18th International Conference on Information Fusion
Washington, DC - July 6-9, 2015

Approximation of Powers of Gaussian Mixtures

Jiří Ajgl, Miroslav Šimandl, Jindřich Duník
European Centre of Excellence – New Technologies for Information Society and
Department of Cybernetics, Faculty of Applied Sciences, University of West Bohemia,
Pilsen, Czech Republic.
Email: jiriajgl@kky.zcu.cz, simandl@kky.zcu.cz

***Abstract*—Gaussian mixtures are used to approximate general probability density functions. In fusion problems, the weighted geometric mean of densities exploits density powers. Since the operation of exponentiation of Gaussian mixtures is not closed, the problem of approximating the mixture powers arises. This paper proposes and inspects several such approximations. The first is based on a limit case, the second is based on an optimisation, the third is given by a numerical solution to the optimisation problem and the last one is given by the exact power of an approximation of the Gaussian mixture. The numerical examples compare component weights and the means and covariance matrices.**

***Index Terms*—state estimation, Gaussian mixtures, information fusion, Integrated Squared Distance, sigma-points**

I. Introduction

Nonlinearities seldom allow estimation problems to be solved exactly. The state estimation [1] of nonlinear systems often employs the Bayesian approach [2] as a tool for making further inference or actions. The system state is considered to be a random variable with a known prior probability density, while the measurements are generated randomly according to known probability distributions. The observation of a measurement changes the knowledge of the state, the Bayesian recursive relations *prescribe* the change. Supposing the prior knowledge of the system is exact, the posterior knowledge is *objective*.

Since general probability density functions cannot be handled, finite representations have been introduced. Nonlinear filters can use analytical approximations [3], [4], numerical computations [5] or simulations [6]. Here, the representation of probability density functions by Gaussian mixtures will be considered, while it will be insignificant, whether the Gaussian mixtures have been obtained directly via mixture reduction [7]–[10] or indirectly via weighted points or weighted samples [11], [12].

While the Bayesian estimation can combine the prior knowledge with measurements to obtain the posterior knowledge, it gets stuck in general fusion problems. That is, it cannot successfully combine two parts of posterior knowledge, where each part is given by the prior knowledge and a part of the measurements. The reason is that the Bayesian estimation is founded on the description of the mechanism that generates the measurements. Therefore, a direct unavailability of the measurements or the lack of the mechanism description renders the objective knowledge unachievable. Consequently, the fusion problems resort to *subjective* beliefs. Instead of describing the system, the beliefs describe human opinions about the system. The choice of a fusion rule from a plethora of them [13]–[16] is an identification problem, i.e. the designer of a fusion system has to decide whether a given fusion rule *describes* the human behaviour adequately. Any subjective belief does not describe the system state objectively and thus, as it was noted at the very beginning, the probability densities are mere *tools* for making further inference or actions.

This paper focuses on the fusion rule given by the weighted geometric mean of densities [17]–[22], since this rule is well explored in literature and established in practice. Besides the simplicity of its form, it is possible to highlight the property of providing the density that minimises an overall distance.

The considered fusion rule provides a probability density function that is proportional to the product of powers of the probability density functions that are being combined, while the sum of the exponents is equal to one. The exponentiation of finite representations of general probability density functions requires some techniques to keep the representation. In the case of densities represented by the weighted samples, the densities that are being combined have to be represented by the same samples before the fusion can be performed, see [23], [24]. In the case of Gaussian mixtures, the powers of the mixtures have to be replaced by other Gaussian mixtures. A simple approximation was proposed in [25]. An approach to approximate the weighted geometric mean of Gaussian mixtures has been proposed recently in [26]. However, the behaviour of the proposed approximations has not been studied in detail, and so there may be a space for improvements.

The goal of this paper is to revise the existing approximations and to inspect them. Namely, the goal is to propose an approximation that is asymptotically exact, to propose an approximation of the power of a Gaussian mixture that is based on the minimisation of a distance criterion and to compare the proposed approximations on a bunch of examples. The power of a Gaussian mixture obtained by a mixture reduction is also to be inspected, namely from the perspective of means and covariances.

The problem is formulated in Section II. The approximation of a nearly unit power of a Gaussian mixture with well separated components and the approximation of the optimisation based approach are presented in Section III. Section IV inspects the approximations for one dimensional mixtures with two and three components and for two dimensional mixtures with two components. Section V summarises the results.

II. Problem settings

Let $\mathbf{x}$ be the state of the system to be estimated and let the dimension of $\mathbf{x}$ be given by $\dim(\mathbf{x}) = n$. Let the probability density p be a Gaussian mixture with N components,

$$p(\mathbf{x}) = \sum_{i=1}^{N} a_i p_i(\mathbf{x}), \quad p_i(\mathbf{x}) = \mathcal{N}(\mathbf{x} : \mathbf{m}_i, \mathbf{P}_i), \tag{1}$$

where a_i, $\mathbf{m}_i$, $\mathbf{P}_i$ with $i = 1, \ldots, N$ are the parameters of the mixture p, it holds $\sum_{i=1}^{N} a_i = 1$ and $\mathcal{N}$ denotes the Gaussian density with the corresponding parameters,

$$\mathcal{N}(\mathbf{x} : \mathbf{m}_i, \mathbf{P}_i) = \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{P}_i|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m}_i)^{\mathrm{T}} \mathbf{P}_i^{-1} (\mathbf{x}-\mathbf{m}_i)}. \tag{2}$$

Denote the ω power of the probability density p by p^{ω}, where ω should be read as an upper index of the symbol p, and define p^{ω} by $p^{\omega}(\mathbf{x}) \propto (p(\mathbf{x}))^{\omega}$. That means, define

$$p^{\omega}(\mathbf{x}) = \frac{1}{k(\omega)} (p(\mathbf{x}))^{\omega}, \quad k(\omega) = \int_{\mathrm{R}^n} (p(\mathbf{x}))^{\omega} \, \mathrm{d}\mathbf{x}, \tag{3}$$

where $k(\omega)$ is a normalising constant dependent on ω.

The central problem of this paper is to approximate the probability density p^{ω} by another Gaussian mixture π with the same number of components N,

$$\pi(\mathbf{x};\omega) = \sum_{i=1}^{N} \alpha_i(\omega) \pi_i(\mathbf{x};\omega), \tag{4a}$$

$$\pi_i(\mathbf{x};\omega) = \mathcal{N}(\mathbf{x} : \mu_i(\omega), \Pi_i(\omega)). \tag{4b}$$

That is, the problem is to find suitable parameters α_i, μ_i, Π_i, where $i = 1, \ldots, N$, for the given exponent ω and to evaluate the used approach.

The following approximations are of the interest. The first one is an approximation of the power density p^{ω}, while the parameters of π are to be given explicitly. The second one is an approximation π that minimises a distance to p^{ω}, while the criterion will be given by the integrated square error for simplicity. The last one is considered for comparison purposes and does not have N components as in (4), but only one. Specifically, the last approximation is the exact power of an approximation of p, namely of the approximation of p by a moment-matched Gaussian density.

