Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Estimation of Parameters of Gaussian Sum Distributed Noises in State-Space Models

Jindřich Duník, Oliver Kost, and Ondřej Straka

Department of Cybernetics and Research Centre NTIS,
Faculty of Applied Sciences, University of West Bohemia,
Univerzitní 8, 306 14 Pilsen, Czech Republic (e-mail: {dunikj,kost,straka30}@kky.zcu.cz)

Abstract: The paper deals with the estimation of noise parameters of a linear time-varying system. In particular, the stress is laid on the state-space models, where the state and measurement noises are described by the Gaussian sum probability density functions. The recently introduced measurement difference method for the estimation of higher-order moments of the state and measurement noises is revised and, subsequently, extended for estimation of the parameters of the noise Gaussian sum densities with a special focus on the densities with two-components. The theoretical results are discussed and illustrated in a numerical example.

Keywords: Noise parameter estimation, State-space models, Gaussian sum density.

1. INTRODUCTION

Knowledge of a system state-space model is a key prerequisite for modern optimal algorithms in the areas of the state estimation, system identification, fault detection, and automatic control. An incorrect model may result in significant deterioration of underlying algorithms output quality or even in their failure.

The state-space model is designed to consistently describe a combination of deterministic and stochastic influences affecting behaviour of the system quantities. While the deterministic part of the model often arises from the first principles based on physical, chemical, or biological laws governing the system behaviour, the description of the stochastic part is often difficult to find by the modelling and has to be identified using the measured data.

From the seventies, a significant research interest has been focused on the design of the methods for estimation of the properties of the stochastic part of the state-space model, i.e., state and measurement noise properties (Mehra, 1970; Bélanger, 1974; Odelson et al., 2006; Särkkä and Nummenmaa, 2009; Duník et al., 2017). Particularly, the stress was laid on the estimation of the *covariance matrices* [1] of the noises, assuming their zero mean. Unfortunately, knowledge of the first two moments is *not* sufficient for noises complete description if they are *not* Gaussian (e.g., heavy-tailed or asymmetric densities) (Kost et al., 2018).

In (Kost et al., 2018), therefore, a method estimating noise higher-order moments and potentially parameters of the state and measurement noise probability density functions (PDFs) has been proposed. The proposed method, further denoted as the *measurement difference method* (MDM), has been designed for the linear time-varying (LTV) state-space model, where the noises are described by Student's t-distribution or skew Student's t-distribution. The Student's t-distribution type of the PDF is extensively studied and used in the area of the object tracking using a time-of-flight based ultra-wideband distance measurement (Nurminen et al., 2018; Roth et al., 2013). Unfortunately, no other *non-Gaussian* noise PDF has been considered.

The goal of this paper is to extend applicability of the MDM for estimation of moments and consequently parameters of the state and measurement noises described by the *Gaussian sum* (GS) PDFs. The GS PDF is a *universal* density used in wide range of signal processing, control, and detection algorithms from the following reasons:

- *Analytical solution*: Majority of optimal signal processing, control, and detection algorithms designed for linear Gaussian models (LGM) are straightforwardly extensible for linear Gaussian sum models (LGSM); for example analytical solution to the Bayesian recursive relations for the LGM results in the Kalman filter (KF), whereas the solution for the LGSM results in the Gaussian sum filter, which can be understood as a bank of parallel KFs (Anderson and Moore, 1979),
- *Accurate model*: A GS PDF can naturally be used for the description of some physical phenomena (Cerón, 2017); e.g., measurement noise affecting radar altimeter measurement (Gustafsson et al., 2002),
- *Approximate model*: Any PDF can be arbitrarily well approximated by a GS PDF (Williams and Maybeck, 2003; Hanebeck et al., 2003; Duník et al., 2018b).

The rest of the paper is organised as follows. In Section 2, the state-space model is defined and the goal of the paper is particularised. Section 3 deals with an overview of the MDM method. In Sections 4 and 5, a transformation of moments and parameters of a random variable with a

[1] The popularity of the methods estimating the noises' covariance matrices, assuming known (typically zero) mean, stems from the fact that a *vast majority* of the modern signal processing and decision making algorithms either *(i)* require the first two moments only (independently of their distribution) or *(ii)* assume the Gaussian distribution of the noises, for which the means and covariance matrices are sufficient statistics.

GS PDF is considered. A numerical evaluation is given in Section 6 and concluding remarks are drawn in Section 7.

2. SYSTEM DEFINITION AND PROBLEM FORMULATION

Let the following discrete-time state-space model of an LTV stochastic dynamic system with additive noises

$$\mathbf{x}_{k+1} = \mathbf{F}_k\mathbf{x}_k + \mathbf{w}_k, \quad k = 0, 1, 2, \ldots, \tau, \tag{1}$$

$$\mathbf{z}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k, \quad k = 0, 1, 2, \ldots, \tau, \tag{2}$$

be considered, where the vectors $\mathbf{x}_k \in \mathbb{R}^{n(x)}$ and $\mathbf{z}_k \in \mathbb{R}^{n(z)}$ represent the immeasurable state of the system and the known measurement at time instant k, respectively. The state and measurement matrices $\mathbf{F}_k \in \mathbb{R}^{n(x)\times n(x)}$ and $\mathbf{H}_k \in \mathbb{R}^{n(z)\times n(x)}$ are *known* and bounded $\forall k$. The system state is assumed to be *observable* $\forall k$. Neither the initial condition $\mathbf{x}_0$ nor the initial condition PDF are known.

The variables $\mathbf{w}_k \in \mathbb{R}^{n(x)}$ and $\mathbf{v}_k \in \mathbb{R}^{n(z)}$ are state and measurement *white* noises with the GS PDFs

$$p_{\mathrm{w}}(\mathbf{w}_k;\boldsymbol{\alpha}_{\mathrm{w}}) = \sum_{i=1}^{N_{\mathrm{w}}} \beta_{\mathrm{w}}^{(i)}\mathcal{N}\{\mathbf{w}_k;\bar{\mathbf{w}}^{(i)},\mathbf{Q}^{(i)}\}, \tag{3}$$

$$p_{\mathrm{v}}(\mathbf{v}_k;\boldsymbol{\alpha}_{\mathrm{v}}) = \sum_{i=1}^{N_{\mathrm{v}}} \beta_{\mathrm{v}}^{(i)}\mathcal{N}\{\mathbf{v}_k;\bar{\mathbf{v}}^{(i)},\mathbf{R}^{(i)}\}, \tag{4}$$

where the notation $\mathcal{N}\{\mathbf{w}_k;\bar{\mathbf{w}}^{(i)},\mathbf{Q}^{(i)}\}$ means the Gaussian PDF of a random variable $\mathbf{w}_k$ with the mean $\bar{\mathbf{w}}^{(i)}$ and the covariance matrix $\mathbf{Q}^{(i)}$, $\beta_{\mathrm{w}}^{(i)}$ is a weight of the i-th GS component satisfying $\sum_{i=1}^{N_{\mathrm{w}}}\beta_{\mathrm{w}}^{(i)} = 1$, and $\boldsymbol{\alpha}_{\mathrm{w}}$ is a set of all GS PDF (3) parameters defined as

$$\boldsymbol{\alpha}_{\mathrm{w}} \triangleq \{N_{\mathrm{w}}, \beta_{\mathrm{w}}^{(1)}, \bar{\mathbf{w}}^{(1)}, \mathbf{Q}^{(1)}, \ldots, \bar{\mathbf{w}}^{(N_{\mathrm{w}})}, \mathbf{Q}^{(N_{\mathrm{w}})}\}. \tag{5}$$

The set of parameters in (4) is defined analogously as

$$\boldsymbol{\alpha}_{\mathrm{v}} \triangleq \{N_{\mathrm{v}}, \beta_{\mathrm{v}}^{(1)}, \bar{\mathbf{v}}^{(1)}, \mathbf{R}^{(1)}, \ldots, \bar{\mathbf{v}}^{(N_{\mathrm{v}})}, \mathbf{R}^{(N_{\mathrm{v}})}\}. \tag{6}$$

The *sets of parameters* $\boldsymbol{\alpha}_{\mathrm{w}}$ and $\boldsymbol{\alpha}_{\mathrm{v}}$ are *unknown* as well as the *moments* of the state and measurement noises.

2.1 Noise Statistics and Parameters Estimation

For more than five decades, an extensive research interest has been devoted to the estimation of the properties of the state and measurement noises. Mainly, the stress was laid on the estimation of the covariance matrices (Duník et al., 2017). These methods are traditionally divided into four groups (Mehra, 1970): correlation methods, maximum likelihood methods, covariance matching methods, and Bayesian methods.

Rather marginal attention has been, however, devoted to the estimation of higher-order moments and parameters of non-Gaussian noise PDFs. In (Kost et al., 2018), the MDM [2] has been proposed for moments and parameters estimation assuming Student's t-distribution of the noises.

2.2 Goal of the Paper

The goal of the paper is to extend the applicability of the MDM to the estimation of the sets of parameters $\boldsymbol{\alpha}_{\mathrm{w}}$ and $\boldsymbol{\alpha}_{\mathrm{v}}$ of the state and measurement noise of the *LGSM* (1)–(4) on the basis of the measured data $\mathbf{z}^{\tau} \triangleq [\mathbf{z}_0, \mathbf{z}_1, \ldots, \mathbf{z}_{\tau}]$ and known matrices $\mathbf{F}_k$ and $\mathbf{H}_k$, $\forall k$.

