcal{N}\{\mathbf{x}; \bar{\mathbf{x}}, \mathbf{P}_x\}$ and $p(\mathbf{v}) = \mathcal{N}\{\mathbf{v}; \bar{\mathbf{v}}, \mathbf{P}_v\}$, respectively, be considered. Transformation of the random variable $\mathbf{x}$ through a *known* nonlinear function $\mathbf{g}(\cdot) : \mathbb{R}^{n_x} \to \mathbb{R}^{n_y}$ with additive variable $\mathbf{v}$ results in the random variable $\mathbf{y} \in \mathbb{R}^{n_y}$ defined as

$$\mathbf{y} = \mathbf{g}(\mathbf{x}) + \mathbf{v} = \mathbf{y}_{\mathrm{g}} + \mathbf{v}, \tag{1}$$

with a generally non-Gaussian distribution $p(\mathbf{y})$. Note that such transformation typically occurs in the state-space models [1].

Respecting the recursive nature of many algorithms employing the nonlinear transformation of a random variable, the true PDF $p(\mathbf{y})$ is approximated by a Gaussian PDF

$$\hat{p}_{\mathrm{G}}(\mathbf{y}) = \mathcal{N}\{\mathbf{y} : \bar{\mathbf{y}}, \mathbf{P}_y\} \tag{2}$$

parametrised by the moments

$$\bar{\mathbf{y}} = \mathrm{E}_{p(\mathbf{x},\mathbf{v})}[\mathbf{y}] = \int \mathbf{g}(\mathbf{x})p(\mathbf{x})d\mathbf{x} + \bar{\mathbf{v}} = \bar{\mathbf{y}}_{\mathbf{g}} + \bar{\mathbf{v}}, \tag{3}$$

$$\mathbf{P}_y = \mathrm{cov}_{p(\mathbf{x},\mathbf{v})}[\mathbf{y}] = \mathrm{E}_{p(\mathbf{x})}[(\mathbf{g}(\mathbf{x}) - \bar{\mathbf{y}}_{\mathrm{g}})(\mathbf{g}(\mathbf{x}) - \bar{\mathbf{y}}_{\mathrm{g}})^T] + \mathbf{P}_v. \tag{4}$$

The approximation $\hat{p}_{\mathrm{G}}(\mathbf{y})$ (2), obtained by the moment matching, is the *best* Gaussian approximation of $p(\mathbf{y})$ according to the Kullback-Leibler distance† [33]. The moments of the random variable $\mathbf{y}$ (1), i.e., the mean $\bar{\mathbf{y}}$ (3) and covariance matrix $\mathbf{P}_y$ (4), can be computed *without* the knowledge of the true PDF $p(\mathbf{y})$, however, *typically* with the aid of certain approximations‡. The methods for approximate moment computation can be divided into two groups; approximation of the nonlinear function $\mathbf{g}(\cdot)$ and approximation of PDFs of $\mathbf{x}$ and $\mathbf{y}$. The former group is based on a polynomial approximation of $\mathbf{g}(\cdot)$ by the Taylor expansion (TE) [1] or Stirling interpolation [35]. The latter group is characterised by the application of various versions of the deterministic and stochastic integration rules [36–39].

The error of the Gaussian approximation $\hat{p}_{\mathrm{G}}(\mathbf{y})$ (2) can be assessed by an approximation error

$$\tilde{p}_{\mathrm{G}}(\mathbf{y}) = \hat{p}_{\mathrm{G}}(\mathbf{y}) - p(\mathbf{y}), \tag{5}$$

which can be understood as a characterisation of the *non-Gaussianity* of $p(\mathbf{y})$. The error is difficult to compute as an exact computation of $p(\mathbf{y})$ is often not possible or, at least, not convenient.

As the PDF $p(\mathbf{y})$ is often *unknown*, no straightforward measure assessing the error $\tilde{p}_{\mathrm{G}}(\mathbf{y})$ (5) can be used. As a consequence, indirect measures, such as the MoNLs and MoNGs [14], are typically used for an assessment of the difference $\tilde{p}_{\mathrm{G}}(\mathbf{y})$ (5). If a possibly large error $\tilde{p}_{\mathrm{G}}(\mathbf{y})$ is indicated by the measure, then the Gaussian PDF $\hat{p}_{\mathrm{G}}(\mathbf{y})$ (2) cannot be considered as a sufficient approximation to the true PDF $p(\mathbf{y})$. Then, the Gaussian PDF $p(\mathbf{x})$ is *split* into a GM§ [1, 2, 4, 21, 22, 30], where each term has smaller covariance matrix than the covariance matrix of the Gaussian PDF $p(\mathbf{x})$, i.e., $\mathbf{P}_x$.

**The notation $p(\mathbf{x}) = \mathcal{N}\{\mathbf{x}; \bar{\mathbf{x}}, \mathbf{P}_x\}$ stands for a Gaussian distribution of the variable $\mathbf{x}$ with the mean $\bar{\mathbf{x}}$ and the covariance matrix $\mathbf{P}_x$.*

†It should be noted that Gaussian approximations with different moments may be used to approximate $p(\mathbf{y})$. Such approximations may be based on minimisation of a norm of the error (5) [3].

‡The moments of $\mathbf{y}$ cannot be computed exactly except for special cases. The special cases include, for example, polynomial functions [34].

§GM PDF can approximate any PDF as closely as desired [1] and inherits some of important advantages of the Gaussian PDF [40].

2.2 Gaussian PDF Splitting: General Algorithm

Splitting a Gaussian PDF $p(\mathbf{x})$ into a GM PDF can be understood as an approximation of $p(\mathbf{x})$ by a GM PDF[¶]

$$p(\mathbf{x}) \approx \hat{p}_{\mathrm{GM},N}(\mathbf{x}) = \sum_{n=1}^{N} \omega^{(n)} \mathcal{N}\{\mathbf{x}; \bar{\mathbf{x}}^{(n)}, \mathbf{P}_x^{(n)}\} \quad (6)$$

with a desired accuracy, where N is the number of the GM terms, $\sum_{n=1}^{N} \omega^{(n)} = 1$, $\omega^{(n)} \geq 0, \forall n$, $\bar{\mathbf{x}}^{(i)} \neq \bar{\mathbf{x}}^{(j)}$ with $i, j \in \{1, \ldots, N\}, i \neq j$, and each particular GM term $\mathcal{N}\{\mathbf{x}; \bar{\mathbf{x}}^{(n)}, \mathbf{P}_x^{(n)}\}$ has a covariance matrix smaller than $\mathbf{P}_x$, i.e., $\mathbf{P}_x^{(n)} \leq \mathbf{P}_x, \forall n$, meaning that the difference $\mathbf{P}_x - \mathbf{P}_x^{(n)}$ is positive semidefinite [1, 2]. In [2], it was proven that the GM PDF (6) converges uniformly to $p(\mathbf{x})$ with increasing N and $\mathbf{P}_x^{(n)}$ going to zero for $n = 1, 2, \ldots N$. For notational convenience, the simplified notation $\mathcal{N}^{(n)}(\mathbf{x}) = \mathcal{N}\{\mathbf{x}; \bar{\mathbf{x}}^{(n)}, \mathbf{P}_x^{(n)}\}$ will be used hereafter.

Then, the transformation of the variable $\mathbf{x}$ with the PDF $\hat{p}_{\mathrm{GM},N}(\mathbf{x})$ (6) via the nonlinear transformation (1) results in the PDF in the form of weighted mixture of densities (denoted by subscript M)

$$\hat{p}_{\mathrm{M},N}(\mathbf{y}) = \sum_{n=1}^{N} \omega^{(n)} p^{(n)}(\mathbf{y}), \quad (7)$$

where $p^{(n)}(\mathbf{y})$ is the PDF of $\mathbf{y} = \mathbf{g}(\mathbf{x}) + \mathbf{v}$, where the variable $\mathbf{x}$ is distributed according to $\mathcal{N}^{(n)}(\mathbf{x})$. Further, each term in (7) is approximated by a Gaussian PDF (by the moment matching, thus, preserving the first two moments of $p^{(n)}(\mathbf{y})$), which results in the GM PDF of the variable $\mathbf{y}$ of the form

$$\hat{p}_{\mathrm{M},N}(\mathbf{y}) \approx \hat{p}_{\mathrm{GM},N}(\mathbf{y}) = \sum_{n=1}^{N} \omega^{(n)} \mathcal{N}\{\mathbf{y}; \bar{\mathbf{y}}^{(n)}, \mathbf{P}_y^{(n)}\} = \sum_{n=1}^{N} \omega^{(n)} \mathcal{N}^{(n)}(\mathbf{y}), \quad (8)$$

with $\bar{\mathbf{y}}^{(n)} = \mathrm{E}_{\mathcal{N}^{(n)}(\mathbf{x})}[\mathbf{g}(\mathbf{x})] + \bar{\mathbf{v}}$ and $\mathbf{P}_y^{(n)} = \mathrm{cov}_{\mathcal{N}^{(n)}(\mathbf{x})}[\mathbf{g}(\mathbf{x})] + \mathbf{P}_v$.