III. Proposed approximations

The three approximation mentioned in Section II are treated in Sections III-A, III-B and III-C, respectively.

A. Approximation based on a limit case

The first approximation is based on the following simple observations.

Lemma 1. *Let q be a mixture density, $q(\mathbf{x}) = \sum_{i=1}^{N} a_i q_i(\mathbf{x})$, such that the components q_i, $i = 1, \ldots, N$, have disjoint supports Ω_i, $\Omega_i = \{\mathbf{x} : p_i(\mathbf{x}) > 0\}$, $\Omega_i \cap \Omega_j = \emptyset$, $i \neq j$, $j = 1, \ldots, N$. Then, the power q^{ω} (3) is given by*

$$q^{\omega}(\mathbf{x})|_{\mathbf{x}\in\Omega_i} = \frac{1}{k(\omega)} (a_i)^{\omega} k_i(\omega) q_i^{\omega}(\mathbf{x}), \tag{5}$$

where q_i^{ω} and $k_i(\omega)$ are given by (3) for the corresponding component q_i, and is equal to zero otherwise, $q^{\omega}(\mathbf{x})|_{\mathbf{x}\notin\Omega} = 0$, $\Omega = \bigcup_{i=1}^{N} \Omega_i$.

Proof. See that $q(\mathbf{x})|_{\mathbf{x}\in\Omega_i} = a_i q_i(\mathbf{x})$. □

Lemma 2. *For Gaussian components p_i, the power p_i^{ω} and the normalising constant $k_i(\omega)$ are given by*

$$p_i^{\omega}(\mathbf{x}) = \mathcal{N}(\mathbf{x} : \mathbf{m}_i, \frac{1}{\omega}\mathbf{P}_i), \tag{6a}$$

$$k_i(\omega) = (2\pi)^{(1-\omega)\frac{n}{2}} \omega^{-\frac{n}{2}} |\mathbf{P}_i|^{\frac{1-\omega}{2}}. \tag{6b}$$

Proof. From (2), it follows

$$(p_i(\mathbf{x}))^{\omega} = \frac{1}{(2\pi)^{\frac{n}{2}\omega} |\mathbf{P}_i|^{\frac{\omega}{2}}} (2\pi)^{\frac{n}{2}} |\frac{1}{\omega}\mathbf{P}_i|^{\frac{1}{2}} \mathcal{N}(\mathbf{x} : \mathbf{m}_i, \frac{1}{\omega}\mathbf{P}_i). \tag{7}$$

and the statement becomes evident. □

Suppose that the components p_i of the Gaussian mixture p (1) fulfil $\Omega_i(\omega,\epsilon) \cap \Omega_j(\omega,\epsilon) = \emptyset$, $i \neq j$, $j = 1, \ldots, N$, where $\Omega_i(\omega,\epsilon) = \{\mathbf{x} : p_i^{\omega}(\mathbf{x}) > \epsilon\}$, for a small value of ϵ. That is, the means $\mathbf{m}_i$, $\mathbf{m}_j$ are significantly distant vectors, taking the covariance matrices $\mathbf{P}_i$, $\mathbf{P}_j$ into account as well. Thus, the Gaussian mixture p is approximately equal to a non-Gaussian mixture that fulfils the assumptions of Lemma 1. Therefore, the following approximation is proposed.

Definition 1. Based on the idea of disjoint supports, the parameters of the approximative mixture π (4) are given by

$$\alpha_i^{DjS}(\omega) = \frac{(a_i)^{\omega} |\mathbf{P}_i|^{\frac{1-\omega}{2}}}{\sum_{j=1}^{N} (a_j)^{\omega} |\mathbf{P}_j|^{\frac{1-\omega}{2}}}, \tag{8}$$

$$\mu_i^{DjS}(\omega) = \mathbf{m}_i, \quad \Pi_i^{DjS}(\omega) = \frac{1}{\omega}\mathbf{P}_i. \tag{9}$$

Remark. Comparing to the approximation proposed in [25], the above approximation takes the component covariance matrices $\mathbf{P}_i$ into account. Both approximations are exact for $\omega = 1$, however, the new one is substantiated better.

B. Approximation based on optimisation

In general, all the parameters of π can be tuned with respect to some criterion. Nevertheless, the number of parameters is quadratic in the dimension n. Further, if the component covariance matrices Π_i are designed by some rule, the number of parameters to be tuned is linear in n. To keep the complexity low, a design of the component means μ_i is also considered. Inspired by (9), the following approximation is proposed.

Definition 2. Based on the idea of minimising the integrated squared error, the parameters of the approximate mixture π (4) are given by

$$\alpha^{ISE}(\omega) = \underset{\alpha:0\leq\alpha_i|_{i=1,\ldots,N}, \sum_{i=1}^{N}\alpha_i=1}{\arg\min} \mathrm{J}^{ISE}(\alpha;\omega), \tag{10}$$

$$\mu_i^{ISE}(\omega) = \mathbf{m}_i, \quad \Pi_i^{ISE}(\omega) = \frac{1}{\omega}\mathbf{P}_i, \tag{11}$$

where the notation $\alpha = [\alpha_1, \ldots, \alpha_N]$ is used and the criterion $\mathrm{J}^{ISE}(\alpha;\omega)$ is given by

$$\mathrm{J}^{ISE}(\alpha;\omega) = \int_{\mathrm{R}^n} (p^{\omega}(\mathbf{x}) - \pi(\mathbf{x};\omega))^2 \,\mathrm{d}\mathbf{x}. \qquad (12)$$

Remark. The component weights α_i are normalised, i.e. the minimisation is constrained by $\sum_{i=1}^{N} \alpha_i = 1$.

Unfortunately, $\mathrm{J}^{ISE}(\alpha;\omega)$ cannot be evaluated analytically. Note that prior the evaluation of (12), the normalising constant $k(\omega)$, which is given by another analytically unsolvable integral (3), has to be computed. Therefore, the following heuristics is proposed.

Definition 3. Based on the sigma point fitting, the parameters of the approximate mixture π (4) are given by

$$\alpha^{SP}(\omega) \propto \underset{\alpha: 0 \le \alpha_i|_{i=1,\ldots,N}}{\arg\min} \; \mathrm{J}^{SP}(\alpha;\omega), \qquad (13)$$

$$\mu_i^{SP}(\omega) = \mathbf{m}_i, \quad \Pi_i^{SP}(\omega) = \frac{1}{\omega}\mathbf{P}_i, \qquad (14)$$

where the symbol $\propto$ means proportional to, i.e. up to a normalising constant, and the criterion $\mathrm{J}^{SP}(\alpha;\omega)$ is given by

$$\mathrm{J}^{SP}(\alpha;\omega) = \sum_{i=1}^{N}\sum_{s=1}^{S_i} w_{i,s}((p(\mathbf{x}_{i,s}))^{\omega} - \pi(\mathbf{x}_{i,s};\omega))^2, \quad (15)$$

with an appropriate choice of the points $\mathbf{x}_{i,s}$, $\dim(\mathbf{x}_{i,s}) = n$, and weights $w_{i,s}$.