The MDM consists of the following three successive steps

(i) Design of a linear measurement predictor and statistical analysis of the measurement prediction error (MPE),
(ii) Sample-based estimation of the MPE moments and estimation of the noise moments,
(iii) Computation of the noise parameters $\boldsymbol{\alpha}_{\mathrm{w}}$ and $\boldsymbol{\alpha}_{\mathrm{v}}$ in (3) and (4), respectively.

In the following sections, the first two steps are briefly reviewed, as they basically remain the same as in the Student's t-distribution case thoroughly discussed in (Kost et al., 2018). The third step is described in detail.

3. MEASUREMENT PREDICTION, MPE ANALYSIS, AND NOISE MOMENT ESTIMATION

The MDM (Kost et al., 2018) is based on the statistical analysis of the estimate error of a linear measurement predictor. Design of the MDM, therefore, starts with the definition of an augmented measurement vector and its one-step prediction. Then, the statistical properties of the MPE are determined and noise moments are estimated.

3.1 Augmented Measurement Vector and its Prediction

Consider the LGSM model (1), (2), available measurements $\mathbf{z}^{\tau}$, and the parameter L selected such that the observability matrix

$$\mathcal{O}_k^L \triangleq \begin{bmatrix} \mathbf{H}_k \\ \mathbf{H}_{k+1}\mathbf{F}_k \\ \vdots \\ \mathbf{H}_{k+L-1}\mathcal{F}_k^{L-1} \end{bmatrix} \in \mathbb{R}^{Ln(z)\times n(x)}, \tag{7}$$

is full rank [3] $\forall k$ and $\mathcal{F}_k^M \in \mathbb{R}^{n(x)\times n(x)}$ is defined as $\mathcal{F}_k^M \triangleq \prod_{i=1}^M \mathbf{F}_{k+M-i} = \mathbf{F}_{k+M-1}\ldots\mathbf{F}_{k+1}\mathbf{F}_k$.

Then, the *augmented measurement vector* $\mathbf{Z}_k^L$ can be expressed as

$$\mathbf{Z}_k^L = \mathcal{O}_k^L\mathbf{x}_k + \boldsymbol{\Gamma}_k^L\mathbf{W}_k^L + \mathbf{V}_k^L, \tag{8}$$

where $k = 0, \ldots, \tau - L + 1$, and the vectors and matrices $\mathbf{Z}_k^L \in \mathbb{R}^{Ln(z)}$, $\mathbf{W}_k^L \in \mathbb{R}^{Ln(x)}$, $\mathbf{V}_k^L \in \mathbb{R}^{Ln(z)}$, $\boldsymbol{\Gamma}_k^L \in \mathbb{R}^{Ln(z)\times Ln(x)}$ are defined by

$$\mathbf{Z}_k^L \triangleq \begin{bmatrix} \mathbf{z}_k \\ \mathbf{z}_{k+1} \\ \vdots \\ \mathbf{z}_{k+L-1} \end{bmatrix}, \mathbf{W}_k^L \triangleq \begin{bmatrix} \mathbf{w}_k \\ \mathbf{w}_{k-1} \\ \vdots \\ \mathbf{w}_{k+L-1} \end{bmatrix}, \mathbf{V}_k^L \triangleq \begin{bmatrix} \mathbf{v}_k \\ \mathbf{v}_{k+1} \\ \vdots \\ \mathbf{v}_{k+L-1} \end{bmatrix},$$

$$\boldsymbol{\Gamma}_k^L \triangleq \begin{bmatrix} \mathbf{0}_{n(z)\times n(x)} & \mathbf{0}_{n(z)\times n(x)} & \cdots & \mathbf{0}_{n(z)\times n(x)} & \mathbf{0}_{n(z)\times n(x)} \\ \mathbf{H}_{k+1} & \mathbf{0}_{n(z)\times n(x)} & \cdots & \mathbf{0}_{n(z)\times n(x)} & \mathbf{0}_{n(z)\times n(x)} \\ \mathbf{H}_{k+2}\mathbf{F}_{k+1} & \mathbf{H}_{k+2} & \cdots & \mathbf{0}_{n(z)\times n(x)} & \mathbf{0}_{n(z)\times n(x)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{H}_{k+L-1}\mathcal{F}_{k+1}^{L-2} & \mathbf{H}_{k+L-1}\mathcal{F}_{k+2}^{L-3} & \cdots & \mathbf{H}_{k+L-1} & \mathbf{0}_{n(z)\times n(x)} \end{bmatrix},$$

and the notation $\mathbf{0}_n$ stands for the zero matrix of indicated dimension.

[2] The MDM belongs into the correlation methods.

[3] The full rank matrix $\mathcal{O}_k^L$ always exists as the system state is supposed to be observable, i.e., the state $\mathbf{x}_k$ is observable from the augmented measurement vector $\mathbf{Z}_k^L$.

A *one-step* prediction of $\mathbf{Z}_k^L$ (8) can be written as

$$\widehat{\mathbf{Z}}_k^L = \mathcal{O}_k^L \mathbf{F}_{k-1} \left(\mathcal{O}_{k-1}^L\right)^\dagger \mathbf{Z}_{k-1}^L, \tag{9}$$

where $(\mathcal{O}_{k-1}^L)^\dagger = \left((\mathcal{O}_{k-1}^L)^T \mathcal{O}_{k-1}^L\right)^{-1} (\mathcal{O}_{k-1}^L)^T \in \mathbb{R}^{n(x)\times Ln(z)}$ is the pseudoinverse of the matrix $\mathcal{O}_{k-1}^L$.

3.2 Augmented measurement vector prediction error

The *augmented measurement vector prediction error* (AMPE) is defined as

$$\widetilde{\mathbf{Z}}_k^L = \mathbf{Z}_k^L - \widehat{\mathbf{Z}}_k^L, k = 1, \ldots, \tau - L + 1. \tag{10}$$

By substitution of (8) and (1) into (9), the AMPE (10) reads

$$\widetilde{\mathbf{Z}}_k^L = \mathcal{A}_k \mathcal{E}_k, \tag{11}$$

where

$$\mathcal{E}_k = \left[(\mathbf{W}_{k-1}^{L+})^T \ (\mathbf{V}_{k-1}^{L+})^T\right]^T \in \mathbb{R}^{L^+(n(x)+n(z))}, \tag{12}$$

$$\mathcal{A}_k = \left[\mathcal{A}_k^{(\mathrm{w})} \ \mathcal{A}_k^{(\mathrm{v})}\right] \in \mathbb{R}^{Ln(z)\times L^+(n(x)+n(z))}, \tag{13}$$

with $L^+ = L + 1$,

$$\mathcal{A}_k^{(\mathrm{w})} = \begin{bmatrix}\mathbf{I}_{Ln(z)} \\ \mathbf{I}_{Ln(z)}\end{bmatrix}^T \begin{bmatrix} \left[\mathcal{O}_k^L, \ \boldsymbol{\Gamma}_k^L\right] \\ \left[-\mathcal{O}_k^L \mathbf{F}_{k-1} \left(\mathcal{O}_{k-1}^L\right)^\dagger \boldsymbol{\Gamma}_{k-1}^L, \mathbf{0}_{Ln(z)\times n(x)}\right] \end{bmatrix},$$

$$\mathcal{A}_k^{(\mathrm{v})} = \begin{bmatrix}\mathbf{I}_{Ln(z)} \\ \mathbf{I}_{Ln(z)}\end{bmatrix}^T \begin{bmatrix} \left[\mathbf{0}_{Ln(z)\times n(z)}, \mathbf{I}_{Ln(z)}\right] \\ \left[-\mathcal{O}_k^L \mathbf{F}_{k-1} \left(\mathcal{O}_{k-1}^L\right)^\dagger, \mathbf{0}_{Ln(z)\times n(z)}\right] \end{bmatrix}.$$

The symbol $\mathbf{I}_n \in \mathbb{R}^{n\times n}$ denotes the identity matrix.

It can be seen, that the matrix $\mathcal{A}_k$ in (11) is a function of the known model matrices $\mathbf{F}_k$ and $\mathbf{H}_k$, thus, $\mathcal{A}_k$ is *known* as well. The AMPE (11) is, therefore, a *linear* function of the state and measurement noises stacked in the vector $\mathcal{E}_k$, of which statistical properties are sought.

Note 1: Whereas, the form of the AMPE (10) is suitable for the prediction error computation on the basis of the measured data, the form (11) can be used for the following AMPE statistical analysis.