In [1], it was shown that with $p(\mathbf{x})$ expressed by the GM PDF (6), the PDF $p(\mathbf{y})$ converges uniformly to the GM PDF (8) as $\mathbf{P}_x^{(n)}$ goes to zero for $n = 1, 2, \ldots N$. Hence, as for a *nonlinear* function $\mathbf{g}(\cdot)$, the error $\tilde{p}_{\mathrm{G}}(\mathbf{y})$ (5) is non-zero, there exists a number of terms N such that the error of the approximate GM PDF defined as

$$\tilde{p}_{\mathrm{GM},N}(\mathbf{y}) = \hat{p}_{\mathrm{GM},N}(\mathbf{y}) - p(\mathbf{y}), \quad (9)$$

satisfies

$$|\tilde{p}_{\mathrm{GM},N}(\mathbf{y})| \leq |\tilde{p}_{\mathrm{G}}(\mathbf{y})|, \quad \forall \mathbf{y} \in \mathbb{R}^{n_y}. \quad (10)$$

Note that by choosing N, the Gaussian sum error $\tilde{p}_{\mathrm{GM},N}(\mathbf{y})$ can be made arbitrarily small. The intuition behind (10) rests in the fact that in a sufficiently small neighbourhood, any smooth nonlinear function will be approximately linear and Gaussian PDFs with smaller covariance matrices remain more Gaussian than the Gaussian PDFs with larger covariance matrices propagated through the nonlinear function [26].

Although, the concept of the Gaussian PDF splitting is relatively simple and well justified, it is composed of several non-trivial steps often requiring a user decision or interaction. These steps, including specification of the number of GM terms N, GM terms means $\bar{\mathbf{x}}^{(n)}$, and GM terms covariance matrices $\mathbf{P}_x^{(n)}$ (i.e, GM terms placement), are detailed below in the following general algorithm.

Algorithm 1: Splitting a Gaussian PDF $p(\mathbf{x})$ to a GM PDF $\hat{p}_{\mathrm{GM}}(\mathbf{x})$

Initialisation: The number of GM terms N in (6) is defined.
Decorrelation: The Gaussian random variable $\mathbf{x}$ with $p(\mathbf{x}) = \mathcal{N}\{\mathbf{x}; \bar{\mathbf{x}}, \mathbf{P}_x\}$ is transformed to a standard normal random variable by

$$\boldsymbol{\chi}_{\mathrm{S}} = \mathbf{S}_x^{-1}(\mathbf{x} - \bar{\mathbf{x}}) \quad (11)$$

leading to $p(\boldsymbol{\chi}_{\mathrm{S}}) = \mathcal{N}\{\boldsymbol{\chi}_{\mathrm{S}} : \mathbf{0}_{n_x}, \mathbf{I}_{n_x}\}$, where $\mathbf{S}_x$ is a factor of the covariance matrix $\mathbf{P}_x$ fulfilling $\mathbf{P}_x = \mathbf{S}_x \mathbf{S}_x^T$, $\mathbf{0}_{n_x}$ is a zero vector of the indicated dimension, and $\mathbf{I}_{n_x}$ is an identity matrix of the indicated dimension.
Splitting: The standard normal distribution $p(\boldsymbol{\chi}_{\mathrm{S}})$ is split into the GM (according to a splitting method given e.g., in [2–4, 10, 27–29])

$$p(\boldsymbol{\chi}_{\mathrm{S}}) \approx \hat{p}_{\mathrm{GM},N}(\boldsymbol{\chi}_{\mathrm{S}}) = \sum_{n=1}^{N} \omega^{(n)} \mathcal{N}\{\boldsymbol{\chi}_{\mathrm{S}}; \bar{\boldsymbol{\chi}}_{\mathrm{S}}^{(n)}, \mathbf{P}_{\boldsymbol{\chi}_{\mathrm{S}}}^{(n)}\} = \sum_{n=1}^{N} \omega^{(n)} \mathcal{N}^{(n)}(\boldsymbol{\chi}_{\mathrm{S}}) \quad (12)$$

with N terms, where $\mathbf{P}_{\boldsymbol{\chi}_{\mathrm{S}}}^{(n)} \leq \mathbf{I}_{n_x}$. Typically, it is assumed that the GM PDF (12) preserves the first two moments of $p(\boldsymbol{\chi}_{\mathrm{S}})$, i.e., $\mathrm{E}_{\hat{p}_{\mathrm{GM},N}(\boldsymbol{\chi}_{\mathrm{S}})}[\boldsymbol{\chi}_{\mathrm{S}}] = \mathbf{0}_{n_x}$ and $\mathrm{cov}_{\hat{p}_{\mathrm{GM},N}(\boldsymbol{\chi}_{\mathrm{S}})}[\boldsymbol{\chi}_{\mathrm{S}}] = \mathbf{I}_{n_x}$. The Gaussian PDF is split along a reference unit vector $\mathbf{r}$. Without a loss of generality, the splitting vector can be defined as $\mathbf{r} = \mathbf{e}_1$, where $\mathbf{e}_1$ is the first column of the matrix $\mathbf{I}_{n_x}$.
Rotation: The random variable $\boldsymbol{\chi}_{\mathrm{S}}$, described by the GM PDF $\hat{p}_{\mathrm{GM},N}(\boldsymbol{\chi}_{\mathrm{S}})$, is multiplied by an orthogonal matrix $\mathbf{Q}_\chi$ (fulfilling $\mathbf{Q}_\chi \mathbf{Q}_\chi^T = \mathbf{I}_{n_x}$ and $\mathbf{Q}_\chi^{-1} = \mathbf{Q}_\chi^T$) resulting in

$$\boldsymbol{\chi} = \mathbf{Q}_\chi \boldsymbol{\chi}_{\mathrm{S}} \quad (13)$$

with the GM PDF

$$\hat{p}_{\mathrm{GM},N}(\boldsymbol{\chi}) = \sum_{n=1}^{N} \omega^{(n)} \mathcal{N}\{\boldsymbol{\chi}; \mathbf{Q}_\chi \bar{\boldsymbol{\chi}}_{\mathrm{S}}^{(n)}, \mathbf{Q}_\chi \mathbf{P}_{\boldsymbol{\chi}_{\mathrm{S}}}^{(n)} \mathbf{Q}_\chi^T\}. \quad (14)$$

The orthogonal matrix $\mathbf{Q}_\chi$ is represented by a rotation or reflection matrix [41]. It is easy to verify that rotation or reflection does not affect the first two moments of (12), i.e., $\mathrm{E}_{\hat{p}_{\mathrm{GM},N}(\boldsymbol{\chi})}[\boldsymbol{\chi}] = \mathbf{0}_{n_x}$ and $\mathrm{cov}_{\hat{p}_{\mathrm{GM},N}(\boldsymbol{\chi})}[\boldsymbol{\chi}] = \mathbf{I}_{n_x}$. However, it affects the placement of the particular terms of the GM PDF and thus, the GM PDF shape. The reference unit vector $\mathbf{r}$ an the orthogonal matrix $\mathbf{Q}_\chi$ determine the splitting direction. Selection and impact of the orthogonal matrix is discussed further.
Correlation: The random variable $\boldsymbol{\chi}$ described with the (approximate) GM PDF (14) is then transformed back according to

$$\mathbf{x} = \mathbf{S}_x \boldsymbol{\chi} + \bar{\mathbf{x}}, \quad (15)$$

which is the inverse transformation to (11), resulting in

$$\hat{p}_{\mathrm{GM},N}(\mathbf{x}; \mathbf{Q}_\chi) = \sum_{n=1}^{N} \omega^{(n)} \mathcal{N}\{\mathbf{x}; \bar{\mathbf{x}}^{(n)}, \mathbf{P}_x^{(n)}\} = \sum_{n=1}^{N} \omega^{(n)} \mathcal{N}\{\mathbf{x}; \mathbf{S}_x \mathbf{Q}_\chi \bar{\boldsymbol{\chi}}_{\mathrm{S}}^{(n)} + \bar{\mathbf{x}}, \mathbf{S}_x \mathbf{Q}_\chi \mathbf{P}_{\boldsymbol{\chi}_{\mathrm{S}}}^{(n)} \mathbf{Q}_\chi^T \mathbf{S}_x^T\}, \quad (16)$$

having $\mathrm{E}_{\hat{p}_{\mathrm{GM},N}(\mathbf{x})}[\mathbf{x}] = \bar{\mathbf{x}}$ and $\mathrm{cov}_{\hat{p}_{\mathrm{GM},N}(\mathbf{x})}[\mathbf{x}] = \mathbf{P}_x$.

[¶] *Roman (non-italics) subscripts are used as a notation distinguishing different forms of PDFs. The italics letters denote variables.*

The algorithm is illustrated in Appendix.

2.3 Relation of Orthogonal Matrix and Splitting Directions

In (16), any orthogonal matrix $\mathbf{Q}_\chi$ can be used without any impact on the GM PDF first two moments. However, the matrix affects the (splitting) vectors along which the standard Gaussian PDF $p(\boldsymbol{\chi})$ and the Gaussian PDF $p(\mathbf{x})$ are split into the GM PDFs $\hat{p}_{\mathrm{GM},N}(\boldsymbol{\chi})$ (14) and $\hat{p}_{\mathrm{GM},N}(\mathbf{x})$ (16), respectively. The former vector is further called as the *intermediate splitting direction*, whereas the latter simply as the *splitting direction*. Both are defined below*.