Remark. Here, the unnormalised component weights α are optimised, the minimisation is not constrained by equality. The normalisation is performed after the optimisation.

In this paper, the following choice of the weights $w_{i,s}$ and points $\mathbf{x}_{i,s}$ is considered,

$$w_{i,s} = \alpha_i, \qquad (16)$$

$$\mathbf{x}_{i,s}(\omega) = \Big[\mathbf{m}_i \quad \mathbf{m}_i - \kappa\tfrac{1}{\omega^{\frac{1}{2}}}\mathbf{S}_i\mathbf{U}_i \quad \mathbf{m}_i + \kappa\tfrac{1}{\omega^{\frac{1}{2}}}\mathbf{S}_i\mathbf{U}_i\Big]_s, \qquad (17)$$

where $[\ldots]_s$ denotes the s-th column of the corresponding matrix, the matrices $\mathbf{S}_i$ are given by $\mathbf{P}_i = \mathbf{S}_i\mathbf{S}_i^{\mathrm{T}}$, matrices $\mathbf{U}_i$ are unitary, $\mathbf{U}_i\mathbf{U}_i^{\mathrm{T}} = \mathbf{I}$, $\mathbf{I}$ denotes the identity matrix of dimension n, and κ is a parameter, $\kappa > 0$. Specifically, the choice $\mathbf{S}_i = \mathbf{U}_i^{\mathbf{D}}\mathbf{D}_i$ given by $\mathbf{P}_i = \mathbf{U}_i^{\mathbf{D}}\mathbf{D}_i\mathbf{D}_i(\mathbf{U}_i^{\mathbf{D}})^{\mathrm{T}}$, where $\mathbf{D}_i$ is a diagonal matrix, is considered, as well as $\mathbf{U}_i = \mathbf{I}$ and $\kappa = 1$. With respect to the construction of the ω dependent points $\mathbf{x}_{i,s}$ (17), it holds $S_i = N \cdot (2n+1)$.

Remark. The sigma point fitting based approximation was proposed independently of [26], where an approximation of the weighted geometric mean of Gaussian mixtures was proposed, while a limit case of the weighted geometric mean leads to the same approximation of the power of a Gaussian mixture. Nevertheless, the placement of the sigma points needs a further discussion, which can be found in Section IV-A.

C. Approximation based on reduction to Gaussian density

The approximation presented in this section is proposed merely in order to inspect the consequences of the mixture reduction, namely in order to compare the power of a mixture with the power of an approximation of the same mixture.

Lemma 3. *The mean* $\mathbf{m}$ *and the covariance matrix* $\mathbf{P}$ *of the Gaussian mixture* p (1) *are given by*

$$\mathbf{m} = \sum_{i=1}^{N} a_i\mathbf{m}_i, \quad \mathbf{P} = \sum_{i=1}^{N} a_i(\mathbf{P}_i + (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^{\mathrm{T}}). \qquad (18)$$

Proof. The statement is a standard result. □

Definition 4. Based on the moment matching reduction of the Gaussian mixture p to the Gaussian density, the approximative density π is a Gaussian mixture with one component, i.e. a Gaussian density, with parameters given by

$$\alpha^{R2G}(\omega) = 1, \quad \mu^{R2G}(\omega) = \mathbf{m}, \quad \Pi^{R2G}(\omega) = \frac{1}{\omega}\mathbf{P}. \quad (19)$$

Remark. This approximation is very poor, the comparison of the proposed parameters with those corresponding to the other approximations is made in Section IV-C.

IV. Discussion and examples

Section IV-A discusses the design of the sigma points in the optimisation based approach. A bunch of examples is prospected in Section IV-B, with the focus on the component weights $\alpha_i(\omega)$. Means and covariance matrices are inspected in Section IV-C. Section IV-D summarises the examples.

A. Design of the sigma points

First, it must be noted that it is necessary to replace identical components p_i into one component with the weight given by the sum of the corresponding weights a_i. In order to arrive at a unique solution to the weights $\alpha_i(\omega)$, the number of different sigma points have to be greater than or equal to the number of components N as well. The main motivation for the choice $S_i = N \cdot (2n+1)$ is the design simplicity.

From the cross-component perspective, the weights (16) should be proportional to the component weights, at least for ω close to 1, in order to keep the relative importance of subsets of the support with respect to the component merging or splitting. For example, compare the case $\mathbb{a} = [a_1, \ldots, a_N] = [\frac{1}{3}, \frac{2}{3}]$, $\mathbb{m} = [\mathbf{m}_1, \ldots, \mathbf{m}_N] = [0, 10]$, $\mathbf{P}_1 = \mathbf{P}_2 = 1$ with the case $\mathbb{a} = [\frac{1}{3}, \frac{1}{3}, \frac{1}{3}]$, $\mathbb{m} = [0, 9.9, 10.1]$, $\mathbf{P}_1 = \mathbf{P}_2 = \mathbf{P}_3 = 1$.

From the inner-component perspective, the weight $w_{i,1}$ could be different from $w_{i,s}$, $s \neq 1$, in general, but the optimal design has not been prospected. It is important to stress that the approximation is a fitting problem, there is no transformation of random variables, and thus, the use of some techniques is not substantiated. Several experiments were done, but they did not provided any definite answer, the design of $w_{i,1}$, was not too important.

The choice of the rotation/reflection matrix $\mathbf{U}_i$ in the design of the points $\mathbf{x}_{i,s}(\omega)$, as well as of the exact value of the

TABLE I
SETTINGS OF EXAMPLES A–E. ONE DIMENSIONAL $\mathbf{x}$, p HAS TWO COMPONENTS p_i, COMPONENT WEIGHTS GIVEN BY $\mathbb{a} = [1/4, 3/4]$.

#	A	B	C	D	E
$\mathbb{m}$	$[0,0]$	$[0,0]$	$[0,1]$	$[0,5]$	$[0,5]$
$\mathbb{P}$	$[1,4]$	$[4,1]$	$[1,1]$	$[1,1]$	$[1,4]$

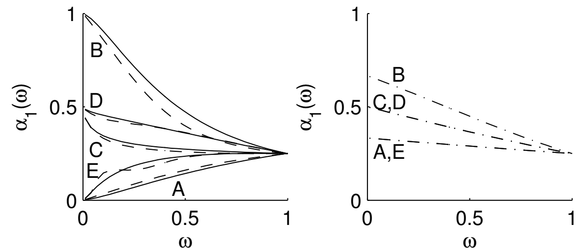

Fig. 1. Component weights $\alpha_1(\omega)$ of the approximate Gaussian mixture $\pi(\mathbf{x};\omega)$: $\alpha_1^{ISE}(\omega)$ (solid lines), $\alpha_1^{SP}(\omega)$ (dashed lines) and $\alpha_1^{DjS}(\omega)$ (dash-dotted lines).

parameter κ, seemed also secondary. Nevertheless, all the design of the sigma points deserves a further attention.