3.3 Raw Moments of AMPE

The AMPE $\widetilde{\mathbf{Z}}_k^L$ (11) is a stochastic process depending linearly on the state and measurement noises. As such, all AMPE moments (either raw or central) are linear functions of state and measurement noise moments (Kost et al., 2018). In particular, the m-th raw moment of the AMPE (11) at time k is given by

$$\mathbf{b}_k^m = \mathsf{E}\left[\left(\widetilde{\mathbf{Z}}_k^L\right)^{\otimes^m}\right] = \mathcal{A}_k^{\otimes^m} \mathsf{E}\left[\mathcal{E}_k^{\otimes^m}\right], \tag{14}$$

where $\mathbf{b}_k^m \in \mathbb{R}^{(Ln(z))^m}$ and the notation $\mathcal{A}^{\otimes^m}$ stands for m-th Kronecker power. As the matrix $\mathcal{A}_k$ (and thus $\mathcal{A}_k^{\otimes^m}$) is known, the m-th raw moment of the AMPE (14), in the vector form, is a *linear* function of the m-th "pure" raw moments of the state and measurement noises

$$\mathbf{M}_{\mathrm{w}}^m \triangleq \mathsf{E}[\mathbf{w}_k^{\otimes^m}], \forall k, \tag{15}$$

$$\mathbf{M}_{\mathrm{v}}^m \triangleq \mathsf{E}[\mathbf{v}_k^{\otimes^m}], \forall k, \tag{16}$$

and of the "cross" raw moments of the state and measurement noises, which can be described by a *known* polynomial function of the lower-order moments, i.e.,

$$\mathbf{M}_{\mathrm{wv}}^m \triangleq f(\mathbf{M}_{\mathrm{w}}^1, \mathbf{M}_{\mathrm{v}}^1, \ldots, \mathbf{M}_{\mathrm{w}}^{m-1}, \mathbf{M}_{\mathrm{v}}^{m-1}). \tag{17}$$

The elements of the vectors $\mathbf{M}_{\mathrm{w}}^m$ (15), $\mathbf{M}_{\mathrm{v}}^m$ (16), and $\mathbf{M}_{\mathrm{wv}}^m$ (17) form the vector $\mathsf{E}\left[\mathcal{E}_k^{\otimes^m}\right]$, however, because of the structure of $\mathcal{E}_k$ (12), the vector $\mathsf{E}\left[\mathcal{E}_k^{\otimes^m}\right]$ in (14) contains *multiple* copies of the vectors (15)–(17).

Defining the vector $\breve{\boldsymbol{\theta}}^m \triangleq [(\mathbf{M}_{\mathrm{w}}^m)^T, (\mathbf{M}_{\mathrm{v}}^m)^T, (\mathbf{M}_{\mathrm{wv}}^m)^T]^T \in \mathbb{R}^{n(\breve{\boldsymbol{\theta}}^m)}$ so that it contains a *single* copy of the unknown m-th moments, the AMPE m-th moment (14) reads

$$\breve{\mathcal{A}}_k^m \breve{\boldsymbol{\theta}}^m = \mathbf{b}_k^m, \tag{18}$$

where the matrix $\breve{\mathcal{A}}_k^m \in \mathbb{R}^{(Ln(z))^m \times n(\boldsymbol{\theta}^m)}$ is

$$\begin{aligned}\breve{\mathcal{A}}_k^m &= \left(\mathcal{A}_k^{\otimes^m}\right) \boldsymbol{\Psi}^m \\ &= [\mathcal{A}_{\mathrm{w},k}^m, \mathcal{A}_{\mathrm{v},k}^m, \mathcal{A}_{\mathrm{wv},k}^m]\end{aligned} \tag{19}$$

and the duplication matrix $\boldsymbol{\Psi}^m \in \mathbb{N}^{(L^+(n(x)+n(z)))^m \times n(\breve{\boldsymbol{\theta}}^m)}$ is designed to fulfil

$$\mathsf{E}\left[\tilde{\mathcal{E}}_k^{\otimes^m}\right] = \boldsymbol{\Psi}^m \breve{\boldsymbol{\theta}}^m. \tag{20}$$

Thus, the matrices $\mathcal{A}_{\mathrm{w},k}^m, \mathcal{A}_{\mathrm{v},k}^m, \mathcal{A}_{\mathrm{wv},k}^m$ are *known.* Design of the duplication matrix is discussed and illustrated in (Duník et al., 2018a; Kost et al., 2018).

The linear equation (18) can be written for all possible time instants k in the following convenient form

$$\mathcal{A}^m \boldsymbol{\theta}^m = \mathbf{b}^m - \mathcal{A}_{\mathrm{wv}}^m \mathbf{M}_{\mathrm{wv}}^m, \tag{21}$$

with

$$\boldsymbol{\theta}^m = [(\mathbf{M}_{\mathrm{w}}^m)^T, (\mathbf{M}_{\mathrm{v}}^m)^T]^T, \tag{22}$$

$$\mathcal{A}^m = \left[\left(\mathcal{A}_1^m\right)^T, \left(\mathcal{A}_2^m\right)^T, \cdots, \left(\mathcal{A}_{\tau-L+1}^m\right)^T\right]^T, \tag{23}$$

$$\mathcal{A}_{\mathrm{wv}}^m = \left[\left(\mathcal{A}_{\mathrm{wv},1}^m\right)^T, \left(\mathcal{A}_{\mathrm{wv},2}^m\right)^T, \cdots, \left(\mathcal{A}_{\mathrm{wv},\tau-L+1}^m\right)^T\right]^T, \tag{24}$$

$$\mathbf{b}^m = \left[\left(\mathbf{b}_1^m\right)^T, \left(\mathbf{b}_2^m\right)^T, \cdots, \left(\mathbf{b}_{\tau-L+1}^m\right)^T\right]^T, \tag{25}$$

where $\mathcal{A}_k^m = [\mathcal{A}_{\mathrm{w},k}^m, \mathcal{A}_{\mathrm{v},k}^m]$.

Note 2: An m-th cross-moment $\mathbf{M}_{\mathrm{wv}}^m$ (17), for $m > 1$, can be written as a function of lower-order noise moments i.e., of $\{\mathbf{M}_{\mathrm{w}}^i\}_{i=1}^{m-1}$ and $\{\mathbf{M}_{\mathrm{v}}^i\}_{i=1}^{m-1}$. For example, $\mathbf{M}_{\mathrm{wv}}^2$ is a function of the noises' first-order moments $\mathbf{M}_{\mathrm{w}}^1$ and $\mathbf{M}_{\mathrm{v}}^1$. Then, the m-th AMPE moment is a *linear* function of the unknown elements of the "pure" m-th noise moments.

3.4 AMPE and Noise Moments Estimation

The matrices $\mathcal{A}^m$ and $\mathcal{A}_{\mathrm{wv}}^m$ in (21) are functions of the *known* system matrices $\mathbf{F}_k$ and $\mathbf{H}_k$, $\forall k$, and of the *known* duplication matrix $\boldsymbol{\Psi}^m$, only. If the vector $\mathbf{b}^m$ (25) were available, then the unknown moments gathered in $\boldsymbol{\theta}^m$ (21) could be estimated by the least-squares (LS) method. The vector $\mathbf{b}^m$, summarising the AMPE moments, is, however, *unknown* (as it depends on the sought noise moments (14)). Nevertheless, it can be *estimated* from the measured data $\mathbf{z}^\tau$. Based on the AMPE moments estimates, the sought noise moments $\boldsymbol{\theta}^m$ are found.

The AMPE and noise moments estimation follows the procedure for $m = 1, 2, \ldots, m_{\max}$:

(i) The AMPE moments for one time instant k, summarised in the vector $\mathbf{b}_k^m$ (14), are estimated on the basis of the AMPE sequence $\{\widetilde{\mathbf{Z}}_k^L\}_{k=1}^{\tau-L+1}$ (10) as

$$\hat{\mathbf{b}}_k^m = \left(\widetilde{\mathbf{Z}}_k^L\right)^{\otimes^m}. \quad (26)$$

(ii) The AMPE moments $\forall k$, summarised in the vector $\mathbf{b}^m$ (25), are estimated as

$$\hat{\mathbf{b}}^m = \left[\left(\hat{\mathbf{b}}_1^m\right)^T, \left(\hat{\mathbf{b}}_2^m\right)^T, \cdots, \left(\hat{\mathbf{b}}_{\tau-L+1}^m\right)^T\right]^T. \quad (27)$$

(iii) Based on the estimate $\hat{\mathbf{b}}^m$ (27) and (21), the LS optimum estimate of the m-th moment of the noises is given by

$$\hat{\boldsymbol{\theta}}^m = [(\hat{\mathbf{M}}_{\mathrm{w}}^m)^T, (\hat{\mathbf{M}}_{\mathrm{v}}^m)^T]^T = (\mathcal{A}^m)^\dagger\left(\hat{\mathbf{b}}^m - \mathbf{A}_{\mathrm{wv}}^m \hat{\mathbf{M}}_{\mathrm{wv}}^m\right), \quad (28)$$

where $\hat{\mathbf{M}}_{\mathrm{wv}}^m$ is a function of the previously computed lower-order moment estimates $\{\hat{\boldsymbol{\theta}}^i\}_{i=1}^{m-1}$ (for $m > 1$ as discussed in Note 2).

Note 3: By an extension of the proof in (Kost et al., 2018), it is possible to show that the noise moment estimates $\hat{\boldsymbol{\theta}}^m$ converge to the true value and the variance of the estimates goes to the zero with increasing number of data τ.

4. COMPUTATION OF GAUSSIAN SUM PARAMETERS FROM MOMENTS

In this section, the computation of the GS PDF set of parameters from the respective moments is discussed and illustrated. The computation is detailed for the measurement noise $\mathbf{v}_k$ only; for the state noise $\mathbf{w}_k$ the calculations are *analogous*.

4.1 GS PDF Parameter Computation

The GS PDF $p_{\mathrm{v}}(\mathbf{v}_k; \boldsymbol{\alpha}_{\mathrm{v}})$ (4) depends on the set of parameters $\boldsymbol{\alpha}_{\mathrm{v}}$ (6), which consists of the number of GS terms N_{v}, the weights $\{\beta_{\mathrm{v}}^{(i)}\}_{i=1}^{N_{\mathrm{v}}}$, the means $\{\bar{\mathbf{v}}^{(i)}\}_{i=1}^{N_{\mathrm{w}}}$, and the covariance matrices $\{\mathbf{R}^{(i)}\}_{i=1}^{N_{\mathrm{v}}}$. Consequently, all the moments $\mathbf{M}_{\mathrm{v}}^m$ (16) nonlinearly depend on the parameters, i.e.,

$$\mathbf{M}_{\mathrm{v}}^m = \mathbf{g}_{\mathrm{v}}^m(\boldsymbol{\alpha}_{\mathrm{v}}), \quad (29)$$

where $\mathbf{g}_{\mathrm{v}}^m(\cdot)$ is a known function.