The intermediate splitting direction of the standard Gaussian PDF $p(\boldsymbol{\chi})$ is defined as a normalized vector pointing from the mean of $p(\boldsymbol{\chi})$ (or $\hat{p}_{\mathrm{GM},N}(\boldsymbol{\chi})$) to the mean of a non-central term of $\hat{p}_{\mathrm{GM},N}(\boldsymbol{\chi})$ (14), i.e., it is defined as

$$\boldsymbol{\nu}_\chi = \mathbf{Q}_\chi \mathbf{r}. \tag{17}$$

Similarly, the splitting direction of the Gaussian PDF $p(\mathbf{x})$ is defined as a normalized vector pointing from the mean $\bar{\mathbf{x}}$ of $\hat{p}_{\mathrm{GM},N}(\mathbf{x})$ (or equivalently $p(\mathbf{x})$) to the mean of a non-central term of $\hat{p}_{\mathrm{GM},N}(\mathbf{x})$ (16), i.e., as

$$\boldsymbol{\nu}_x = \frac{\bar{\mathbf{x}} - \bar{\mathbf{x}}^{(n)}}{\|\bar{\mathbf{x}} - \bar{\mathbf{x}}^{(n)}\|_2}, \tag{18}$$

where $n \in \{1, \ldots, N\}$, $\bar{\mathbf{x}}^{(n)} \neq \bar{\mathbf{x}}$, and $\|\cdot\|_2$ is the Euclidean norm.

It can be seen that the intermediate splitting direction $\boldsymbol{\nu}_\chi$ (17) *determines* the splitting direction $\boldsymbol{\nu}_x$ (18) and vice versa. In fact, both directions are the same up to the rotation stemming from the the covariance matrix, i.e., the relation between both directions can be expressed as

$$\mathbf{U}_x \boldsymbol{\nu}_\chi = \boldsymbol{\nu}_x. \tag{19}$$

where $\mathbf{U}_x$ is the orthogonal matrix coming from the singular value decomposition (SVD) of the covariance matrix $\mathbf{P}_x$. Note that if the covariance matrix $\mathbf{P}_x$ is diagonal, then both directions (17) and (18) are identical.

2.4 Splitting Direction Selection

In principle, any splitting direction can be used in computation of GM PDF $\hat{p}_{\mathrm{GM},N}(\boldsymbol{\chi})$ (14), however, some directions may significantly reduce the approximation error $\tilde{p}_{\mathrm{GM},N}(\mathbf{y})$ (9), but other not. In the literature, various DoU, DoNLs, and DoNGs can be found.

The DoU [4] defines splitting direction vector $\boldsymbol{\nu}_x$ (18) as a (normalised) eigenvector corresponding to the largest eigenvalue of $\mathbf{P}_x$. The DoU $\boldsymbol{\nu}_x^{\mathrm{U}}$ is, therefore, a column of the orthogonal matrix $\mathbf{U}_x$ defined as

$$\boldsymbol{\nu}_x^{\mathrm{U}} = [\mathbf{U}_x]_{i^*}, \text{ with } i^* = \arg\max_i [\mathbf{D}_x]_{i,i}, \tag{20}$$

where $\mathbf{D}_x$ is a diagonal matrix with the eigenvalues of $\mathbf{P}_x$ on diagonal, i.e., $\mathbf{S}_x = \mathbf{U}_x\sqrt{\mathbf{D}_x}$, and the notation $[\mathbf{D}_x]_{i,j}$ denotes an element of the matrix $\mathbf{D}_x$ in i-th row and j-th column. The notation $[\mathbf{U}_x]_i$ stands for the i-th column of the matrix $\mathbf{U}_x$.

The DoNLs [21, 29–32] point to such a direction where the nonlinearity of the function $\mathbf{g}(\cdot)$ in (1) is, according to certain criterion, the largest. For example, the direction proposed in [29] is defined as

$$\boldsymbol{\nu}_x^{\mathrm{NL}} = \frac{\sum_{i=0}^{n_x-1} \eta_i \boldsymbol{\phi}_i}{\|\sum_{i=0}^{n_x-1} \eta_i \boldsymbol{\phi}_i\|_2}, \tag{21}$$

where $\boldsymbol{\phi}_i = \frac{\mathcal{X}_{2i+1} - \mathcal{X}_0}{\|\mathcal{X}_{2i+1} - \mathcal{X}_0\|_2}$ is the tested nonlinearity direction and $\eta_i = 0.5\|\mathcal{Y}_{2i+1} + \mathcal{Y}_{2i+2} - 2\mathcal{Y}_0\|$ is respective degree of nonlinearity being an approximation of the Hessian† of the function $\mathbf{g}(\cdot)$ in (1). The σ-points $\{\mathcal{X}_j\}_{j=0}^{2n_x}$ and the transformed σ-points $\{\mathcal{Y}_j\}_{j=0}^{2n_x}$ are defined as

$$\mathcal{X}_0 = \hat{\mathbf{x}}, \tag{22}$$

$$\mathcal{X}_{2i+1} = \bar{\mathbf{x}} + \sqrt{\kappa - n_x}[\mathbf{S}_x]_i, \tag{23}$$

$$\mathcal{X}_{2i+2} = \bar{\mathbf{x}} - \sqrt{\kappa - n_x}[\mathbf{S}_x]_i, \tag{24}$$

where $i = 0, 1, \ldots, n_x - 1$, κ is a user-defined scaling parameter, and $\mathcal{Y}_j = \mathbf{g}(\mathcal{X}_j), \forall j$. If the covariance matrix factor $\mathbf{S}_x$ is computed by the SVD, then the nonlinearity directions are aligned with the principal axes of the covariance matrix $\mathbf{P}_x$ defined by the columns of the matrix $\mathbf{U}_x$.

The DoNG [22] is designed to find a direction of the largest non-Gaussianity. The non-Gaussianity is defined using a third-order moment of the transformed random variable $\mathbf{y}$ (1). The DoNG is given by

$$\boldsymbol{\nu}_x^{\mathrm{NG}_3} = \frac{\sum_{i=1}^{n_x} \rho_i \boldsymbol{\phi}_{i-1}}{\|\sum_{i=1}^{n_x} \rho_i \boldsymbol{\phi}_{i-1}\|_2}, \tag{25}$$

where the degree of non-Gaussianity in the respective direction ρ_i is given as $\rho_i = \max\{|\mathbf{M}^{\mathrm{c}}_{y_g y_g x_i}|, |\mathbf{M}^{\mathrm{c}}_{y_g x_i x_i}|\}$. The sets of third-order joint moments of $\mathbf{x}$ and $\mathbf{y}$ are defined

$$\mathbf{M}^{\mathrm{c}}_{y_g y_g x_i} = \mathrm{E}_{p(\mathbf{x})}[(\mathbf{g}(\mathbf{x}) - \bar{\mathbf{y}}_{\mathbf{g}})(\mathbf{g}(\mathbf{x}) - \bar{\mathbf{y}}_{\mathbf{g}})^T(x_i - \bar{x}_i)], \tag{26}$$

$$\mathbf{M}^{\mathrm{c}}_{y_g x_i x_i} = \mathrm{E}_{p(\mathbf{x})}[(\mathbf{g}(\mathbf{x}) - \bar{\mathbf{y}}_{\mathbf{g}})(x_i - \bar{x}_i)^2], \tag{27}$$

where x_i is i-th element of the vector $\mathbf{x}$. The moments (26), (27) cannot be generally computed exactly and an approximation based on, e.g., the unscented transformation (UT) or numerical integration rules, should be used.

The PDF $p(\mathbf{y})$, and thus the non-Gaussianity of $\mathbf{y}$ (5), is determined by the properties of the to-be-transformed random variable $\mathbf{x}$, i.e., by the PDF $p(\mathbf{x})$, and characteristics of the nonlinear function $\mathbf{g}(\cdot)$. However, the DoU $\boldsymbol{\nu}_x^{\mathrm{U}}$ (20) takes into account properties of $\mathbf{x}$ only and the DoNL $\boldsymbol{\nu}_x^{\mathrm{NL}}$ (21) uses mainly the properties of the nonlinear function $\mathbf{g}(\cdot)$. The DoNG $\boldsymbol{\nu}_x^{\mathrm{NG}_3}$ (25) assesses directly the properties of the transformed variable $\mathbf{y}$ but it ignores the "pure" third-order moments of $\mathbf{y}$, i.e., $M^{\mathrm{c}}_{y_i y_j y_k}$, which are expected to contain a significant portion of information regarding the non-Gaussianity of $\mathbf{y}$. Moreover, note that all the introduced splitting directions are influenced by the scale factor (units) of the individual elements of the variables $\mathbf{x}$ and $\mathbf{y}$, which is inconvenient.

2.5 Goal of the Paper

The *goal* of the paper is to design a technique for computation of the splitting direction of the Gaussian PDF $p(\mathbf{x})$ based on a *directional* assessment of the moments of the transformed random variable $\mathbf{y}$. The proposed directions, which extends the family of DoNGs, are based on the normalised third-order moment, thus, they are unitless. The directions allows efficient and computationally inexpensive splitting of the Gaussian random variable PDF resulting, in the end, in a smaller error of the PDF of nonlinearly transformed Gaussian variable.

**In a view of introduced terminology, the reference unit vector* $\mathbf{r}$ *can be called as the basic splitting direction.*

†The Hessian of the nonlinear function is related to the curvature of the nonlinear function [42].

3 Computation of Skewness-based Splitting Direction

The proposed splitting directions are based on the measures of non-Gaussianity. The MoNGs have been proposed and utilised for a few last decades to quantify (or measure) a distance of a non-Gaussian PDF from its, in some sense optimal, Gaussian approximation [12, 43–45]. With respect to the notation of this paper, a MoNG assesses a difference between $p(\mathbf{y})$ (1) and $\hat{p}_{\mathrm{G}}(\mathbf{y})$ (2), i.e., the error $\tilde{p}_{\mathrm{G}}(\mathbf{y})$ (5). Compared to the MoNLs, the MoNGs, thus, asses the effect of the nonlinear transformation on the Gaussian variable.