The key factor of the design is the scaling by $\frac{1}{\omega^{\frac{1}{2}}}$ in (17). If all powers p^ω, $0 < \omega < 1$, are needed, the dependence on ω becomes troublesome. This arises in fusion problems, but the scaling performs much better for low powers ω, e.g. $0 < \omega \leq \frac{1}{2}$, which always occur in the fusion. Anyway, the scaling by $\frac{1}{\omega}$ occurs in the components π_i in (14).

B. Comparison of the component weights

In this section, the approximations that are based on the limit case for disjoint supports, on minimising the integrated square error and on the sigma point fitting are compared, see Definitions 1, 2 and 3 respectively. Since the means $\mu_i(\omega)$ and covariance matrices $\Pi_i(\omega)$ are the same in these cases, only the weights $\alpha_i(\omega)$ are compared for the exponents that are needed in fusion problems, i.e. for $0 \leq \omega \leq 1$.

For one dimensional state $\mathbf{x}$, the normalising constant $k(\omega)$ (3) was computed numerically as $(\mathbf{x}_1 - \mathbf{x}_0)\sum_{s=0}^{100}(p(\mathbf{x}_s))^\omega$, where the points $\mathbf{x}_s$ were linearly spaced between $\min\{\mathbf{m}_i - 3\sqrt{\mathbf{P}_i} : i = 1,\ldots,N\}$ and $\max\{\mathbf{m}_i + 3\sqrt{\mathbf{P}_i} : i = 1,\ldots,N\}$. For two dimensional state $\mathbf{x}$, an orthogonal grid was used, with an analogous integration within the individual dimensions. A quadratic programming solver was used for the minimisation in (10), replacing the probability density p^ω by the uniform probability mass function defined in the points $\mathbf{x}_s$ of the above constructed grid. The minimisation in (13) was executed by the same quadratic programming solver.

In the following examples A–X, various aspects of the exponentiation of Gaussian mixtures have been inspected. The approximation based on the integrated square error minimisation is perceived as the benchmark approximation.

Table I determines the basic examples A–E, i.e. the case of one dimensional state $\mathbf{x}$, $n = 1$, and a Gaussian mixture with two components, $N = 2$, while $\mathbb{P}$ denotes $[\mathbf{P}_1, \mathbf{P}_2]$. The component weights are unequal, $a_1 \neq a_2$, while the same and different means and variances are studied.

Fig. 1 shows the exponent ω dependent weights of the first components of the approximate mixtures π. It can be observed that for the limit case $\omega \to 0$, the optimisation based weight $\alpha_1^{ISE}(0)$ is equal to 1 for $\mathbf{P}_1 > \mathbf{P}_2$, equal to 0.5 for $\mathbf{P}_1 = \mathbf{P}_2$ and equal to 0 for $\mathbf{P}_1 < \mathbf{P}_2$, irrespective of the values of the component weights a_1, a_2 or the distance of the means $\mathbf{m}_1$ and $\mathbf{m}_2$, see example B, examples C and D and examples A and E respectively. Having determined the qualitative relation of $\mathbf{P}_1$, and $\mathbf{P}_2$, the values of the parameters $\mathbf{m}_i$ and $\mathbf{P}_i$ govern the progress of $\alpha_i(\omega)$, compare examples A and E and examples C and D.

TABLE II
SETTINGS OF EXAMPLES F–J. TWO DIMENSIONAL $\mathbf{x}$, p HAS TWO COMPONENTS p_i, COMPONENT WEIGHTS GIVEN BY $\mathbb{a} = [1/4, 3/4]$, EQUAL MEANS – $\mathbf{m}_1 = \mathbf{m}_2 = [0,0]^{\mathrm{T}}$.

#	F	G	H	I	J
$\mathrm{diag}(\mathbf{P}_1)$	$[1,1]$	$[2,1]$	$[4,1]$	$[1,4]$	$[1,4]$
$\mathrm{diag}(\mathbf{P}_2)$	$[1,4]$	$[1,4]$	$[1,4]$	$[2,1]$	$[1,1]$

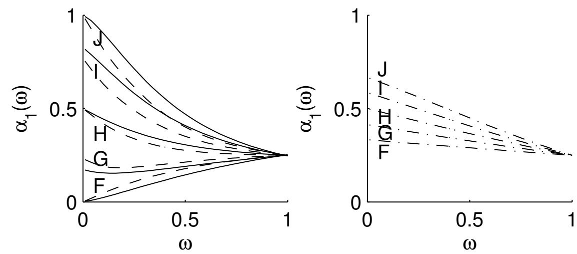

Fig. 2. Component weights $\alpha_1(\omega)$: $\alpha_1^{ISE}(\omega)$ (solid lines), $\alpha_1^{SP}(\omega)$ (dashed lines) and $\alpha_1^{DjS}(\omega)$ (dash-dotted lines).

Further, Fig. 1 compares the weights $\alpha_1^{ISE}(\omega)$ and $\alpha_1^{SP}(\omega)$ and $\alpha_1^{DjS}(\omega)$. Except the example B, the weights $\alpha_1^{SP}(\omega)$ are close to the weights $\alpha_1^{ISE}(\omega)$, although they were computed with lower computation demands. As it can be expected, the approximation based on the limit case works well only for ω close to 1 and for $\mathbf{m}_1$ substantially different from $\mathbf{m}_2$, see $\alpha_1^{DjS}(\omega)$ especially for the examples A and C.

Table II determines examples F–J, which consider two dimensional state $\mathbf{x}$, $n = 2$, and focus on the same component means, $\mathbf{m}_1 = \mathbf{m}_2$. The component weights a_i are again unequal, while diagonal covariance matrices $\mathbf{P}_i$ are studied.

Fig. 2 shows that in the limit case $\omega \to 0$, it holds $\alpha_1^{ISE}(0) = 1$ for $\mathbf{P}_1 \geq \mathbf{P}_2$ and $\alpha_1^{ISE}(0) = 0$ for $\mathbf{P}_1 \leq \mathbf{P}_2$, where both inequalities are meant in the positive-semidefinite sense, see examples J and F respectively. Unlike in the one dimension, the weight $\alpha_1^{ISE}(0)$ can take any value between 0 and 1, see G and I, and it is equal to 0.5 for the symmetric example H. As it can be expected, the weights $\alpha_1^{SP}(\omega)$ have difficulties with following $\alpha_1^{ISE}(\omega)$, if the first component, i.e. the one with the lower weight a_i, has relatively small covariance matrix $\mathbf{P}_i$, see examples I and J and recall B, Fig. 1. Again, the weights $\alpha_1^{DjS}(\omega)$ have a limited applicability.

TABLE III
SETTINGS OF EXAMPLES K–L. TWO DIMENSIONAL $\mathbf{x}$, p HAS TWO COMPONENTS p_i, COMPONENT WEIGHTS GIVEN BY $\mathbb{a} = [1/4, 3/4]$, UNEQUAL MEANS – $\mathbf{m}_1 = [0, 0]^{\mathrm{T}}$, $\mathbf{m}_2 = [5, 0]^{\mathrm{T}}$.