Theoretically, given the *true* moments $\mathbf{M}_{\mathrm{v}}^m, m = 1, 2, \ldots, m_{\max}$, the set of parameters $\boldsymbol{\alpha}_{\mathrm{v}}$ can be computed from the system of $m_{\max}$ nonlinear equations (29). However, such system of equations may *not* have a unique solution even for $m_{\max} \to \infty$.

4.2 Two-Component GS PDF Parameter Computation

The attention has been, thus, focused on a solution to (29), $\forall m$, for $\boldsymbol{\alpha}_{\mathrm{v}}$ subject to certain *assumptions* on the computed set of parameters. Typically, the proposed solutions are developed for a *two-component* GS PDF, i.e., for $N_{\mathrm{v}} = 2$, where the sought set

$$\boldsymbol{\alpha}_{\mathrm{v}} = \{\beta_{\mathrm{v}}^{(1)}, \bar{\mathbf{v}}^{(1)}, \mathbf{R}^{(1)}, \bar{\mathbf{v}}^{(2)}, \mathbf{R}^{(2)}\} \quad (30)$$

consists of 5 *unknown* parameters as $\beta_{\mathrm{v}}^{(2)} = 1 - \beta_{\mathrm{v}}^{(1)}$. The assumption of two-component GS PDF $p_{\mathrm{v}}(\mathbf{v}_k; \boldsymbol{\alpha}_{\mathrm{v}})$ is adopted in this paper as well.

4.3 State-of-the-Art Solutions

The fundamental solution to the two-component GS PDF set of parameters (30) computation, denoted as the *method of moments* (MoM), was proposed by Pearson in 1894. The MoM was, however, proposed for a *scalar* [4] random variable only and is based on the analytical solution to a suitably reformulated system of nonlinear equations for $m_{\max} \geq 5$ (Cerón, 2017).

Various extensions of the MoM have been proposed since focusing on the parameter estimation for a vector variable. The proposed solutions are, however, tied with rather *stringent* assumptions (Cerón, 2017), typically related to the known means $\bar{\mathbf{v}}^{(1)}$ and $\bar{\mathbf{v}}^{(2)}$ (or at least certain elements of the *both* means). The remaining parameters, including $\beta_{\mathrm{v}}^{(1)}, \mathbf{R}^{(1)}, \mathbf{R}^{(2)}$, can be found from (29) analytically for $m_{\max} = 3$ (Cerón, 2017).

However, the assumption of known means is too restrictive for many practical applications, and, thus, in the following part, two solutions *relaxing* this assumption are developed. The solutions are detailed for two-dimensional variable $\mathbf{v}_k$, i.e., $n(v) = 2$, but it can be straightforwardly extended for a dimension $n(v) \geq 3$.

4.4 Proposed Full Solution

The proposed full solution computes the *full* set of 5 unknown parameters of $\boldsymbol{\alpha}_{\mathrm{v}}$ (30). It means, for two dimensional random variable, the solution computes 11 unique elements of the parameters in $\boldsymbol{\alpha}_{\mathrm{v}}$ (one unknown weight, each of the means has two unknown elements, each of the covariance matrices has three unknown elements).

The full solution stems from a solution to the system of 5 matrix equations for the raw moments $\mathbf{M}_{\mathrm{v}}^m$(29), $m = 1, \ldots, 5$, of the GS PDF $p_{\mathrm{v}}(\mathbf{v}_k; \boldsymbol{\alpha}_{\mathrm{v}})$. The matrix equations result in 20 *scalar* equations for *unique* elements of the raw moments $\mathbf{M}_{\mathrm{v}}^m$. The scalar moment equations are derived and summarised in (31)–(43), where the time index k is omitted, i.e., $\mathbf{v} = \mathbf{v}_k$, and the following shorthand notation is used; $\mathbf{v}_i$ for the i-th element of the vector $\mathbf{v}$, $\mathbf{R}_{i,j}$ for an element of the matrix $\mathbf{R}$ at i-th row and j-th column, $B_{\mathrm{v}}^{(1)} = \beta_{\mathrm{v}}^{(1)}, B_{\mathrm{v}}^{(2)} = 1 - \beta_{\mathrm{v}}^{(1)}$, $\sum$ for sum over $i = 1, 2$, and $m_{\mathrm{v}}^{i,j} \triangleq \mathsf{E}[(\mathbf{v}_1)^i(\mathbf{v}_2)^j]$ represents an element of m-th raw moment $\mathbf{M}_{\mathrm{v}}^m$ with $m = i + j$. It can be seen that the higher-order raw moment elements (31)–(43) of the GS PDF $p_{\mathrm{v}}(\mathbf{v}_k; \boldsymbol{\alpha}_{\mathrm{v}})$ are nonlinear functions of sought parameters $\boldsymbol{\alpha}_{\mathrm{v}}$ (30). Given the moments $\mathbf{M}_{\mathrm{v}}^m$ or $m_{\mathrm{v}}^{i,j}$, the sought parameters can be determined using a numerical solution to (31)–(43) (e.g., using the MATLAB routine `lsqnonlin`). Although a theoretical proof of convergence is still an open problem [5], thorough numerical simulations have shown, that a numerical solution to the system of equations (31)–(43) converges to true $\boldsymbol{\alpha}_{\mathrm{v}}$.

It is worth noting that the solution is based on 5 matrix moment equations (for $\mathbf{M}_{\mathrm{v}}^m$) resulting in 20 scalar moment equations (for $m_{\mathrm{v}}^{i,j}$) for computation of 11 unknown GS

[4] The MoM developed by Pearson is suitable not only for the scalar models but also for vector models with independent noise elements.

[5] Besides special cases, the conditions for convergence of estimated GS PDF parameters have not been provided yet (Cerón, 2017).

$$m_{\rm v}^{1,0} = \sum B_{\rm v}^{(i)} \bar{\mathbf{v}}_1^{(i)},\ m_{\rm v}^{0,1} = \sum B_{\rm v}^{(i)} \bar{\mathbf{v}}_2^{(i)}, \tag{31}$$

$$m_{\rm v}^{2,0} = \sum B_{\rm v}^{(i)} \big((\bar{\mathbf{v}}_1^{(i)})^2 + \mathbf{R}_{1,1}^{(i)}\big),\ m_{\rm v}^{0,2} = \sum B_{\rm v}^{(i)} \big((\bar{\mathbf{v}}_2^{(i)})^2 + \mathbf{R}_{2,2}^{(i)}\big),\ m_{\rm v}^{1,1} = \sum B_{\rm v}^{(i)} \big(\bar{\mathbf{v}}_1^{(i)}\bar{\mathbf{v}}_2^{(i)} + \mathbf{R}_{1,2}^{(i)}\big), \tag{32}$$

$$m_{\rm v}^{3,0} = \sum B_{\rm v}^{(i)} \big((\bar{\mathbf{v}}_1^{(i)})^3 + 3\mathbf{R}_{1,1}^{(i)}\bar{\mathbf{v}}_1^{(i)}\big),\ m_{\rm v}^{0,3} = \sum B_{\rm v}^{(i)} \big((\bar{\mathbf{v}}_2^{(i)})^3 + 3\mathbf{R}_{2,2}^{(i)}\bar{\mathbf{v}}_2^{(i)}\big), \tag{33}$$

$$m_{\rm v}^{2,1} = \sum B_{\rm v}^{(i)} \big((\bar{\mathbf{v}}_1^{(i)})^2\bar{\mathbf{v}}_2^{(i)} + \mathbf{R}_{1,1}^{(i)}\bar{\mathbf{v}}_2^{(i)} + 2\mathbf{R}_{1,2}^{(i)}\bar{\mathbf{v}}_1^{(i)}\big),\ m_{\rm v}^{1,2} = \sum B_{\rm v}^{(i)} \big((\bar{\mathbf{v}}_2^{(i)})^2\bar{\mathbf{v}}_1^{(i)} + \mathbf{R}_{2,2}^{(i)}\bar{\mathbf{v}}_1^{(i)} + 2\mathbf{R}_{1,2}^{(i)}\bar{\mathbf{v}}_2^{(i)}\big), \tag{34}$$

$$m_{\rm v}^{4,0} = \sum B_{\rm v}^{(i)} \big((\bar{\mathbf{v}}_1^{(i)})^4 + 6\mathbf{R}_{1,1}^{(i)}(\bar{\mathbf{v}}_1^{(i)})^2 + 3(\mathbf{R}_{1,1}^{(i)})^2\big),\ m_{\rm v}^{0,4} = \sum B_{\rm v}^{(i)} \big((\bar{\mathbf{v}}_2^{(i)})^4 + 6\mathbf{R}_{2,2}^{(i)}(\bar{\mathbf{v}}_2^{(i)})^2 + 3(\mathbf{R}_{2,2}^{(i)})^2\big), \tag{35}$$

$$m_{\rm v}^{3,1} = \sum B_{\rm v}^{(i)} \big((\bar{\mathbf{v}}_1^{(i)})^3\bar{\mathbf{v}}_2^{(i)} + 3\mathbf{R}_{1,1}^{(i)}\bar{\mathbf{v}}_1^{(i)}\bar{\mathbf{v}}_2^{(i)} + 3\mathbf{R}_{1,2}^{(i)}(\bar{\mathbf{v}}_1^{(i)})^2 + 3\mathbf{R}_{1,1}^{(i)}\mathbf{R}_{1,2}^{(i)}\big), \tag{36}$$