In this paper, the *normalised* third-order central moment, denoted as (co)-skewness [46], is selected and computed along the principal axes of the to-be-transformed variable. Based on the central moment values, the splitting direction is then computed.

3.1 Normalised Third-Order Moment

The third-order central moment of $\mathbf{y}$ is defined as [47]

$$\mathbf{M}_y^{\mathrm{c}} = \mathrm{E}_{p(\mathbf{x})}[(\mathbf{g}(\mathbf{x}) - \bar{\mathbf{y}}_{\mathrm{g}})(\mathbf{g}(\mathbf{x}) - \bar{\mathbf{y}}_{\mathrm{g}})^T \otimes (\mathbf{g}(\mathbf{x}) - \bar{\mathbf{y}}_{\mathrm{g}})^T] + \mathbf{M}_v^{\mathrm{c}} \tag{28}$$

where the symbol $\otimes$ denotes the Kronecker product [47] and $\mathbf{M}_v^{\mathrm{c}} = \mathbf{0}_{n_y \times n_y^2}$ due to the Gaussian PDF $p(\mathbf{v})$. The matrix $\mathbf{M}_y^{\mathrm{c}} \in \mathbb{R}^{n_y \times n_y^2}$ (28) can be understood as a set of all $N_{yyy} = \binom{n_y - 2}{3}$ partial third-order central moments

$$M^{\mathrm{c}}_{y_i y_j y_k} = [\mathbf{M}_y^{\mathrm{c}}]_{i, j\times n_y + k} = \mathrm{E}_{p(\mathbf{x})}[(y_i - \bar{y}_i)(y_j - \bar{y}_j)(y_k - \bar{y}_k)], \tag{29}$$

where $i, j, k \in \{1, 2, \ldots, n_y\}$. The (co-)skewness $\mathbf{M}_y \in \mathbb{R}^{n_y \times n_y^2}$ of the random variable $\mathbf{y}$ is then defined by a set of all N_{yyy} normalised third-order central moments

$$M_{y_i y_j y_k} = [\mathbf{M}_y]_{i, j\times n_y + k} = \frac{M^{\mathrm{c}}_{y_i y_j y_k}}{\sqrt{[\mathbf{P}_y]_{i,i}[\mathbf{P}_y]_{j,j}[\mathbf{P}_y]_{k,k}}}. \tag{30}$$

The normalized moment (30) is *unit-less* and it can take any positive or negative value, although in practical settings it rarely exceeds 3 in the absolute value [44]. Note that the (co)-skewness of a Gaussian PDF is zero. The skewness is, thus, a suitable measure of non-Gaussianity of a transformed random variable PDF $p(\mathbf{y})$. The skewness-based MoNG can be defined as

$$\mu_{\mathrm{G}} \triangleq \max_{\{i,j,k\}} |M_{y_i y_j y_k}|. \tag{31}$$

The PDF is considered to be far from a Gaussian PDF $\hat{p}_{\mathrm{G}}(\mathbf{y}) = \mathcal{N}\{\mathbf{y} : \bar{\mathbf{y}}, \mathbf{P}_y\}$ (2) if the MoNG μ_{G} (31) is greater than a pre-defined threshold. By splitting the PDF $p(\mathbf{x})$ into the GM PDF $\hat{p}_{\mathrm{GM},N}(\mathbf{x}; \mathbf{Q}_\mathcal{X})$ (16), the MoNG related to the GM approximation of $p(\mathbf{y})$, i.e., to $\hat{p}_{\mathrm{GM},N}(\mathbf{y})$, defined as

$$\mu_{\mathrm{GM}}(\mathbf{Q}_\mathcal{X}) \triangleq \max_n \max_{\{i,j,k\}} |M^{(n)}_{y_i y_j y_k}| = \max_n \mu_{\mathrm{G}}^{(n)}, \tag{32}$$

where $M^{(n)}_{y_i y_j y_k}$ is the skewness of the n-th term of $\hat{p}_{\mathrm{M},N}(\mathbf{y})$ (7), should be lower than μ_{G} (30). The aim is, therefore, to find either a rotation (or reflection) matrix $\mathbf{Q}_\mathcal{X}$ or the respective rotation vector $\boldsymbol{\nu}_\mathcal{X}$ (17) that leads to the *smallest* $\mu_{\mathrm{GM}}(\mathbf{Q}_\mathcal{X})$ (32). Below, two techniques for DoNG computation are developed, which differ in accuracy and computational complexity.

3.2 Skewness-based Direction of Non-Gaussianity

The (exact) DoNG is defined as

$$\boldsymbol{\nu}_x^{\mathrm{NG_S}} = \mathbf{U}_x \mathbf{Q}_\mathcal{X}^{\mathrm{NG_S}} \mathbf{r}, \tag{33}$$

where

$$\mathbf{Q}_\mathcal{X}^{\mathrm{NG_S}} = \arg\min_{\mathbf{Q}_\mathcal{X}} \mu_{\mathrm{GM}}(\mathbf{Q}_\mathcal{X}). \tag{34}$$

By the minimisation of the criterion (34) such an orthogonal (rotation or reflection) matrix used in the construction of the GM PDF $\hat{p}_{\mathrm{GM},N}(\mathbf{x}; \mathbf{Q}_\mathcal{X})$ (16) is sought, which results in the GM PDF $\hat{p}_{\mathrm{M},N}(\mathbf{y}) = \hat{p}_{\mathrm{M},N}(\mathbf{y}; \mathbf{Q}_\mathcal{X}^{\mathrm{NG_S}})$ (7) with the minimum value of $\mu_{\mathrm{GM}}(\mathbf{Q}_\mathcal{X})$ among all possible $\hat{p}_{\mathrm{M},N}(\mathbf{y}; \mathbf{Q}_\mathcal{X})$ of the same structure. In other words, the mixture PDF $\hat{p}_{\mathrm{M},N}(\mathbf{y}; \mathbf{Q}_\mathcal{X}^{\mathrm{NG_S}})$ (7), where all terms are closer to Gaussian PDFs, is selected and then reasonably well approximated by the Gaussian mixture PDF $\hat{p}_{\mathrm{GM},N}(\mathbf{y}; \mathbf{Q}_\mathcal{X}^{\mathrm{NG_S}})$ (8). Therefore, the DoNG (33) directly minimises the measure (31). However, an analytical solution to (34) does not exist and a numerical evaluation shall be used instead.

Any numerical solution requires evaluation of the criterion (34), i.e., evaluation of the measure $\mu_{\mathrm{GM}}(\mathbf{Q}_\mathcal{X})$, for multiple choices of $\mathbf{Q}_\mathcal{X}$, which may be impractical even for low-dimensional set-ups; for each of N terms of the mixture PDF $\hat{p}_{\mathrm{M},N}(\mathbf{y}; \mathbf{Q}_\mathcal{X})$ (7), $N_{yyy} = \binom{n_y - 2}{3}$ third-order moments (30) of $\mathbf{y}$ have to be evaluated. It means $N \times N_{yyy}$ evaluations of $M_{y_i y_j y_k}$ (30) per a single choice of $\mathbf{Q}_\mathcal{X}$.

It should also be mentioned that a direct minimisation of the criterion (34) w.r.t. the matrix $\mathbf{Q}_\mathcal{X}$ is quite difficult and should be avoided. The reason is twofold; the rotation (or reflection) matrix has to be an orthogonal matrix, thus it is difficult to perform the minimisation respecting the matrix structure, and n_x^2 elements need to be found. Instead, the rotation matrix may be parametrised by a set of $n_\theta = n_x(n_x - 1)/2$ rotation angles $\theta_1, \ldots, \theta_{n_\theta}$ [48]. The rotation angles correspond to a minimal representation of a general rotation in $\mathbb{R}^{n_x}$.

3.3 Approximate Skewness-based Direction of Non-Gaussianity

To reduce the computational complexity of the DoNG computation, an *empirical* determination of the DoNG $\boldsymbol{\nu}_x$ (18) is proposed. The proposed efficient approximate direction is based on the quantification and assessment of the MoNG along each principal direction of $p(\mathbf{x})$. The principal directions of $p(\mathbf{x})$ are given by the eigenvectors of the $\mathbf{P}_x$ further denoted as $\mathbf{u}_i = [\mathbf{U}_x]_i$, $i = 1, \ldots, n_x$.

The skewness-based DoNG is motivated by the following idea. Suppose that the Gaussian PDF $p(\mathbf{x})$ is "squeezed", i.e., the covariance matrix $\mathbf{P}_x$ is reduced, along a principal direction, random variable $\mathbf{x}$ described by the squeezed PDF is transformed via (1), and an indicative MoNG of the transformed variable is computed. If the indicative MoNG is significantly lower than the MoNG based on the original PDF $p(\mathbf{x})$, i.e., than μ_{G} (31), then splitting of the Gaussian PDF $p(\mathbf{x})$ along this direction is reasonable; it may significantly reduce the MoNG of the resulting GM PDF, i.e., reduce $\mu_{\mathrm{GM}}(\mathbf{Q}_\mathcal{X})$ (32). On the other hand, if squeezing of the PDF $p(\mathbf{x})$ does not lead to the MoNG reduction and the indicative MoNG is of the same value as μ_{G}, then splitting of $p(\mathbf{x})$ along this direction may lead in negligible reduction of $\mu_{\mathrm{GM}}(\mathbf{Q}_\mathcal{X})$ w.r.t. μ_{G}.