#	K	L	M	N	O	P
$\mathrm{diag}(\mathbf{P}_1)$	$[1, 1]$	$[1, 1]$	$[1, 1]$	$[4, 1]$	$[2, 4]$	$[4, 4]$
$\mathrm{diag}(\mathbf{P}_2)$	$[4, 4]$	$[2, 4]$	$[1, 4]$	$[1, 4]$	$[1, 1]$	$[1, 1]$

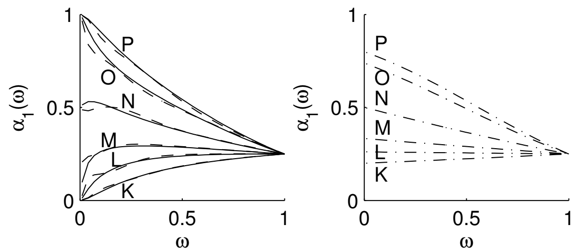

Fig. 3. Component weights $\alpha_1(\omega)$: $\alpha_1^{ISE}(\omega)$ (solid lines), $\alpha_1^{SP}(\omega)$ (dashed lines) and $\alpha_1^{DjS}(\omega)$ (dash-dotted lines).

TABLE IV
SETTINGS OF EXAMPLES Q–S. ONE DIMENSIONAL $\mathbf{x}$, p HAS THREE COMPONENTS p_i, COMPONENT WEIGHTS GIVEN BY $\mathbb{a} = [1/3, 1/3, 1/3]$.

#	Q	R	S
$\mathbb{m}$	$[1, 2, 3]$	$[1, 4, 7]$	$[1, 2, 3]$
$\mathbb{P}$	$[2, 1, 1]$	$[2, 1, 1]$	$[1, 2, 1]$

Table III determines examples K–P, which continue inspecting the case $n = 2$, $N = 2$, $a_1 \neq a_2$, while focusing on unequal component means, $\mathbf{m}_1 \neq \mathbf{m}_2$. Again, only diagonal covariance matrices $\mathbf{P}_i$ are studied.

Fig. 3 shows the same behaviour of $\alpha_1^{ISE}(\omega)$ in the limit case $\omega \to 0$ as earlier, compare $\mathbf{P}_i$ for examples K and L and for examples O and P. Also, a comparison with Fig. 2 can be made, namely the comparison of examples M with F and of examples N with H. Further, it can be observed that for distant means $\mathbf{m}_i$, the weights $\alpha_1^{SP}(\omega)$ follow the weights $\alpha_1^{ISE}(\omega)$ well.

Note that the approximation proposed in [25], i.e. $\alpha_i(\omega) \propto (a_i)^\omega$, can be found in Figs. 1, 2 and 3 as well, since it is equal to $\alpha_i^{DjS}(\omega)$ for examples C/D, H and N respectively. It is evident that the inclusion of the covariance matrices $\mathbf{P}_i$ into the formula improves the results.

The examples Q–X return to one dimensional state $\mathbf{x}$, $n = 1$, but the considered number of components N is risen to 3. The means $\mathbf{m}_i$ are unequal, while small and medium distances of $\mathbf{m}_i$ are considered.

Table IV determines examples Q–S, wherein the component weights a_i are all equal. One of the components have a larger variance than the other two, while the distance of the means $\mathbf{m}_i$ and the placement of the large variance component is studied.

Fig. 4 shows that the large variance component has weight $\alpha_i^{ISE}(\omega)$ that limits to 1 for $\omega \to 0$. It can be concluded that this property is independent of the dimension or the number of components. If the large variance component is at the periphery, the weights $\alpha_i^{SP}(\omega)$ follow the weights $\alpha_i^{ISE}(\omega)$ well, see examples Q and R. The distance of the component means $\mathbf{m}_i$ influences the value of the exponent ω, in which the middle component weight $\alpha_2^{ISE}(\omega)$ attains the zero value. On the other hand, if the large variance component is placed in the middle of the small variance components, see example S, the approximation $\alpha_i^{SP}(\omega)$ does not work correctly. The cause is that with the exponent ω decreasing from 1, the weight α_2^{ISE} has to decrease first, as the central part of the power p^ω becomes covered by the outer widening components $\pi_i(\mathbf{x}; \omega)$, but finally, the weight α_2^{ISE} has to approach 1, as the tails of the power becomes paramount. Also, the omission of the equality constraint from the criterion, compare (10) with (13), manifests itself here, see the zero/one and one half/zero values of the weights $\alpha_1^{SP}(\omega)$ and $\alpha_3^{SP}(\omega)$ versus the weights $\alpha_2^{SP}(\omega)$. Last, note that the weights $\alpha_i^{DjS}(\omega)$ are poor especially in the examples Q and S, since the distance of the means $\mathbf{m}_i$ is small.

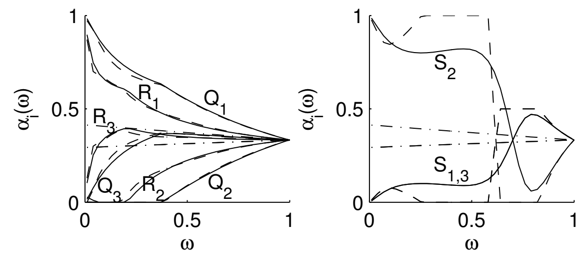

Fig. 4. Component weights $\alpha_i(\omega)$, $i = 1, 2, 3$: $\alpha_i^{ISE}(\omega)$ (solid lines), $\alpha_i^{SP}(\omega)$ (dashed lines) and $\alpha_i^{DjS}(\omega)$ (dash-dotted lines).

TABLE V
SETTINGS OF EXAMPLES T–X. ONE DIMENSIONAL $\mathbf{x}$, p HAS THREE COMPONENTS p_i, COMPONENT VARIANCES GIVEN BY $\mathbb{P} = [1, 1, 1]$.

#	T	U	V	W	X
$\mathbb{m}$	$[1, 2, 3]$	$[1, 4, 7]$	$[1, 4, 7]$	$[1, 4, 7]$	$[1, 4, 7]$
$5 \cdot \mathbb{a}$	$[3, 1, 1]$	$[3, 1, 1]$	$[1, 2, 2]$	$[1, 3, 1]$	$[2, 1, 2]$

Table V determines examples T–X, wherein the component variances $\mathbf{P}_i$ are equal. One of the components has a different weight than the other two, while the distance of the means $\mathbf{m}_i$ and the placement of the different weight component is studied.

Fig. 5 shows that for the equal variance components, the weight $\alpha_2^{ISE}(\omega)$ of the middle component attains the zero value for some value of the exponent ω. This value of ω depends on the values of $\mathbf{m}_i$ and a_i. However, the weights $\alpha_1^{ISE}(0)$, $\alpha_3^{ISE}(0)$ of the outer components are the same in the limit, irrespective of $\mathbf{m}_i$ and a_i. The weights $\alpha_i^{SP}(\omega)$ follow the weights $\alpha_i^{ISE}(\omega)$ well, the weights $\alpha_i^{DjS}(\omega)$ are poor in the example T for all exponents ω, since the distance of the means $\mathbf{m}_i$ is small, and acceptable for ω close to 1 in the other examples.