$$m_{\rm v}^{1,3} = \sum B_{\rm v}^{(i)} \big(\bar{\mathbf{v}}_1^{(i)}(\bar{\mathbf{v}}_2^{(i)})^3 + 3\mathbf{R}_{2,2}^{(i)}\bar{\mathbf{v}}_1^{(i)}\bar{\mathbf{v}}_2^{(i)} + 3\mathbf{R}_{1,2}^{(i)}(\bar{\mathbf{v}}_2^{(i)})^2 + 3\mathbf{R}_{2,2}^{(i)}\mathbf{R}_{1,2}^{(i)}\big), \tag{37}$$

$$m_{\rm v}^{2,2} = \sum B_{\rm v}^{(i)} \big((\bar{\mathbf{v}}_1^{(i)})^2(\bar{\mathbf{v}}_2^{(i)})^2 + 4\mathbf{R}_{1,2}^{(i)}\bar{\mathbf{v}}_1^{(i)}\bar{\mathbf{v}}_2^{(i)} + \mathbf{R}_{2,2}^{(i)}(\bar{\mathbf{v}}_1^{(i)})^2 + \mathbf{R}_{1,1}^{(i)}(\bar{\mathbf{v}}_2^{(i)})^2 + \mathbf{R}_{1,1}^{(i)}\mathbf{R}_{2,2}^{(i)} + 2(\mathbf{R}_{1,2}^{(i)})^2\big), \tag{38}$$

$$m_{\rm v}^{5,0} = \sum B_{\rm v}^{(i)} \big((\bar{\mathbf{v}}_1^{(i)})^5 + 10\mathbf{R}_{1,1}^{(i)}(\bar{\mathbf{v}}_1^{(i)})^3 + 15(\mathbf{R}_{1,1}^{(i)})^2\bar{\mathbf{v}}_1^{(i)}\big), m^{0,5} = \sum B_{\rm v}^{(i)} \big((\bar{\mathbf{v}}_2^{(i)})^5 + 10\mathbf{R}_{2,2}^{(i)}(\bar{\mathbf{v}}_2^{(i)})^3 + 15(\mathbf{R}_{2,2}^{(i)})^2\bar{\mathbf{v}}_2^{(i)}\big), \tag{39}$$

$$m_{\rm v}^{4,1} = \sum B_{\rm v}^{(i)} \big((\bar{\mathbf{v}}_1^{(i)})^4\bar{\mathbf{v}}_2^{(i)} + 6\mathbf{R}_{1,1}^{(i)}(\bar{\mathbf{v}}_1^{(i)})^2\bar{\mathbf{v}}_2^{(i)} + 4\mathbf{R}_{1,2}^{(i)}(\bar{\mathbf{v}}_1^{(i)})^3 + 3(\mathbf{R}_{1,1}^{(i)})^2\bar{\mathbf{v}}_2^{(i)} + 12\mathbf{R}_{1,1}^{(i)}\mathbf{R}_{1,2}^{(i)}\bar{\mathbf{v}}_1^{(i)}\big), \tag{40}$$

$$m_{\rm v}^{1,4} = \sum B_{\rm v}^{(i)} \big(\bar{\mathbf{v}}_1^{(i)}(\bar{\mathbf{v}}_2^{(i)})^4 + 6\mathbf{R}_{2,2}^{(i)}\bar{\mathbf{v}}_1^{(i)}(\bar{\mathbf{v}}_2^{(i)})^2 + 4\mathbf{R}_{1,2}^{(i)}(\bar{\mathbf{v}}_2^{(i)})^3 + 3(\mathbf{R}_{2,2}^{(i)})^2\bar{\mathbf{v}}_1^{(i)} + 12\mathbf{R}_{2,2}^{(i)}\mathbf{R}_{1,2}^{(i)}\bar{\mathbf{v}}_2^{(i)}\big), \tag{41}$$

$$m_{\rm v}^{3,2} = \sum B_{\rm v}^{(i)} \big((\bar{\mathbf{v}}_1^{(i)})^3(\bar{\mathbf{v}}_2^{(i)})^2 + 3\mathbf{R}_{1,1}^{(i)}\bar{\mathbf{v}}_1^{(i)}(\bar{\mathbf{v}}_2^{(i)})^2 + 6\mathbf{R}_{1,2}^{(i)}(\bar{\mathbf{v}}_1^{(i)})^2\bar{\mathbf{v}}_2^{(i)} + \mathbf{R}_{2,2}^{(i)}(\bar{\mathbf{v}}_1^{(i)})^3 + 3\mathbf{R}_{1,1}^{(i)}\mathbf{R}_{2,2}^{(i)}\bar{\mathbf{v}}_1^{(i)} + 6(\mathbf{R}_{1,2}^{(i)})^2\bar{\mathbf{v}}_1^{(i)} + 6\mathbf{R}_{1,1}^{(i)}\mathbf{R}_{1,2}^{(i)}\bar{\mathbf{v}}_2^{(i)}\big), \tag{42}$$

$$m_{\rm v}^{2,3} = \sum B_{\rm v}^{(i)} \big((\bar{\mathbf{v}}_1^{(i)})^2(\bar{\mathbf{v}}_2^{(i)})^3 + 3\mathbf{R}_{2,2}^{(i)}(\bar{\mathbf{v}}_1^{(i)})^2\bar{\mathbf{v}}_2^{(i)} + 6\mathbf{R}_{1,2}^{(i)}\bar{\mathbf{v}}_1^{(i)}(\bar{\mathbf{v}}_2^{(i)})^2 + \mathbf{R}_{1,1}^{(i)}(\bar{\mathbf{v}}_2^{(i)})^3 + 3\mathbf{R}_{2,2}^{(i)}\mathbf{R}_{1,1}^{(i)}\bar{\mathbf{v}}_2^{(i)} + 6(\mathbf{R}_{1,2}^{(i)})^2\bar{\mathbf{v}}_2^{(i)} + 6\mathbf{R}_{2,2}^{(i)}\mathbf{R}_{1,2}^{(i)}\bar{\mathbf{v}}_1^{(i)}\big), \tag{43}$$

PDF parameters of $\boldsymbol{\alpha}_{\rm v}$. Although, the first 4 matrix moment equations result in a sufficient number of 14 scalar moment equations (31)–(38), the corresponding GS PDF parameters estimate may *not* converge to the true parameters. This observation leads to a conclusion, that for the estimate convergence, is necessary to solve overdetermined system of equations.

Note 4: If the state or measurement noise PDF is Gaussian (which is a limit case of the GS PDF), then the vector of parameters is determined by the first two moments.

Note 5: The system of equations (31)–(43) was derived using a comparison of the GS PDF a central and raw moment and an application of Isserlis' theorem. An exemplary derivation is given in Appendix A.

4.5 Proposed Efficient Partial Solution

Solution to the system of equations (31)–(43) is a numerical optimisation over 11 dimensional space, which may be a computationally demanding procedure. Under certain reasonable assumptions, it is, however, possible to significantly reduce the complexity by a combination of an analytical and a numerical solution to (31)–(43).

Compared to the state-of-the-art solution requiring knowledge of both means, the efficient solution requires *only* one mean, e.g., $\bar{\mathbf{v}}^{(1)}$. Then, the set of 4 unknown parameters is

$$\boldsymbol{\alpha}_{\rm v} = \{\beta_{\rm v}^{(1)}, \mathbf{R}^{(1)}, \bar{\mathbf{v}}^{(2)}, \mathbf{R}^{(2)}\}. \tag{44}$$

and it contains 9 unique elements of unknown parameters. Such GS PDF is particularly useful in modelling noise behaviour with two modes; a *nominal mode* is represented by the first term of the GS centred at zero, i.e., $\bar{\mathbf{v}}^{(1)} = \mathbf{0}_{2\times1}$, (but with unknown covariance matrix $\mathbf{R}^{(1)}$ and weight $\beta_{\rm v}^{(1)}$), an *outlier mode* is represented by the second term with unknown mean $\bar{\mathbf{v}}^{(2)}$ and covariance matrix $\mathbf{R}^{(2)}$ describing a "rare-normal" behaviour of the noise. As an example, application of this two-dimensional and bimodal GS PDF noise model can be found in the radar altimeter based terrain-aided navigation, where the first mode describes the measurement noise while flying over terrain without trees (terrain modelled properly without any bias) and the second mode while flying over the terrain with trees (terrain modelled incorrectly as the trees are not modelled in the terrain map) (Schön et al., 2005).

The efficient partial solution is based on the observation, that the unknown parameters $\mathbf{R}^{(1)}, \bar{\mathbf{v}}^{(2)}, \mathbf{R}^{(2)}$ can be uniquely (and analytically) determined from the first three matrix moment equations leading to (31)–(34) under assumption of the known weight $\beta_{\rm v}^{(1)}$. The weight is, however, unknown and, thus, the equations (31)–(34) are analytically solved for multiple choices of $\beta_{\rm v}^{(1)}$ specified by the grid points $\{\beta_{{\rm v},n}^{(1)}\}_{n=1}^N$. From the resulting N candidates of $\mathbf{R}_n^{(1)}, \bar{\mathbf{v}}_n^{(2)}, \mathbf{R}_n^{(2)}$, such parameters are selected, which lead to fourth order moments closest to true $m_{\rm v}^{4,0}, m_{\rm v}^{0,4}, \ldots, m_{\rm v}^{2,2}$. Particularly, the computation of $\boldsymbol{\alpha}_{\rm v}$ (44) is as follows:

(i) Define an equally distributed grid covering the region where the weight $\beta_{\rm v}^{(1)}$ can lie, i.e., determine N grid points $\{\beta_{{\rm v},n}^{(1)}\}_{n=1}^N$ such that $0 < \beta_{{\rm v},1}^{(1)} < \beta_{{\rm v},2}^{(1)} < \ldots < \beta_{{\rm v},N}^{(1)} < 1$. Number of the points is driven by required accuracy and available computational power.