The indicative MoNG along the direction of the eigenvector corresponding to the n-th element of the vector $\mathbf{x}$, $n = 1, \ldots, n_x$, is defined as

$$\mu_{\mathrm{G},n} = \max_{i,j,k} \frac{M_{y_i y_j y_k}}{M_{y_i y_j y_k, n}} \tag{35}$$

where $i, j, k = 1, \ldots, n_y$ and $M_{y_i y_j y_k, n}$ is the skewness of $p_j(\mathbf{y})$ computed for a Gaussian PDF "squeezed" in the direction of the n-th

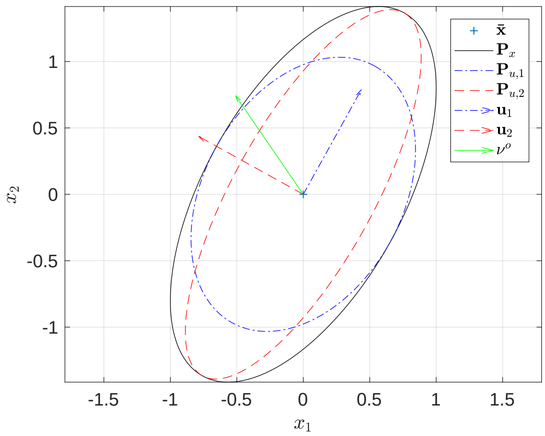

Fig. 3: Illustration of a covariance matrix squeezed along the principal directions.

eigenvector, further denoted as

$$p_n(\mathbf{x}) = \mathcal{N}\{\mathbf{x}; \bar{\mathbf{x}}, \mathbf{P}_{u,n}\} \tag{36}$$

with

$$\mathbf{P}_{u,n} = \mathbf{S}_x \mathbf{P}_{\chi,n} \mathbf{S}_x^T, \tag{37}$$

$$\mathbf{P}_{\chi,n} = \mathbf{I}_{n_x} - \delta \mathbf{e}_n \mathbf{e}_n^T, \tag{38}$$

and $\delta \in (0,1)$. That is $\mathbf{P}_{u,n} \leq \mathbf{P}_x$. As a consequence, $M_{y_i y_j y_k, n}$ related to $\mathbf{P}_{u,n}$ is typically lower than $M_{y_i y_j y_k}$ based on $\mathbf{P}_x$. If $\left|\frac{M_{y_i y_j y_k}}{M_{y_i y_j y_k, n}}\right| < 1$, then $\mu_{\mathrm{G},n}$ is set to zero as this direction does not result in reduction of the non-Gaussianity. Illustration of "squeezing" of the covariance matrix $\mathbf{P}_x$ for a two-dimensional variable $\mathbf{x}$ along the principal directions of $p(\mathbf{x})$ given by eigenvectors $\mathbf{u}_n$, $n = 1, 2$, of $\mathbf{P}_x$, i.e., $\mathbf{P}_{u,n}$, is given in Fig. 3.

The required splitting direction is defined as the MoNG weighted average of the eigenvectors, i.e., as

$$\boldsymbol{\nu}_x^{\mathrm{NG_{SA}}} = \sum_{n=1}^{n_x} \frac{\mu_{\mathrm{G},n}}{\|[\mu_{\mathrm{G},1}, \mu_{\mathrm{G},2}, \ldots, \mu_{\mathrm{G},n_x}]\|} \mathbf{u}_n. \tag{39}$$

The proposed splitting direction (39) is based on the belief that the direction $\boldsymbol{\nu}_x^{\mathrm{NG_{SA}}}$ (39) points in such a direction, where splitting of $p(\mathbf{x})$ into $\hat{p}_{\mathrm{GM},N}(\mathbf{x}; \mathbf{Q}_\chi^{\mathrm{NG_{SA}}})$ leads to a significant decrease of $\mu_{\mathrm{GM}}(\mathbf{Q}_\chi^{\mathrm{NG_{SA}}})$ (32) of $\hat{p}_{\mathrm{M},N}(\mathbf{y})$ (7). To compute the direction $\boldsymbol{\nu}_x^{\mathrm{NG_{SA}}}$, only $n_y \times n_x$ evaluations of the third-order moment are required. Note that a possible DoNG $\boldsymbol{\nu}_x^{\mathrm{NG_{SA}}}$ is illustrated in Fig. 3.

To be able to compute the GM PDF $\hat{p}_{\mathrm{GM},N}(\mathbf{x}; \mathbf{Q}_\chi)$ (16), it is necessary to translate the splitting direction $\boldsymbol{\nu}_x^{\mathrm{NG_{SA}}}$ (39) into an orthogonal matrix $\mathbf{Q}_\chi^{\mathrm{NG_{SA}}}$. According to (19), the splitting direction $\boldsymbol{\nu}_x^{\mathrm{NG_{SA}}}$ and the matrix $\mathbf{Q}_\chi^{\mathrm{NG_{SA}}}$ are tied via the additional rotation caused by the $\mathbf{P}_x$ in the correlation step of Algorithm 1 within (16). In particular, their relation can be expressed as

$$\mathbf{S}_x \mathbf{Q}_\chi^{\mathrm{NG_{SA}}} \mathbf{r} \propto \boldsymbol{\nu}_x^{\mathrm{NG_{SA}}}. \tag{40}$$

From (40), it follows that

$$\mathbf{Q}_\chi^{\mathrm{NG_{SA}}} \mathbf{r} = \underbrace{\frac{\mathbf{S}_x^{-1} \boldsymbol{\nu}_x^{\mathrm{NG_{SA}}}}{\|\mathbf{S}_x^{-1} \boldsymbol{\nu}_x^{\mathrm{NG_{SA}}}\|}}_{\triangleq \boldsymbol{\nu}_{S_x}^{\mathrm{NG_{SA}}}}, \tag{41}$$

where $\mathbf{r}$, $\boldsymbol{\nu}_{S_x}^{\mathrm{NG_{SA}}}$ are known *unit* vectors and $\mathbf{Q}_\chi^{\mathrm{NG_{SA}}}$ is the sought unknown orthogonal matrix. In principle, there exist infinite many matrices $\mathbf{Q}_\chi^{\mathrm{NG_{SA}}}$ fulfilling the equality (41). A possible solution stems from the Householder reflection, which results in the reflection matrix [41]

$$\mathbf{Q}_\chi^{\mathrm{NG_{SA}}} = \mathbf{I}_{n_x} - 2\frac{(\mathbf{r} - \boldsymbol{\nu}_{S_x}^{\mathrm{NG_{SA}}})(\mathbf{r} - \boldsymbol{\nu}_{S_x}^{\mathrm{NG_{SA}}})^T}{(\mathbf{r} - \boldsymbol{\nu}_{S_x}^{\mathrm{NG_{SA}}})^T(\mathbf{r} - \boldsymbol{\nu}_{S_x}^{\mathrm{NG_{SA}}})}. \tag{42}$$

Another solution is based on the construction of a rotation matrix $\mathbf{Q}_\chi^{\mathrm{NG_{SA}}}$ based on the direction cosine matrix (DCM) or the Rodrigues rotation formula [48].

Finally note that an alternative and even more efficient splitting direction, inspired by the idea given in [29], can be defined as a principal direction of $p(\mathbf{x})$, i.e., $\mathbf{u}_j$, which corresponds to maximal $\mu_{\mathrm{G},j}$, i.e., the one fulfilling $\mu_{\mathrm{G},j} \geq \mu_{\mathrm{G},i}, \forall i$. As consequence, the respective matrix $\mathbf{Q}_\chi^{\mathrm{NG_{SA}}}$ is a permutation matrix.

3.4 MoNG-based Splitting Algorithm Summary

Let a Gaussian PDF $p(\mathbf{x}) = \mathcal{N}\{\mathbf{x}; \bar{\mathbf{x}}, \mathbf{P}_x\}$ be supposed. Then, the algorithm for the proposed MoNG-based Gaussian PDF splitting is given by the following steps:

Algorithm 2: MoNG-based Directional Splitting of a Gaussian PDF $p(\mathbf{x})$

Non-Gaussianity detection: Compute the first three moments of the transformed variable $\mathbf{y}$ (1), i.e., the mean $\bar{\mathbf{y}}$ (3), the covariance matrix $\mathbf{P}_y$ (4), and the (co-)skewness $\mathbf{M}_y$ (30), and compute the MoNG (31). The Gaussian PDF $\hat{p}_{\mathrm{G}}(\mathbf{y})$ (2) is considered to be a *sufficiently accurate* approximation to $p(\mathbf{y})$ if $\mu_{\mathrm{G}} \leq \mathrm{MoNG_{thld}}$, where $\mathrm{MoNG_{thld}} \in \mathbb{R}$ is a defined threshold (discussed below). If $\mu_{\mathrm{G}} \leq \mathrm{MoNG_{thld}}$, then the algorithm stops. If $\mu_{\mathrm{G}} \geq \mathrm{MoNG_{thld}}$, then a significant non-Gaussianity of $p(\mathbf{y})$ is detected and the algorithm continues with the next step.
Non-Gaussianity direction computation: Determine the direction defining the splitting direction of the Gaussian PDF $p(\mathbf{x})$ which results in the largest decrease of the MoNG. The direction is defined by the rotation matrix $\mathbf{Q}_\chi$ in $\hat{p}_{\mathrm{GM},N}(\mathbf{x}; \mathbf{Q}_\chi)$. The matrix $\mathbf{Q}_\chi$ can be computed directly using the optimisation of (34), then $\mathbf{Q}_\chi = \mathbf{Q}_\chi^{\mathrm{NG_S}}$, or indirectly using optimisation of (39) and evaluation of $\mathbf{Q}_\chi^{\mathrm{NG_{SA}}}$ (42), then $\mathbf{Q}_\chi = \mathbf{Q}_\chi^{\mathrm{NG_{SA}}}$.
Gaussian PDF directional splitting: Find a GM PDF approximation $\hat{p}_{\mathrm{GM},N}(\mathbf{x}; \mathbf{Q}_\chi)$ (16) of $p(\mathbf{x})$ using Algorithm 1.
GM PDF transformation: Transform the variable $\mathbf{x}$ described by $\hat{p}_{\mathrm{GM},N}(\mathbf{x}; \mathbf{Q}_\chi)$ (8) via the nonlinear transformation (1), which leads to the mixture PDF $\hat{p}_{\mathrm{M},N}(\mathbf{y}; \mathbf{Q}_\chi)$ (7). Approximation of the mixture PDF by the GM PDF, results in the GM PDF of the variable $\mathbf{y}$ defined by $\hat{p}_{\mathrm{GM},N}(\mathbf{y}; \mathbf{Q}_\chi)$ (8) where each term $\mathcal{N}^{(n)}(\mathbf{y})$ is associated with $\mu_{\mathrm{G}}^{(n)}$ lower than MoNG respective to $p(\mathbf{y})$.
GM PDF verification: Compute the MoNG for $\hat{p}_{\mathrm{GM},N}(\mathbf{y}; \mathbf{Q}_\chi)$, i.e., $\mu_{\mathrm{GM}}(\mathbf{Q}_\chi)$ (32). If $\mu_{\mathrm{GM}}(\mathbf{Q}_\chi) \leq \mathrm{MoNG_{thld}}$, the algorithm stops. Otherwise, the algorithm continues by the step "Non-Gaussianity direction computation" for each particular term of $\hat{p}_{\mathrm{M},N}(\mathbf{y}; \mathbf{Q}_\chi)$ having MoNG greater than the pre-defined threshold $\mathrm{MoNG_{thld}}$.

3.5 Notes on DoNG

Note 1: The proposed DoNGs are based on computation of the first three moments of the nonlinearly transformed random variable. The moments are typically computed using approximate techniques involving σ-point based transformations [1, 36, 39]. Nevertheless, for certain nonlinear functions $\mathbf{g}(\cdot)$ in (1), mainly for polynomial functions, the moments can be computed *exactly* by a numerical integration rules [39].

Note 2: The skewness-based DoNGs and MoNG explicitly take into account the influence of the additive Gaussian variable $\mathbf{v}$ on $p(\mathbf{y})$.

Note 3: Thorough non-Gaussianity assessment of $p(\mathbf{y})$ would require computation of infinite many moments of $p(\mathbf{y})$, which is impractical. The skewness, as a one of infinite many moments, is typically considered to be sufficient for non-Gaussianity assessment of $p(\mathbf{y})$ in practical situations [12, 43, 45]. Nevertheless, the higher order moments of $p(\mathbf{y})$ can analogously be monitored and used for DoNG computation. However, a problem with using higher order moments may rest in their computation using a limited set of σ-points. Thus, without increasing the number of σ-points (imposing higher computational complexity) their computation may be of limited accuracy.

Note 4: The proposed directional splitting algorithm allows, through the step "GM PDF verification", recursive splitting of each GM term of $\hat{p}_{\mathrm{GM},N}(\mathbf{x};\mathbf{Q}_\chi)$ which causes $\mu_{\mathrm{GM}}(\mathbf{Q}_\chi)$ being greater than the threshold. Generally, the splitting direction (and also the number of Gaussian terms in GM PDF) can be different for each split term. An alternative approach to the recursive splitting is the splitting along multiple directions simultaneously which was discussed e.g., in [10, 21]. Extension of the proposed skewness-based splitting algorithm for multi-directional splitting is straightforward.

Note 5: By application of the splitting technique, the number of GM components grows, which might be a limiting factor w.r.t. allowed computational load. To reduce the number of GM components, various techniques, proposed e.g., in [3, 4], can be used.

4 Numerical Illustration

The splitting directions are illustrated using the example of transformation of polar coordinates to Cartesian coordinates. The transformation is defined as

$$\mathbf{y} = \begin{bmatrix} y_1 \ [km] \\ y_2 \ [km] \end{bmatrix} = \begin{bmatrix} x_1 \cos(x_2) \\ x_1 \sin(x_2) \end{bmatrix}, \qquad (43)$$

where the to-be-transformed variable $\mathbf{x} = [x_1 \ [km], x_2 \ [rad]]^T$ is described by the Gaussian PDF with moments

$$\bar{\mathbf{x}} = \begin{bmatrix} 5 \\ \bar{x}_2 \end{bmatrix}, \ \mathbf{P}_x = \begin{bmatrix} 0.2^2 & 0 \\ 0 & (12\frac{\pi}{180})^2 \end{bmatrix} \qquad (44)$$

and $\bar{x}_2 \in (0, \pi/2)[rad]$. Four splitting directions are compared, namely

(i) DoU $\boldsymbol{\nu}_x^{\mathrm{U}}$ (20),
(ii) DoNL $\boldsymbol{\nu}_x^{\mathrm{NL}}$ (21),
(iii) the *proposed* efficient DoNGs $\boldsymbol{\nu}_x^{\mathrm{NG_{SA}}}$ (39), and
(iv) the optimal direction $\boldsymbol{\nu}_x^{\mathrm{opt}}$ defined below.

According to Algorithm 1, the Gaussian PDF $p(\mathbf{x})$ is split into the GM PDF $\hat{p}_{\mathrm{GM},N}(\mathbf{x};\mathbf{Q}_\chi)$ (14), where the rotation matrix is computed as

$$\mathbf{Q}_\chi = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \qquad (45)$$

and the rotation angle is computed for each particular splitting direction $\boldsymbol{\nu} = [\nu_1, \nu_2]^T$ (considering the reference unit vector $\mathbf{r} = \mathbf{e}_1$) according to

$$\theta = \arctan(\frac{\nu_2}{\nu_1}). \qquad (46)$$

Note that the construction of the GM PDF of the normalised Gaussian PDF $\hat{p}_{\mathrm{GM},N}(\boldsymbol{\chi}_{\mathrm{s}})$ (12) is based on the method provided in [27] with $N = 3$.

The suitability of the splitting direction is assessed using the integral error (IE) of approximate GM PDF $\hat{p}_{\mathrm{GM},3}(\mathbf{y};\mathbf{Q}_\chi)$, i.e., $\tilde{p}_{\mathrm{GM},3}(\mathbf{y})$ (9), given by

$$\tilde{p}^{\mathrm{IE}}_{\mathrm{GM},3} = \int (\tilde{p}_{\mathrm{GM},3}(\mathbf{y};\mathbf{Q}_\chi))^2 d\mathbf{y}. \qquad (47)$$

The *lower* the value of the IE is, the *better* GM PDF $\hat{p}_{\mathrm{GM},3}(\mathbf{y};\mathbf{Q}_\chi)$ approximation of the true $p(\mathbf{y})$ is. Thus, the optimal splitting direction $\boldsymbol{\nu}_x^{\mathrm{opt}}$ is selected according to

$$\boldsymbol{\nu}_x^{\mathrm{opt}} = \arg\min_{\boldsymbol{\nu}} \tilde{p}^{\mathrm{IE}}_{\mathrm{GM},3}. \qquad (48)$$

Hence, the optimal direction $\boldsymbol{\nu}_x^{\mathrm{opt}}$ results in the minimal* IE of the approximate GM PDF $\hat{p}_{\mathrm{GM},3}(\mathbf{y};\mathbf{Q}_\chi)$. Note that the criterion $\tilde{p}^{\mathrm{IE,opt}}_{\mathrm{GM},3}$ (47) was evaluated using the Monte-Carlo integration with 10^7 samples.

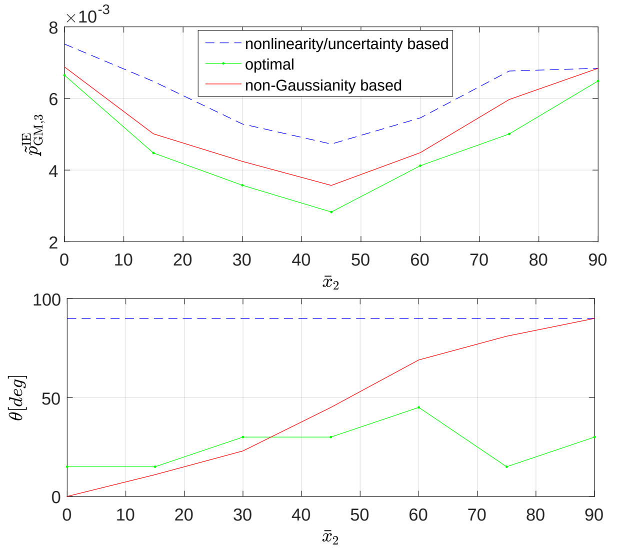

Fig. 4: Illustration of integral error of particular GM PDF and splitting direction.