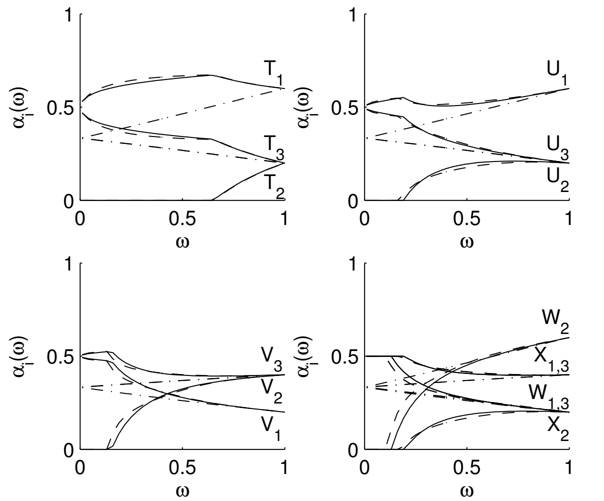

Fig. 5. Component weights $\alpha_i(\omega)$, $i = 1,2,3$: $\alpha_i^{ISE}(\omega)$ (solid lines), $\alpha_i^{SP}(\omega)$ (dashed lines) and $\alpha_i^{DjS}(\omega)$ (dash-dotted lines).

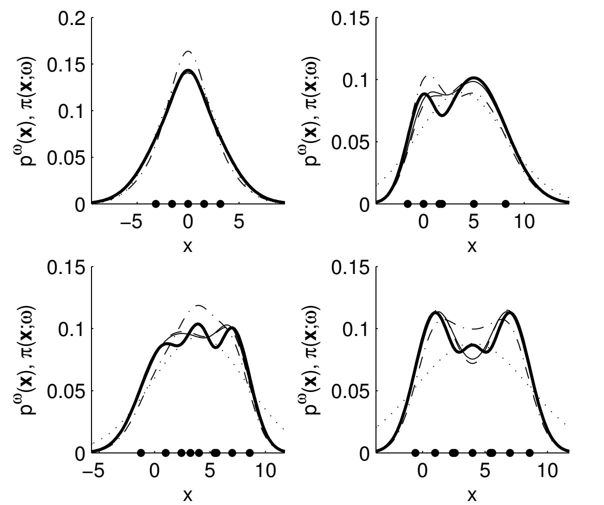

Fig. 6. Power densities $p^\omega(\mathbf{x})$ (thick solid lines) and the approximative Gaussian mixtures $\pi^{ISE}(\mathbf{x};\omega)$ (solid lines), $\pi^{SP}(\mathbf{x};\omega)$ (dashed lines), $\pi^{DjS}(\mathbf{x};\omega)$ (dash-dotted lines), Gaussian density $\pi^{R2G}(\mathbf{x};\omega)$ (dotted lines) and the sigma points $\mathbf{x}_{i,s}$ for the examples A (top left), E (top right), R (bottom left) and X (bottom right).

The comparison of the component weights is closed by the illustration of the power density $p^\omega(\mathbf{x})$ and the approximative densities $\pi(\mathbf{x};\omega)$ for the exponent $\omega = 0.4$.

Fig. 6 illustrates the examples A, E, R and X. It shows that the limit case based approximation $\pi^{DjS}(\mathbf{x};\omega)$ overweights the components with small variances, see examples A and E, and it also overweights the central part of the density, see examples R and X. It is evident that the power of the moment matching based Gaussian density, $\pi^{R2G}(\mathbf{x};\omega)$, is a poor approximation. The sigma points $\mathbf{x}_{i,s}$ that were used for the fitting based approach are illustrated as well. Note that the sigma points can overlap for some pairs (i,s).

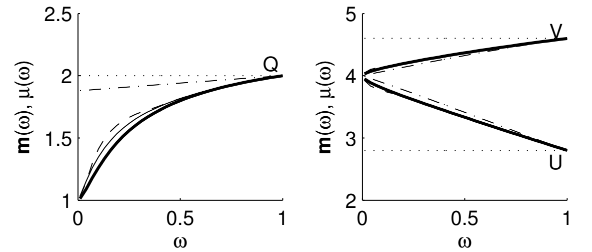

Fig. 7. Density means $\mathbf{m}(\omega)$ (thick solid lines) and $\mu(\omega)$: $\mu^{ISE}(\omega)$ (solid lines), $\mu^{SP}(\omega)$ (dashed lines), $\mu^{DjS}(\omega)$ (dash-dotted lines) and $\mu^{R2G}(\omega)$ (dotted lines) for examples Q (left) and U and V (right).

C. Comparison of the means and covariance matrices

In this section, the three approximation compared in the above section are completed with the reduction based approach, see the approximation given by Definition 4.

The means $\mathbf{m}(\omega)$ and the covariance matrices $\mathbf{P}(\omega)$ that are given by the power densities p^ω, $\mathbf{m}(\omega) = \int_{\mathrm{R}^n} \mathbf{x}\, p^\omega(\mathbf{x})\,\mathrm{d}\mathbf{x}$, $\mathbf{P}(\omega) = \int_{\mathrm{R}^n} (\mathbf{x}-\mathbf{m}(\omega))(\mathbf{x}-\mathbf{m}(\omega))^{\mathrm{T}} p^\omega(\mathbf{x})\,\mathrm{d}\mathbf{x}$, are computed numerically by using the same points $\mathbf{x}_s$ that were used to compute the normalising constant $k(\omega)$, see the beginning of Section IV-B. With respect to the construction of the approximate mixtures $\pi(\mathbf{x};\omega)$, the means $\mu(\omega)$ and the covariance matrices $\Pi(\omega)$ can be computed as

$$\mu(\omega) = \sum_{i=1}^{N} \alpha_i(\omega)\mathbf{m}_i, \tag{20}$$

$$\Pi(\omega) = \sum_{i=1}^{N} \alpha_i(\omega)\Big(\frac{1}{\omega}\mathbf{P}_i + (\mathbf{m}_i - \mu(\omega))(\mathbf{m}_i - \mu(\omega))^{\mathrm{T}}\Big). \tag{21}$$

Fig. 7 draws the exponent ω dependent means related to the power densities and their approximations. It can be observed that the mean $\mathbf{m}(\omega)$ do change with the exponent ω and that the optimisation based means $\mu^{ISE}(\omega)$, $\mu^{SP}(\omega)$ can follow it quite well, as in the considered examples Q, U and V, but the limit case based mean $\mu^{DjS}(\omega)$ can follow it only by chance. Note also that the mean $\mu^{R2G}(\omega)$ is constant by construction. In the example Q, the weight $\alpha_1^{ISE}(\omega)$ corresponding to the large variance component limits to 1 for $\omega \to 0$, and therefore, it holds $\mu^{ISE}(0) = \mu_1^{ISE}(0)$. Since it holds $\alpha_1^{ISE}(0) = \alpha_3^{ISE}(0)$ in both the examples U and V irrespective of a_i, both the means $\mu^{ISE}(0)$ are equal to $\frac{1}{2}(\mathbf{m}_1 + \mathbf{m}_3)$.