(ii) Repeat the following steps for $n = 1, 2, \ldots, N$:
- Recall $\bar{\mathbf{v}}^{(1)} = \mathbf{0}_{2\times1}$ is known, select the grid point $\beta_{{\rm v},n}^{(1)}$ for given n, and compute $B_{{\rm v},n}^{(1)} = \beta_{{\rm v},n}^{(1)}, B_{{\rm v},n}^{(2)} = 1 - \beta_{{\rm v},n}^{(1)}$.

- Compute elements of the n-th candidate mean $\bar{\mathbf{v}}_n^{(2)} = [\bar{\mathbf{v}}_{1,n}^{(2)}; \bar{\mathbf{v}}_{2,n}^{(2)}]^T$ from (31) as
$$\bar{\mathbf{v}}_{1,n}^{(2)} = \left(m_{\mathrm{v}}^{1,0}\right) \Big/ \left(B_{\mathrm{v},n}^{(2)}\right), \bar{\mathbf{v}}_{2,n}^{(2)} = \left(m_{\mathrm{v}}^{0,1}\right) \Big/ \left(B_{\mathrm{v},n}^{(2)}\right). \quad (45)$$
- Compute elements of the n-th candidate covariance matrix $\mathbf{R}_n^{(2)} = \begin{bmatrix} \mathbf{R}_{1,1,n}^{(2)} & \mathbf{R}_{1,2,n}^{(2)} \\ \mathbf{R}_{1,2,n}^{(2)} & \mathbf{R}_{2,2,n}^{(2)} \end{bmatrix}$ from (33), (34) as
$$\mathbf{R}_{1,1,n}^{(2)} = \left(m_{\mathrm{v}}^{3,0} - B_{\mathrm{v},n}^{(2)} (\bar{\mathbf{v}}_{1,n}^{(2)})^3\right) \Big/ \left(3\bar{\mathbf{v}}_{1,n}^{(2)} B_{\mathrm{v},n}^{(2)}\right), \quad (46)$$
$$\mathbf{R}_{2,2,n}^{(2)} = \left(m_{\mathrm{v}}^{0,3} - B_{\mathrm{v},n}^{(2)} (\bar{\mathbf{v}}_{2,n}^{(2)})^3\right) \Big/ \left(3\bar{\mathbf{v}}_{2,n}^{(2)} B_{\mathrm{v},n}^{(2)}\right), \quad (47)$$
$$\mathbf{R}_{1,2,n}^{(2)} = \frac{m_{\mathrm{v}}^{2,1} - B_{\mathrm{v},n}^{(2)} \left((\bar{\mathbf{v}}_{1,n}^{(2)})^2 \bar{\mathbf{v}}_{2,n}^{(2)} + \mathbf{R}_{1,1,n}^{(2)} \bar{\mathbf{v}}_{2,n}^{(2)}\right)}{2\bar{\mathbf{v}}_{1,n}^{(2)} B_{\mathrm{v},n}^{(2)}}. \quad (48)$$
- Compute elements of the n-th candidate covariance matrix $\mathbf{R}_n^{(1)} = \begin{bmatrix} \mathbf{R}_{1,1,n}^{(1)} & \mathbf{R}_{1,2,n}^{(1)} \\ \mathbf{R}_{1,2,n}^{(1)} & \mathbf{R}_{2,2,n}^{(1)} \end{bmatrix}$ from (32) as
$$\mathbf{R}_{1,1,n}^{(1)} = \frac{m_{\mathrm{v}}^{2,0} - B_{\mathrm{v},n}^{(2)} ((\bar{\mathbf{v}}_{1,n}^{(2)})^2 + \mathbf{R}_{1,1,n}^{(2)})}{B_{\mathrm{v},n}^{(1)}}, \quad (49)$$
$$\mathbf{R}_{2,2,n}^{(1)} = \frac{m_{\mathrm{v}}^{0,2} - B_{\mathrm{v},n}^{(2)} ((\bar{\mathbf{v}}_{2,n}^{(2)})^2 + \mathbf{R}_{2,2,n}^{(2)})}{B_{\mathrm{v},n}^{(1)}}, \quad (50)$$
$$\mathbf{R}_{1,2,n}^{(1)} = \frac{m_{\mathrm{v}}^{1,1} - B_{\mathrm{v},n}^{(2)} (\bar{\mathbf{v}}_{1,n}^{(2)} \bar{\mathbf{v}}_{2,n}^{(2)} + \mathbf{R}_{1,2,n}^{(2)})}{B_{\mathrm{v},n}^{(1)}}. \quad (51)$$
- Compute the n-th candidate of 4-th order moments elements, further denoted as $\check{m}_{\mathrm{v},n}^{4,0}, \check{m}_{\mathrm{v},n}^{0,4}, \check{m}_{\mathrm{v},n}^{3,1}, \check{m}_{\mathrm{v},n}^{1,3}$, and $\check{m}_{\mathrm{v},n}^{2,2}$ according to (35)–(38), where $\bar{\mathbf{v}}^{(1)} = \mathbf{0}_{2\times 1}$, $\bar{\mathbf{v}}^{(1)}$ is substituted with $\bar{\mathbf{v}}_n^{(1)}$, $\mathbf{R}^{(1)}$ with $\mathbf{R}_n^{(1)}$, $\mathbf{R}^{(2)}$ with $\mathbf{R}_n^{(2)}$, and $B_{\mathrm{v}}^{(1)} = \beta_{\mathrm{v},n}^{(1)}, B_{\mathrm{v}}^{(2)} = 1 - \beta_{\mathrm{v},n}^{(1)}$.
- Compute a distance between the n-th candidate of 4-th order moments and given (true) 4-th order moments as
$$d_n = \left\| \begin{matrix} \left[\check{m}_{\mathrm{v},n}^{4,0} \; \check{m}_{\mathrm{v},n}^{0,4} \; \check{m}_{\mathrm{v},n}^{3,1} \; \check{m}_{\mathrm{v},n}^{1,3} \; \check{m}_{\mathrm{v},n}^{2,2} \right]^T \\ - \left[m_{\mathrm{v}}^{4,0} \; m_{\mathrm{v}}^{0,4} \; m_{\mathrm{v}}^{3,1} \; m_{\mathrm{v}}^{1,3} \; m_{\mathrm{v}}^{2,2} \right]^T \end{matrix} \right\|, \quad (52)$$
where the notation $\|\cdot\|$ denotes a vector norm.

(iii) Select optimal candidate GS PDF parameters index according to
$$n^* = \arg\min_n d_n \quad (53)$$
and determine sought GS PDF parameter set as
$$\boldsymbol{\alpha}_{\mathrm{v}} = \{\beta_{\mathrm{v},n^*}^{(1)}, \mathbf{R}_{n^*}^{(1)}, \bar{\mathbf{v}}_{n^*}^{(2)}, \mathbf{R}_{n^*}^{(2)}\}. \quad (54)$$

4.6 Properties of Proposed Solutions

The full solution computes the whole set of five parameters (30) on the basis of five moment matrix equations using a numerical optimisation over 11 dimensional space.

The partial efficient solution computes the reduced set of four parameters (44) on the basis of four moment matrix equations using a combined analytical and numerical optimisation over one dimensional bounded space. This solution is, thus, more computationally feasible, but at the expense of the generality.

5. ESTIMATION OF GAUSSIAN SUM PARAMETERS FROM ESTIMATED MOMENTS

The solutions discussed in the previous section compute the noise GS PDF set of parameters under the assumption of given noise (higher-order) moments. The noise moments are, however, generally not known, but they can be estimated by the method reviewed in Section 3 and used instead of the true ones. The final algorithm of the MDM for the parameter estimation of the noise GS PDF reads:

(i) Measure data $\mathbf{z}_k, \forall k$.
(ii) Based on the known system matrices $\mathbf{F}_k, \mathbf{H}_k, \forall k$ estimate $m_{\max}$ moments of the state and measurement noise according to the algorithm given in Section 3, i.e., determine $\hat{\mathbf{M}}_{\mathrm{w}}^m$, $\hat{\mathbf{M}}_{\mathrm{v}}^m$ for $m = 1, \ldots, m_{\max}$.
(iii) Estimate the GS PDF sets of parameters $\boldsymbol{\alpha}_{\mathrm{w}}$ and $\boldsymbol{\alpha}_{\mathrm{v}}$ according to the solutions proposed in Section 4, where the true moments $\mathbf{M}_{\mathrm{w}}^m$, $\mathbf{M}_{\mathrm{v}}^m$ are substituted with their estimates $\hat{\mathbf{M}}_{\mathrm{w}}^m$, $\hat{\mathbf{M}}_{\mathrm{v}}^m$.

Note 6: The proposed candidate selection in Section 4.5 is based on the comparison of "expected" candidate 4-th moments with observed ones. Besides, the 4-th order moments, other higher-order moments can be used as well. However, the GS PDF candidate can also be selected on the basis of empirical PDFs as follows

- Estimate candidate parameters for the state and measurement noises.
- Based on the candidate noise GS PDFs generate noises' samples and compute "expected" samples of the AMPE according to (11). The samples form the candidate AMPE empirical distribution.
- Compare the candidate AMPE empirical distribution with the measured empirical distribution, which is given by the measurement-based AMPE samples (10). Select such candidate AMPE empirical distribution which is, in some sense, closest to the measured empirical distribution.