Table 1 Computational complexity for splitting directions computation.

	computational complexity $[sec]$	averaged $\tilde{p}^{\mathrm{IE}}_{\mathrm{GM},3}$
DoU	0	6.2×10^{-3}
DoNL	0.01	6.2×10^{-3}
DoNG	0.07	5.3×10^{-3}
optimal	2	4.7×10^{-3}

The performance of four splitting directions in terms of the IE and the respective angles θ for varying $\bar{x}_2$ are plotted in Fig. 4. It can be seen, that the DoU $\boldsymbol{\nu}_x^{\mathrm{U}}$ (20) and DoNL $\boldsymbol{\nu}_x^{\mathrm{NL}}$ (21) are the same with the angle $\theta = 90[deg]$ and constant for all values of $\bar{x}_2$. On the other hand, the *proposed* DoNG $\boldsymbol{\nu}_x^{\mathrm{NG_{SA}}}$ (39) and the optimal one $\boldsymbol{\nu}_x^{\mathrm{opt}}$ (48) *varies* with $\bar{x}_2$. The figure also indicates that the DoNG results in *lower* IE then the standard DoU and DoNL (ca. 20 per cent). The optimal splitting direction provides even lower IE, however, with *much higher* computational complexity. The computational complexity is shown in Table 1 together with averaged values of the IE $\tilde{p}^{\mathrm{IE}}_{\mathrm{GM},3}$ (47) over all values of $\bar{x}_2$.

It can be seen that the computation of the optimal direction $\boldsymbol{\nu}_x^{\mathrm{opt}}$ (48) is expensive as it relies on the numerical optimisation concerning the whole PDFs, i.e., $p(\mathbf{y})$ has to be approximately computed. In

**Gaussian PDF can be split in direction $\theta \in (0, 360)[deg]$. However, because symmetry, it is sufficient to find the splitting direction in the interval $\theta \in (0, 180)[deg]$.*

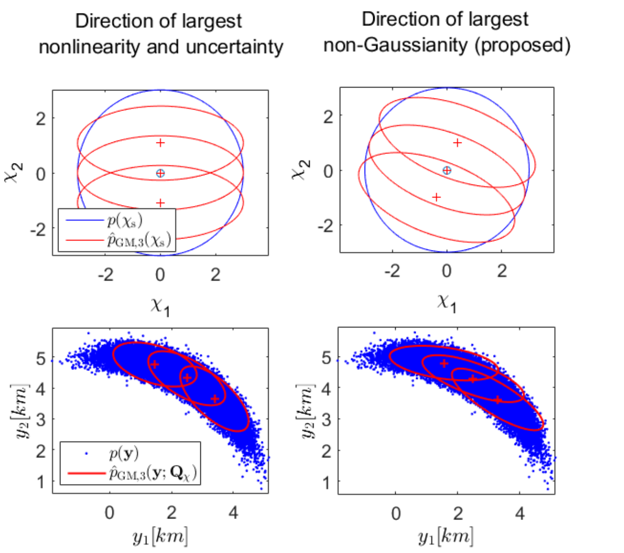

Fig. 5: Splitting direction impact on location of GM PDF terms for $\bar{x}_2 = 60[deg]$.

the considered scenario, computation of the optimal direction took almost $2[sec]$ per a value of $\bar{x}_2$. A significant reduction of the computational complexity can be achieved by utilisation of the DoU, DoNL, and DoNG. From these DoU $\boldsymbol{\nu}_x^{\mathrm{U}}$ (20) is the most efficient as it is based on the assessment of the covariance matrix $\mathbf{P}_x$ only (and it is the same for all $\bar{x}_2$). The DoNL $\boldsymbol{\nu}_x^{\mathrm{NL}}$ (21) is based on the evaluation of the nonlinear function in (43) at 5 points and subsequent processing, which is fast. The computational complexity of the DoNG is largely affected by the used integration method for the moment computation. In this example, the stochastic integration rule (or alternatively the randomised unscented transformation) [39] was used which resulted in higher complexity than the DoNL. The DoNL and DoNG complexity is, however, of the same order. Note also that the splitting direction computation is just one part of Algorithm 1 used in the step "Rotation". The remaining steps, which are the same for all splitting direction, took about 0.01 $[sec]$. The numerical evaluation of the IE (47) required approximately 5 $[sec]$.

For the sake of completeness, a visual illustration of different DoU/DoNL and DoNG for one particular value $\bar{x}_2 = 60[deg]$ is given in Fig. 5. In the figure, the top plots show $p(\boldsymbol{\chi}_\mathrm{s})$ in (11) and the location of the particular terms of its GM PDF approximation $\hat{p}_{\mathrm{GM},N}(\boldsymbol{\chi})$ (14). The bottom plots show the GM terms of $\hat{p}_{\mathrm{GM},3}(\mathbf{y};\mathbf{Q}_\chi)$ together with the points representing the "true" PDF $p(\mathbf{y})$.

5 Concluding Remarks

In the paper, the splitting of the Gaussian density of a random variable, which is supposed to undergo a nonlinear transformation, into the Gaussian mixture PDF was discussed. The aim of the splitting is to reduce error of Gaussian approximation of the PDF of the transformed random variable. Novel approach for the splitting direction specification was proposed. The resulting directions are based on a directional assessment of the moments, namely the co-skewness, of the transformed Gaussian variable. Compared to the other splitting directions, the introduced ones try to directly minimise the non-Gaussianity of transformed Gaussian mixture terms, which is a vital assumption of many recursive algorithms for estimation and decision making. The proposed directions have been analysed and compared using a numerical illustration. It was shown that employment of the skewness-based splitting direction allows to reduce an approximation error of Gaussian-mixture PDF.

Acknowledgement

This work was supported by the German Research Foundation (DFG) under grant HA 3789/14-1, the Czech Science Foundation (GAČR) under grant GA 16-19999J, and the Czech Ministry of Education, Youth and Sports under grant LO1506.

6 References

1 B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Prentice Hall, New Jersey, 1979.
2 H. W. Sorenson and D. L. Alspach, "Recursive Bayesian estimation using Gaussian sums," *Automatica*, vol. 7, pp. 465–479, 1971.
3 J. L. Williams and P. S. Maybeck, "Cost-function-based Gaussian mixture reduction for target tracking," in *Proceedings of the 6th International Conference on Information Fusion*, Queensland, Australia, 7 2003.
4 F. Faubel, J. McDonough, and D. Klakow, "The split and merge unscented Gaussian mixture filter," *IEEE Signal Processing Letters*, vol. 16, no. 9, pp. 786–789, 2009.
5 T. S. Söderström and P. G. Stoica, *System Identification*. Prentice Hall, 1989.
6 L. Král and M. Šimandl, "Neural network based bicriterial dual control with multiple linearization," in *Proceedings of the 10th IFAC Workshop on Adaptation and Learning in Control and Signal Processing*, Antalya, Turkey, 8 2010.
7 M. Flídr, O. Straka, and M. Šimandl, "Pruning and merging strategies in receding horizon bicriterial dual controller with multiple linearization," in *Proceedings of the 19th Mediterranean Conference on Control & Automation*, Corfu, Greece, 6 2011.
8 S. X. Ding, *Model-Based Fault Diagnosis Techniques: Design Schemes, Algortihms and Tools*. Springer, 2008.
9 D. R. Gizda, P. Singla, and M. K. Jah, "An approach for nonlinear uncertainty propagation: Application to orbital mechanics," in *Proceedings of the AIAA Guidance, Navigation, and Control Conference*, Aug. 2009.
10 K. J. DeMars, R. H. Bishop, and M. K. Jah, "Entropy-based approach for uncertainty propagation of nonlinear dynamical systems," *Journal of Guidance, Control, and Dynamics*, vol. 36, no. 4, pp. 1047–1057, 2013.
11 K. Reif, S. Günther, E. Yaz, and R. Unbehauen, "Stochastic stability of the discerete-time extended Kalman filter," *IEEE Transactions on Automatic Control*, vol. 44, no. 4, pp. 714–728, 1999.
12 J. Duník and M. Šimandl, "Nonlinearity and non-Gaussianity measures for stochastic dynamic systems," in *Proceedings of the 16th International Conference on Information Fusion*, Istanbul, 2013.
13 Y. Liu and X. R. Li, "Measure of nonlinearity for estimation," *IEEE Transactions on Signal Processing*, vol. 63, no. 9, pp. 2377–2388, 2015.
14 J. Duník, O. Straka, M. Mallick, and E. Blasch, "Survey of nonlinearity and non-Gaussianity measures for state estimation," in *Proceedings of the 19th International Conference on Information Fusion*, Heidelberg, Germany, Jul. 2016.
15 M. Raitoharju, R. Piché, J. Ala-Luhtala, and S. Ali-Löytty, "Partitioned update Kalman filter," *Journal of Advances in Information Fusion*, vol. 11, no. 1, pp. 3–14, 2016.
16 M. Raitoharju, A. F. García-Fernández, and R. Piché, "Kullback âĂŞ- Leibler divergence approach to partitioned update Kalman filter," *Signal Processing*, vol. 130, pp. 289–298, 2017.
17 G. Terejanu, P. Singla, T. Singh, and P. D. Scott, "Uncertainty propagation for nonlinear dynamics systems using Gaussian sum mixture models," *Journal of Guidance, Control, and Dynamics*, vol. 31, no. 6, pp. 1623–1633, 2008.
18 ——, "Adaptive Gaussian sum filter for nonlinear Bayesian estimation," *IEEE Transactions on Automatic Control*, vol. 56, no. 9, pp. 2151–2156, 2011.
19 K. J. DeMars, Y. Cheng, and M. K. Jah, "Collision probability with Gaussian mixture orbit uncertainty," *Journal of Guidance, Control, and Dynamics*, vol. 37, no. 3, pp. 979–958, 2014.
20 O. Straka, J. Duník, and M. Šimandl, "Structure adaptation of nonlinear filters based on non-Gaussianity measures," in *Proceedings of the 2015 American Control Conference*, Chicago, IL, USA, 2015.
21 M. Raitoharju, S. Ali-Löytty, and R. Piché, "Binomial Gaussian mixture filter," *EURASIP Journal on Advances in Signal Processing*, vol. 36, pp. 1–18, 2015.
22 O. Straka, J. Duník, and I. Punčochář, "Direcional splitting for structure adaptation of Bayesian filters," in *Proceedings of the American Control Conference*, Boston, MA, USA, 7 2016.
23 H. W. Sorenson and D. L. Alspach, "Approximation of density function by a sum of gaussians for nonlinear bayesian estimation," in *Proceedings of the 1st Symposium for Nonlinear Estimation Theory and Its Applications*, San Diego, 1970, pp. 88–90.
24 T. Schön, F. Gustafsson, and P. Nordlund, "Marginalized particle filters for mixed linear/nonlinear state-space models," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2279–2289, 2005.
25 V. Šmídl and M. Gašperin, "Rao-Blackwellized point mass filter for reliable state estimation," in *16th International Conference on Information Fusion*, Istanbul, Turkey, 2013.
26 J. T. Horwood, N. D. Aragon, and A. B. Poore, "Gaussian sum filters for space surveillance: Theory and simulations," *Journal of Guidance, Control, and Dynamics*, vol. 34, no. 6, pp. 1839–1851, 2011.
27 U. D. Hanebeck, K. Briechle, and A. Rauh, "Progressive Bayes: A new framework for nonlinear state estimation," in *Proceedings of SPIE, Vol. 5099, AeroSense Symposium*, Orlando, Florida, 2003, pp. 256–267.
28 C. Couvreur, "The EM algorithm: A guided tour," in *Proceedings of the Second IEEE European Workshop on Computer-Intensive Methods in Control and Signal Processing*, 1997, pp. 115–120.