To compare the variances, a quantity $\Delta(\omega)$ is introduced,

$$\Delta(\omega) = \frac{\Pi(\omega)}{\frac{1}{\omega}\mathbf{P}} - 1, \tag{22}$$

where $\mathbf{P}$ is the variance of p. That is, $\Delta(\omega)$ is the relative difference from the variance given by $\Pi^{R2G}(\omega)$. In the case of two dimensional state $\mathbf{x}$, the quantities $\Delta_j(\omega) = \frac{\Pi|_{j,j}(\omega)}{\frac{1}{\omega}\mathbf{P}|_{j,j}} - 1$, $j = 1,2$, will be compared, while the notation $|_{j,j}$ refers to the j-th element on the diagonal. For the power density p^ω, the quantity $\Delta^p(\omega)$ will be given analogically by $\frac{\mathbf{P}(\omega)}{\frac{1}{\omega}\mathbf{P}} - 1$.

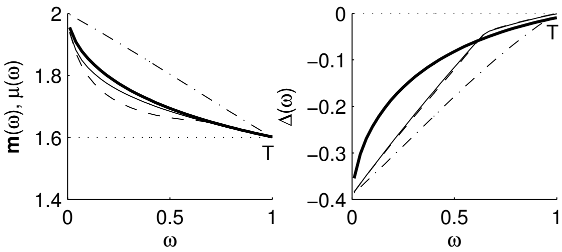

Fig. 8. Density mean $\mathbf{m}(\omega)$ (thick solid line) and $\mu(\omega)$: $\mu^{ISE}(\omega)$ (solid line), $\mu^{SP}(\omega)$ (dashed line), $\mu^{DjS}(\omega)$ (dash-dotted line) and $\mu^{R2G}(\omega)$ (dotted line) for example T (left), corresponding lines for $\Delta(\omega)$ (right).

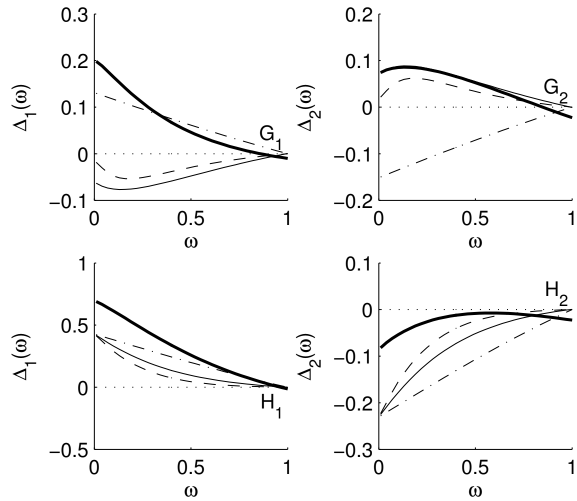

Fig. 9. Relative differences of variances $\Delta_j(\omega)$: $\Delta^p(\omega)$ (thick solid lines), $\Delta^{ISE}(\omega)$ (solid lines), $\Delta^{SP}(\omega)$ (dashed lines), $\Delta^{DjS}(\omega)$ (dash-dotted lines) and $\Delta^{R2G}(\omega)$ (dotted lines). $\Delta_1(\omega)$ (left figures), $\Delta_2(\omega)$ (right figures). Example G (top figures), example H (bottom figures).

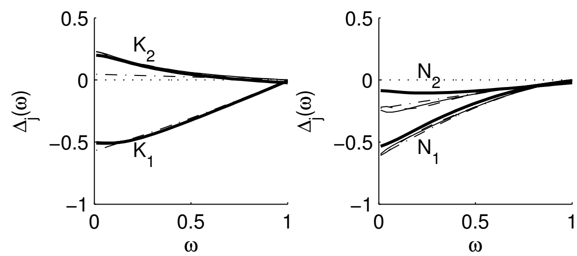

Fig. 10. Relative differences of variances $\Delta_j(\omega)$: $\Delta^p(\omega)$ (thick solid lines), $\Delta^{ISE}(\omega)$ (solid lines), $\Delta^{SP}(\omega)$ (dashed lines), $\Delta^{DjS}(\omega)$ (dash-dotted lines) and $\Delta^{R2G}(\omega)$ (dotted lines). Example K (left), example N (right).

It can be observed that for $\omega \to 0$, the quadratic term $(\mathbf{m}_i - \mu(\omega))(\mathbf{m}_i - \mu(\omega))^{\mathrm{T}}$ can be neglected in (21). In such a case, the covariance matrix $\Pi(\omega)$ is given by the component weights and covariance matrices as $\frac{1}{\omega}\sum_{i=1}^{N}\alpha_i(\omega)\mathbf{P}_i$. That means that $\Pi^{ISE}(\omega)$ can be much larger or smaller than the reduction based covariance matrix $\Pi^{R2G}(\omega)$, see Fig. 6 again.

Fig. 8 illustrates the case when $\Pi^{ISE}(\omega)$ is lower than $\Pi^{R2G}(\omega)$, i.e. when $\Delta^{ISE}(\omega) < 0$. Also, it must be noted that in the example T, the optimisation based variances $\Pi^{ISE}(\omega)$, $\Pi^{SP}(\omega)$ follow the variance $\mathbf{P}(\omega)$ as long as the weight $\alpha_2(\omega)$ of the middle component is nonzero, see also Fig. 3. For $\alpha_2(\omega) = 0$, the other two optimised parameters, $\alpha_1(\omega)$ and $\alpha_3(\omega)$, does not suffice to provide an approximate mixture with an adequate variance.

Fig. 9 illustrates other examples, wherein any approximation is a compromise. Both in G and H, the means $\mathbf{m}_i$ are the same and neither $\mathbf{P}_1 \geq \mathbf{P}_2$ nor $\mathbf{P}_1 \leq \mathbf{P}_2$ holds. It can be observed that no $\Pi(\omega)$ follows the covariance matrix $\mathbf{P}(\omega)$. While comparing $\Delta_j^p(\omega)$ for G and H for $\omega \to 0$, note that despite the qualitative similarity of the examples, the covariance matrices $\mathbf{P}(\omega)$ given by the power density p^ω satisfies $\mathbf{P}(\omega) \geq \frac{1}{\omega}\mathbf{P}$ for G, but not for H, see $\Delta_2^p(\omega)$ and disregard numerical errors for ω close to 1. In other words, the reduction based approach gives qualitatively different results.