This approach can also be used in the proposed full solution as a verification of the estimated parameters $\boldsymbol{\alpha}_{\mathrm{v}}$.

6. NUMERICAL ILLUSTRATION

In this section, the performance of the developed MDM is illustrated. Let the time-varying LGSM (1), (2) with $n(x) = 1$, $n(z) = 2$ defined as
$$F_k = 0.9 + 0.1\sin(5k/\tau), \mathbf{H}_k = \begin{bmatrix} 2 + \sin(13k/\tau) \\ \cos(9k/\tau) \end{bmatrix}, \quad (55)$$
be considered with $k = 0, \ldots, \tau$ with $\tau = 10^5$ and $\tau = 10^6$.

For the simulation purposes, the state noise is Gaussian
$$p_{\mathrm{w}}(w_k; \boldsymbol{\alpha}_{\mathrm{w}}) = \mathcal{N}\{w_k; \bar{w}, Q\}, \quad (56)$$
with the mean $\bar{w} = 1$ and the variance $Q = 1$, i.e., the set of parameters is $\boldsymbol{\alpha}_{\mathrm{w}} = \{\bar{w}, Q\}$ and the measurement noise $\mathbf{v}_k = [\mathbf{v}_{k,1}, \mathbf{v}_{k,2}]^T$ is described by the bi-modal GS PDF (4) with
$$\boldsymbol{\alpha}_{\mathrm{v}} = \{\beta_{\mathrm{v}}^{(1)}, \bar{\mathbf{v}}^{(1)}, \mathbf{R}^{(1)}, \bar{\mathbf{v}}^{(2)}, \mathbf{R}^{(2)}\}, \quad (57)$$
where the parameters are $\beta_{\mathrm{v}}^{(1)} = 0.8$, $\bar{\mathbf{v}}^{(1)} = \left[\begin{smallmatrix} 4 \\ -3 \end{smallmatrix}\right]$, $\mathbf{R}^{(1)} = \left[\begin{smallmatrix} 3 & 0.5 \\ 0.5 & 2 \end{smallmatrix}\right]$, $\bar{\mathbf{v}}^{(2)} = \left[\begin{smallmatrix} 6 \\ 7 \end{smallmatrix}\right]$, $\mathbf{R}^{(2)} = \left[\begin{smallmatrix} 4 & 2 \\ 2 & 4 \end{smallmatrix}\right]$.

The sets of parameters $\boldsymbol{\alpha}_{\mathrm{w}}$ and $\boldsymbol{\alpha}_{\mathrm{v}}$ are estimated from the measured data by the MDM, summarised in Section 5, using the proposed *full solution* [6].

[6] MATLAB optimisation routine `lsqnonlin` was used with randomly generated initial parameters.

Table 1. True and estimated moments and parameters of the state noise $\mathbf{w}_k$.

	$m_{\mathrm{w}}^{1,0}$	$m_{\mathrm{w}}^{2,0}$	$m_{\mathrm{w}}^{3,0}$	$m_{\mathrm{w}}^{4,0}$	$m_{\mathrm{w}}^{5,0}$	$\bar{w}$	Q
True	1	2	4	10	26	1	1
Avg. of est. ($\tau = 10^5$)	1	2	3.999	10.003	26.012	1	1
Avg. of est. ($\tau = 10^6$)	1	2	4	10.001	25.998	1	1
STD of est. ($\tau = 10^5$)	0.004	0.034	0.118	0.628	2.74	0.004	0.032
STD of est. ($\tau = 10^6$)	0.001	0.011	0.038	0.2	0.874	0.001	0.01

Table 2. True and estimated selected moments of the measurement noise $\mathbf{v}_k$.

	$m_{\mathrm{v}}^{1,0}$	$m_{\mathrm{v}}^{0,1}$	$m_{\mathrm{v}}^{2,0}$	$m_{\mathrm{v}}^{1,1}$	$m_{\mathrm{v}}^{3,0}$	$m_{\mathrm{v}}^{2,1}$	$m_{\mathrm{v}}^{4,0}$	$m_{\mathrm{v}}^{3,1}$	$m_{\mathrm{v}}^{5,0}$	$m_{\mathrm{v}}^{4,1}$
True	4.4	-1	23.2	-0.4	137.6	18.4	898.4	234	6358.4	2341.6
Avg. of est. ($\tau = 10^5$)	4.399	-1	23.196	-0.403	137.575	18.369	898.257	233.739	6357.979	2340.359
Avg. of est. ($\tau = 10^6$)	4.4	-1	23.199	-0.402	137.591	18.387	898.303	233.842	6355.802	2339.178
STD of est. ($\tau = 10^5$)	0.031	0.014	0.295	0.098	2.693	0.762	27.854	8.023	317.64	90.45
STD of est. ($\tau = 10^6$)	0.01	0.004	0.091	0.03	0.824	0.24	8.525	2.53	98.342	27.662

Table 3. True and estimated parameters of the measurement noise $\mathbf{v}_k$.

	$\beta_{\mathrm{v}}^{(1)}$	$\bar{\mathbf{v}}^{(1)}$		$\bar{\mathbf{v}}^{(2)}$		$\mathbf{R}^{(1)}$			$\mathbf{R}^{(2)}$		
True	0.8	4	-3	6	7	3	0.5	2	4	2	4
Avg. of est. ($\tau = 10^5$)	0.803	3.989	-2.988	6.061	7.078	3.04	0.493	2.025	3.901	1.865	3.848
Avg. of est. ($\tau = 10^6$)	0.801	3.999	-2.997	6.015	7.02	3.002	0.497	2.005	3.975	1.965	3.957
STD of est. ($\tau = 10^5$)	0.025	0.132	0.099	0.359	0.417	0.456	0.071	0.214	0.575	0.729	0.805
STD of est. ($\tau = 10^6$)	0.012	0.065	0.039	0.165	0.178	0.224	0.029	0.089	0.264	0.326	0.315

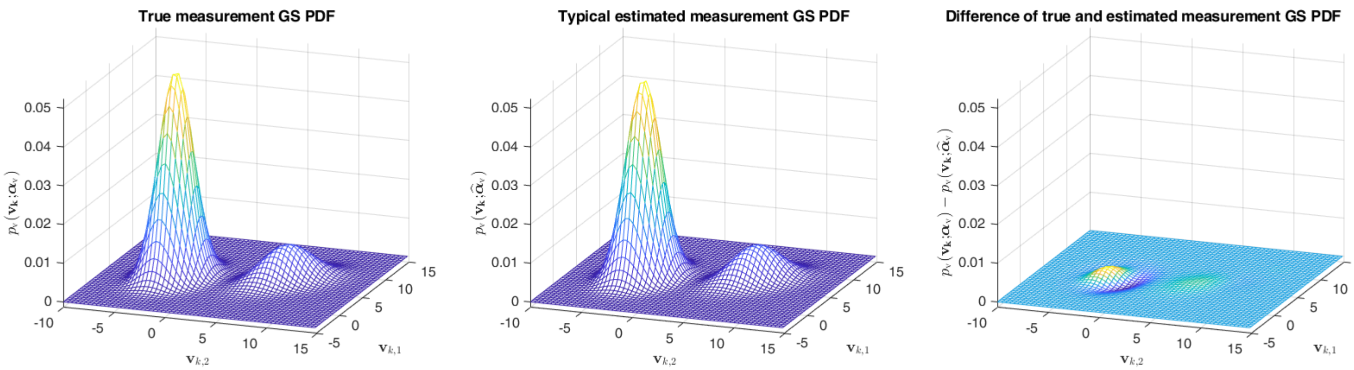

Fig. 1. Illustration of the true GS PDF, its typical estimate, and the respective error.

The resulting estimates for $M = 10^3$ Monte-Carlo (MC) simulations are shown in Tables 1–3. In Table 1, true and estimated moments $\mathbf{M}_{\mathrm{w}}^{m}, \forall m$, and the set of parameter $\boldsymbol{\alpha}_{\mathrm{w}}$ of the state noise w_k are given. The estimate quality is characterised by two criteria:

- *Average value of the MC estimates* illustrating biasedness of the MDM estimates,
- *Standard deviation (STD) of the MC estimates.*

Similarly, selected [7] moments elements and all parameters of the measurement noise $\mathbf{v}_k$ described by the GS PDF are summarised in Tables 2 and 3, respectively. For the sake of completeness, the GS PDF $p_{\mathrm{v}}(\mathbf{v}_k; \boldsymbol{\alpha}_{\mathrm{v}})$ (4) with true parameters (57) is illustrated in Fig. 1 together with the GS PDF $p_{\mathrm{v}}(\mathbf{v}_k; \hat{\boldsymbol{\alpha}}_{\mathrm{v}})$, where $\hat{\boldsymbol{\alpha}}_{\mathrm{v}}$ is a "typical" estimate of the set $\boldsymbol{\alpha}_{\mathrm{v}}$ with the parameters $\hat{\beta}_{\mathrm{v}}^{(1)} = 0.808$, $\hat{\bar{\mathbf{v}}}^{(1)} = \left[\begin{smallmatrix} 3.984 \\ -2.956 \end{smallmatrix}\right]$, $\hat{\mathbf{R}}^{(1)} = \left[\begin{smallmatrix} 3.125 & 0.486 \\ 0.486 & 2.108 \end{smallmatrix}\right]$, $\hat{\bar{\mathbf{v}}}^{(2)} = \left[\begin{smallmatrix} 6.089 \\ 7.119 \end{smallmatrix}\right]$, $\hat{\mathbf{R}}^{(2)} = \left[\begin{smallmatrix} 3.848 & 1.802 \\ 1.802 & 3.812 \end{smallmatrix}\right]$. In addition, the figure also visualise a difference between these two GS PDFs.