29 F. Faubel and D. Klakow, “Further improvement of the adaptive level of detail transform: Splitting in direction on the nonlinearity,” in *Proceedings of the 18th European Signal Processing Conference*, Aalborg, Denmark, 8 2010.

30 M. F. Huber, “Adaptive Gaussian mixture filter based on statistical linearisation,” in *Proceedings of the 14th International Conference on Information Fusion*, Chicago, IL, USA, Jul. 2011.

31 M. Raitoharju and S. Ali-Löytty, “An adaptive derivative free method for Bayesian posterior approximation,” *IEEE Signal Processing Letters*, vol. 19, no. 2, pp. 87–90, 2012.

32 F. Havlak and M. Campbell, “Discrete and continuous probabilistic anticipation for autonomous robots in urban environments,” *IEEE Transactions on Robotics*, vol. 30, no. 2, pp. 461–474, 2013.

33 J.-F. Cardoso, “Dependence, correlation and Gaussianity in independent component analysis,” *Journal of Machine Learning Research*, vol. 4, pp. 1177–1203, 2003.

34 A. H. Stroud, *Approximate calculation of multiple integrals*. Prentice-Hall Englewood Cliffs, NJ, 1971.

35 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.

36 S. J. Julier and J. K. Uhlmann, “Unscented filtering and nonlinear estimation,” *IEEE Review*, vol. 92, no. 3, pp. 401–421, 2004.

37 Y. Wu, D. Hu, M. Wu, and X. Hu, “A numerical-integration perspective on Gaussian filters,” *IEEE Transactions on Signal Processing*, vol. 54, no. 8, pp. 2910–2921, 2006.

38 J. Steinbring and U. D. Hanebeck, “LRKF revisited: The smart sampling Kalman filter (S^2KF),” *Journal of Advances in Information Fusion*, vol. 9, no. 2, pp. 106–123, 2014.

39 J. Duník, O. Straka, M. Šimandl, and E. Blasch, “Random-point-based filters: Analysis and comparison in target tracking,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 2, pp. 303–308, 2015.

40 M. A. Carreire-Perpinan, “Mode-finding for mixtures of Gaussian distributions,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1318–1323, 2000.

41 G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. The Johns Hopkins University Press, 11 1996.

42 M. Mallick, “Differential geometry measures of nonlinearity with applications to ground target tracking,” in *Proceedings of the 7th International Conference on Information Fusion*, Stockholm, Sweden, Jun. 2004.

43 R. A. Groeneveld and G. Meeden, “Measuring skewness and kurtosis,” *Journal of the Royal Statistical Society. Series D (The Statistician)*, vol. 33, no. 4, pp. 391–399, 1984.

44 P. von Hippel, *Skewness*. Springer, 2010, International Encyclopedia of Statistical Science, Ed. M. Lovric.

45 A. Hyvärinen and E. Oja, “Independent component analysis: Algorithms and applications,” *Neural Networks*, vol. 13, pp. 411–430, 2000.

46 M. B. Miller, *Mathematics and Statistics for Financial Risk Management*. Wiley, 2014.

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

48 G. Strang, *Introduction to Linear Algebra: 4th Edition*. Wellesley - Cambridge Press, 2009.

Appendix: Illustration of Splitting Algorithm

Illustration of Algorithm 1 for a two-dimensional variable $\mathbf{x}$ is shown in Figure 6 with $\mathbf{Q}_\chi = \mathbf{I}_2$ (upper figure) and $\mathbf{Q}_\chi = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ sin(\theta) & \cos(\theta) \end{bmatrix}, \theta = 60[deg]$ (lower figure). The red symbols '+' denote the means of the GM terms whereas the ellipses illustrates the covariance matrices of the GM terms. We can see that the rotation matrix affects both the means and the covariance matrices of the terms of not only $\hat{p}_{\mathrm{GM},N}(\boldsymbol{\chi}; \mathbf{Q}_\chi)$ (14) but also of $\hat{p}_{\mathrm{GM},N}(\mathbf{x}; \mathbf{Q}_\chi)$ (16). In turn, the matrix $\mathbf{Q}_\chi$ influences the properties of the mixture PDF $\hat{p}_{\mathrm{M},N}(\mathbf{y})$ and its GM PDF approximation $\hat{p}_{\mathrm{GM},N}(\mathbf{y})$. Note that, the intermediate splitting direction $\boldsymbol{\nu}_\chi = \mathbf{Q}_\chi \mathbf{r}$ (17) and the splitting direction $\boldsymbol{\nu}_x$ (18) are illustrated in Figure 6 as well.

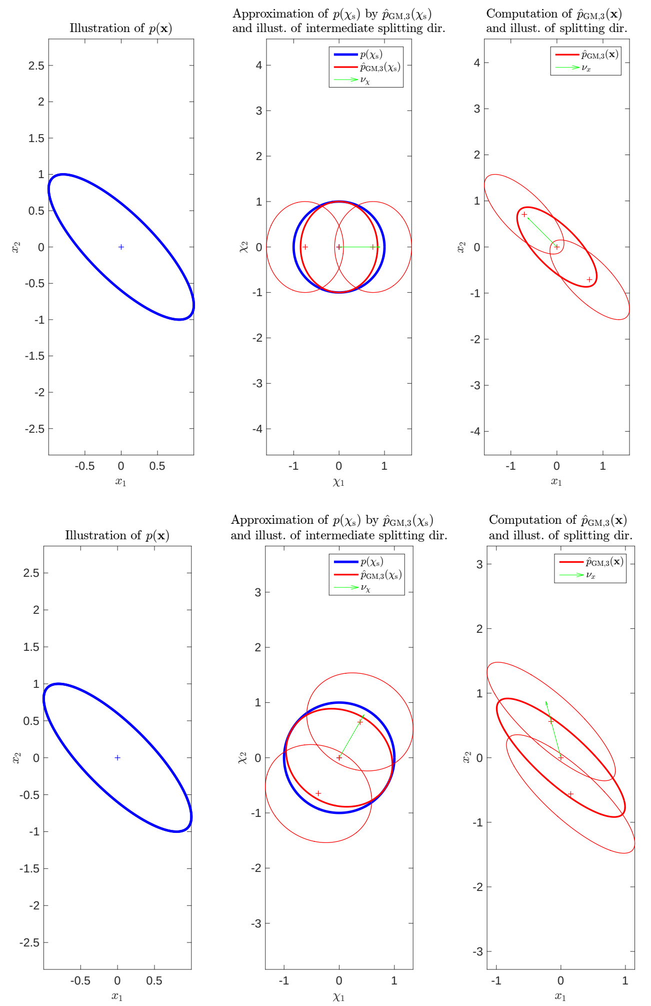

Fig. 6: Illustration of Algorithm 1 for Gaussian PDF splitting (*left fig.*: illustration of original to-be-split PDF $p(\mathbf{x})$; *middle fig.*: illustration of standard normal PDF $p(\boldsymbol{\chi})$ and its (rotated) GM PDF approximation; *right fig.*: illustration of GM PDF $\hat{p}_{\mathrm{GM},N}(\mathbf{x})$ approximating $p(\mathbf{x})$).