Fig. 10 shows another example, K, wherein neither $\mathbf{P}(\omega) \geq \frac{1}{\omega}\mathbf{P}$ nor $\mathbf{P}(\omega) \leq \frac{1}{\omega}\mathbf{P}$ hold. Nevertheless, the covariance matrices $\Pi(\omega)$ corresponding to the optimisation based approximations are close to the power density based covariance matrix $\mathbf{P}(\omega)$. The example N suggests that the ability to follow $\mathbf{P}(\omega)$ can be direction dependent. See that the difference of the means $\mathbf{m}_2 - \mathbf{m}_1$ is 5 in the first dimension of the state, but it is zero in the second one, $\mathbf{m}_2 - \mathbf{m}_1 = [5, 0]^{\mathrm{T}}$, and compare $\Delta_j^{ISE}(\omega)$, $\Delta_j^{SP}(\omega)$ with $\Delta_j^p(\omega)$ for $j = 1$ and for $j = 2$.

D. Lessons learned from the examples

Comparing to the integrated square error optimisation based approximation, the limit case based approximation, that pretends that the components are significantly separated, overweights some components. Especially, it overweights the components with small variances. Among the components with the same variances, it overweights the components within the central part of the density. For multidimensional state, the covariance matrices of the components can be incomparable, in which case it cannot be said which influence will prevail.

If the component means are significantly different, the sigma point fitting based approximation is close to the integrated square error optimisation based approximation. If a large variance component is located between small variance components, the fitting based approach can provide a poor approximation.

For Gaussian mixtures with similar components or for a multidimensional state, even the optimisation based approaches can provide approximative mixtures whose related covariance matrices are significantly different from the ones related to the power density. Also, the quality can vary among the individual dimensions of the state.

To approximate a power of a Gaussian mixture by a Gaussian density, it is inappropriate to approximate the mixture by a Gaussian density first and to compute the power of the Gaussian density in the next step. Therefore, the use of a mixture reduction algorithm before the exponentiation critically influences the results.

V. Summary

It has been proposed to approximate a power of a Gaussian mixture by a Gaussian mixture with the same number of components, while the components are given by scaling the components of the original mixture. From the perspective of the component weights of the approximative mixture, two approximations have been compared against an optimisation based approximation. The approximation based on a limit case have a limited applicability, the specific numerical solution to the optimisation based approximation works well for mixtures with dissimilar components. From the perspective of the means and covariance matrices, the three approximations and the reduction based approximation have been compared against the theoretical values. The reduction based approximation is poor in most cases, while even the optimisation based approximation can lead to such densities that does not provide sufficiently exact covariance matrices.

Acknowledgements

This work was supported by the Czech Science Foundation, project no. P103–13–07058J.

References

[1] D. Simon, *Optimal State Estimation: Kalman, H-infinity, and nonlinear approaches*. Wiley, 2006.

[2] V. Peterka, "Bayesian approach to system identification," in *Trends and Progress in System Identification*, P. Eykhoff, Ed. Pergamon Press, 1981, pp. 239–304.

[3] D. Alspach and H. Sorenson, "Nonlinear Bayesian estimation using Gaussian sum approximations," *IEEE Transactions on Automatic Control*, vol. 17, no. 4, pp. 439–448, August 1972.

[4] M. Šimandl and J. Královec, "Filtering, prediction and smoothing with Gaussian sum representation," in *Proceedings of the 12th IFAC Symposium on System Identification*, Santa Barbara, California, USA, June 2000, pp. 1157–1162.

[5] M. Šimandl, J. Královec, and T. Söderström, "Advanced point-mass method for nonlinear state estimation," *Automatica*, vol. 42, no. 7, pp. 1135–1145, July 2006.

[6] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*, ser. Statistics for Engineering and Information Science. Springer, 2001.

[7] M. West, "Approximating posterior distributions by mixtures," *Journal of the Royal Statistical Society: Series B*, vol. 55, no. 2, pp. 409–422, 1993.

[8] M. Šimandl and O. Straka, "Distance-based pruning for Gaussian sum method in non-Gaussian system state estimation," in *Proceedings of the 8th IASTED International Conference on Intelligent Systems and Control*, Cambridge, USA, October–November 2005, pp. 96–101.

[9] A. Runnalls, "Kullback–Leibler approach to Gaussian mixture reduction," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 3, pp. 989–999, 2007.

[10] D. F. Crouse, P. Willett, K. Pattipati, and L. Svensson, "A look at Gaussian mixture reduction algorithms," in *Proceedings of the 14th International Conference on Information Fusion*, Chicago, Illinois, USA, July 2011, pp. 570–577.

[11] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, ser. Monographs on Statistics and Applied Probability. London: Chapman and Hall, 1986, vol. 26.

[12] G. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, 2nd ed. Wiley, 2008.

[13] C. Genest and J. V. Zidek, "Combining probability distributions: A critique and an annotated bibliography," *Statistical Science*, vol. 1, no. 1, pp. 114–148, 1986.

[14] F. Smarandache and J. Dezert, Eds., *Advances and Applications of DSmT for Information Fusion*. American Research Press, 2004.

[15] ——, *Advances and Applications of DSmT for Information Fusion*. American Research Press, 2006, vol. 2.

[16] R. P. S. Mahler, *Statistical Multisource-Multitarget Information Fusion*. Artech House, 2007.

[17] R. Kulhavý and F. J. Kraus, "On duality of regularized exponential and linear forgetting," *Automatica*, vol. 32, no. 10, pp. 1403–1415, October 1996.

[18] T. Heskes, "Selecting weighting factors in logarithmic opinion pools," in *Advances in Neural Information Processing Systems*. The MIT Press, 1998, pp. 266–272.

[19] S. J. Julier, T. Bailey, and J. K. Uhlmann, "Using exponential mixture models for suboptimal distributed data fusion," in *Proceedings of the 2006 IEEE Nonlinear Statistical Signal Processing Workshop*, Cambridge, UK, September 2006, pp. 160–163.

[20] W. J. Farrell and C. Ganesh, "Generalized Chernoff fusion approximation for practical distributed data fusion," in *Proceedings of the 12th International Conference on Information Fusion*, Seattle, Washington, USA, July 2009.

[21] J. Ajgl and M. Šimandl, "On conservativeness of posterior density fusion," in *Proceedings of the 16th International Conference on Information Fusion*, Istanbul, Turkey, July 2013.

[22] ——, "Design of a robust fusion of probability densities," in *2015 American Control Conference*, Chicago, IL, USA, July 2015.

[23] ——, "Particle based probability density fusion with differential Shannon entropy criterion," in *Proceedings of the 14th International Conference on Information Fusion*, Chicago, IL, USA, July 2011.

[24] ——, "Marginal marginalised particle filter," in *2013 American Control Conference*, Washington, DC, USA, June 2013.

[25] S. J. Julier, "An empirical study into the use of Chernoff information for robust, distributed fusion of Gaussian mixture models," in *Proceedings of the 9th International Conference on Information Fusion*, Florence, Italy, July 2006.

[26] M. Günay, U. Orguner, and M. Demirekler, "Approximate Chernoff fusion of Gaussian mixtures using sigma-points," in *Proceedings of the 17th International Conference on Information Fusion*, Salamanca, Spain, July 2014.