The tables demonstrate that the MDM provides *unbiased* estimates of the moments and parameters of the noise density in the form of the Gaussian and GS PDF. The estimates also *converge* to the true values with increasing number of data τ.

Note 7: Tables 1 and 2 reveal that estimate STDs grow with increasing order m. Therefore, the less estimated moments is used for the parameter estimation, the lower STD of the parameter estimate can be expected. Simulations also indicate that the MDM with the *partial efficient solution* to parameters computation reduces the STD of the estimate STD almost by a factor of four (these results not shown in the paper because of the space reasons).

[7] First five moments of two-dimensional random variable result in 20 unique moment elements (please see (31)–(43)) out of which only 10 are given in Table 2 because of the space reasons.

7. CONCLUDING REMARKS

The paper dealt with the estimation of the parameters of the GS PDF assumed noises of linear time-varying state-space models. The proposed solution extends applicability of the measurement difference method, which estimates the (higher-order) moments of the state and measurement noise from the measured data. The extension of the method lies in the explicit derivation of the relations and techniques allowing transformation of the noise moments into the noise parameters. The theoretical results were thoroughly discussed and illustrated in a numerical simulation.

ACKNOWLEDGEMENT

This work was supported by the Czech Science Foundation (GACR) under grant GA 20-06054J.

Appendix A. DERIVATION OF GS PDF HIGHER-ORDER RAW MOMENT

Because of lengthy and tedious derivation of higher-order moment elements for the GS PDF, the final relations were summarised only by (31)–(43). However, for the sake of clarity a sketch of the derivation is given for 4-th order raw moment element (36) defined as

$$m_{\mathrm{v}}^{3,1} = \mathsf{E}[\mathbf{v}_1^3\mathbf{v}_2] = \mathsf{E}_{p(\mathbf{v})}[\mathbf{v}_1^3\mathbf{v}_2] = \sum_{i=1}^{N_{\mathrm{v}}} \beta_{\mathrm{v}}^{(i)} \mathsf{E}_{\mathcal{N}_{\mathrm{v}}^{(i)}}[\mathbf{v}_1^3\mathbf{v}_2], \quad \text{(A.1)}$$

where $\mathcal{N}_{\mathrm{v}}^{(i)}$ denotes $\mathcal{N}\{\mathbf{v}_k; \bar{\mathbf{v}}^{(i)}, \mathbf{R}^{(i)}\}$. The derivation starts from expansion of the i-th term central moment counterpart of the GS PDF

$$\begin{aligned}\mathsf{E}_{\mathcal{N}_{\mathrm{v}}^{(i)}}[(\mathbf{v}_1 - \bar{\mathbf{v}}_1^{(i)})^3(\mathbf{v}_2 - \bar{\mathbf{v}}_2^{(i)})] = \mathsf{E}_{\mathcal{N}_{\mathrm{v}}^{(i)}}[\mathbf{v}_1^3\mathbf{v}_2 - \mathbf{v}_1^3\bar{\mathbf{v}}_2^{(i)} \\ - 3\mathbf{v}_1^2\mathbf{v}_2\bar{\mathbf{v}}_1^{(i)} + 3\mathbf{v}_1^2\bar{\mathbf{v}}_1^{(i)}\bar{\mathbf{v}}_2^{(i)} + 3\mathbf{v}_1\mathbf{v}_2(\bar{\mathbf{v}}_1^{(i)})^2 \\ - 3\mathbf{v}_1(\bar{\mathbf{v}}_1^{(i)})^2\bar{\mathbf{v}}_2^{(i)} - (\bar{\mathbf{v}}_1^{(i)})^3\mathbf{v}_2 + (\bar{\mathbf{v}}_1^{(i)})^3\bar{\mathbf{v}}_2^{(i)}], \quad \text{(A.2)}\end{aligned}$$

which can be further simplified to the form

$$\begin{aligned}\mathsf{E}_{\mathcal{N}_{\mathrm{v}}^{(i)}}[(\mathbf{v}_1 - \bar{\mathbf{v}}_1^{(i)})^3(\mathbf{v}_2 - \bar{\mathbf{v}}_2^{(i)})] = \mathsf{E}_{\mathcal{N}_{\mathrm{v}}^{(i)}}[\mathbf{v}_1^3\mathbf{v}_2] - (\bar{\mathbf{v}}_1^{(i)})^3\bar{\mathbf{v}}_2^{(i)} \\ - 3\bar{\mathbf{v}}_1^{(i)}\bar{\mathbf{v}}_2^{(i)}\mathbf{R}_{1,1}^{(i)} - 3(\bar{\mathbf{v}}_1^{(i)})^2\mathbf{R}_{1,2}^{(i)}. \quad \text{(A.3)}\end{aligned}$$

Application of Isserlis' theorem to the central moment (left-hand side of (A.3)) leads to

$$\mathsf{E}_{\mathcal{N}_{\mathrm{v}}^{(i)}}[(\mathbf{v}_1 - \bar{\mathbf{v}}_1^{(i)})^3(\mathbf{v}_2 - \bar{\mathbf{v}}_2^{(i)})] = 3\mathbf{R}_{1,1}^{(i)}\mathbf{R}_{1,2}^{(i)}. \quad \text{(A.4)}$$

Combination of (A.3) and (A.4) leads to the final form of the i-th GS PDF 4-th order moment

$$\begin{aligned}\mathsf{E}_{\mathcal{N}_{\mathrm{v}}^{(i)}}[\mathbf{v}_1^3\mathbf{v}_2] = (\bar{\mathbf{v}}_1^{(i)})^3\bar{\mathbf{v}}_2^{(i)} + 3\mathbf{R}_{1,1}^{(i)}\bar{\mathbf{v}}_1^{(i)}\bar{\mathbf{v}}_2^{(i)} + 3\mathbf{R}_{1,2}^{(i)}(\bar{\mathbf{v}}_1^{(i)})^2 \\ + 3\mathbf{R}_{1,1}^{(i)}\mathbf{R}_{1,2}^{(i)}. \quad \text{(A.5)}\end{aligned}$$

The weighted sum of (A.5) in (A.1), $\forall i$, results in (36).

REFERENCES

Anderson, B.D.O. and Moore, J.B. (1979). *Optimal Filtering.* Prentice Hall, New Jersey.

Bélanger, P.R. (1974). Estimation of noise covariance matrices for a linear time-varying stochastic process. *Automatica*, 10(3), 267–275.

Cerón, C.E.A. (2017). *Algebraic Statistics of Gaussian Mixtures.* Ph.D. thesis, Technical University Berlin.

Duník, J., Straka, O., and Kost, O. (2018a). Design of measurement difference autocovariance method for estimation of process and measurement noise covariances. *Automatica*, 90, 16–24.

Duník, J., Straka, O., Kost, O., and Havlík, J. (2017). Noise covariance matrices in state-space models: A survey and comparison - part I. *International Journal of Adaptive Control and Signal Processing*, 31, 1505–1543.

Duník, J., Straka, O., Noack, B., Steinbring, J., and Hanebeck, U.D. (2018b). Directional splitting of Gaussian density in non-linear random variable transformation. *IET Signal Processing*, 12(9), 1073–1081.

Gustafsson, F., Gunnarsson, F., Bergman, N., Forssell, U., Jansson, J., Karlsson, R., and Nordlund, P.J. (2002). Particle filters for positioning, navigation, and tracking. *IEEE Transactions on Signal Processing*, 50(2), 425–437.

Hanebeck, U.D., Briechle, K., and Rauh, A. (2003). Progressive bayes: A new framework for nonlinear state estimation. In *Proceedings of SPIE, Vol. 5099, AeroSense Symposium*, 256–267. Orlando, Florida.

Kost, O., Duník, J., and Straka, O. (2018). Noise moment and parameter estimation of state-space model. In *Proceedings of the 18th IFAC Symposium on System Identification.* Stockholm, Sweden.

Mehra, R.K. (1970). On the identification of variances and adaptive filtering. *IEEE Transactions on Automatic Control*, 15(2), 175–184.

Nurminen, H., Ardeshiri, T., Piché, R., and Gustafsson, F. (2018). Robust inference for state-space models with skewed measurement noise. *IEEE Signal Processing Letters*, 22(11), 1898–1902.

Odelson, B.J., Rajamani, M.R., and Rawlings, J.B. (2006). A new autocovariance least-squares method for estimating noise covariances. *Automatica*, 42(2), 303–308.

Roth, M., Özkan, E., and Gustafsson, F. (2013). A Student's-t filter for heavy tailed process and measurement noise. In *Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing.* Vancouver, Canada.

Särkkä, S. and Nummenmaa, A. (2009). Recursive noise adaptive Kalman filtering by variational Bayesian approximations. *IEEE Transactions on Automatic Control*, 54(3), 596–600.

Schön, T., Gustafsson, F., and Nordlund, P. (2005). Marginalized particle filters for mixed linear/nonlinear state-space models. *IEEE Transactions on Signal Processing*, 53(7), 2279–2289.

Williams, J.L. and Maybeck, P.S. (2003). Cost-function-based Gaussian mixture reduction for target tracking. In *Proceedings of the 6th International Conference on Information Fusion.* Queensland, Australia.