

# On the Reduction of Gaussian inverse Wishart Mixtures

Karl Granström

Division of Automatic Control  
Department of Electrical Engineering  
Linköping University, SE-581 83, Linköping, Sweden  
Email: karl@isy.liu.se

Umut Orguner

Department of Electrical and Electronics Engineering  
Middle East Technical University  
06531, Ankara, Turkey  
Email: umut@eee.metu.edu.tr

**Abstract**—This paper presents an algorithm for reduction of Gaussian inverse Wishart mixtures. Sums of an arbitrary number of mixture components are approximated with single components by analytically minimizing the Kullback-Leibler divergence. The Kullback-Leibler *difference* is used as a criterion for deciding whether or not two components should be merged, and a simple reduction algorithm is given. The reduction algorithm is tested in simulation examples in both one and two dimensions. The results presented in the paper are useful in extended target tracking using the random matrix framework.

**Index Terms**—Gaussian inverse Wishart, mixture reduction, extended target, random matrix, Kullback-Leibler divergence.

## I. INTRODUCTION

In a broad variety of signal processing and sensor fusion problems the state variables are modeled using mixtures. A mixture is a weighted sum of distributions, where the weights are positive. In case the weights sum to one, the mixture is also a distribution. If the weights do not sum to one, the mixture can be called intensity. The individual distributions are called components, a common component choice is the Gaussian distribution, leading to Gaussian mixtures (GM).

In target tracking, GMs are used in e.g. the Multi-hypothesis Tracking (MHT) filter [1], and the Gaussian Mixture PHD-filters [2]–[4]. To keep the complexity at a tractable level, the number of components must be kept at a minimum, leading to the mixture reduction problem. Mixture reduction consists of approximating the original mixture with a reduced mixture, such that the reduced mixture has (considerably) fewer components, while the difference between the two mixtures, defined by some measure, is kept to a minimum.

Several methods for GM reduction have been presented. One solution is pruning, i.e. removing components whose weight is below some threshold (and re-normalizing the weights, if needed). While being very simple, pruning means that the information contained in the pruned components is completely lost. A possibly better choice is to merge components, because merging, to some extent, attempts to preserve some information from each of the merged components. For GM merging, there are top-down algorithms which successively remove components from the original mixture, and there are bottom-up algorithms which successively add components to the reduced

mixture<sup>1</sup>. In terms of the difference measure applied, there are local algorithms which consider only a subset of the available mixture information, and global algorithms that consider all available mixture information.

Examples of GM reduction algorithms include Salmond's (local, top-down) [5], Williams' (global, top-down) [6], Runnalls' (localized version of global measure, top-down) [7], Huber's (global, bottom-up) [8], and Schieferdecker's (global, top-down) [9]. A nice overview of the existing literature is given by Crouse *et al.* [10]. A local top-down approach to reduction of gamma distribution mixtures is presented in [11].

Gaussian inverse Wishart (GIW) densities have recently been introduced as a representation for extended targets [12]. The inverse Wishart distribution is a matrix-variate distribution, which can be used to model the distribution of a Gaussian covariance matrix. For a detailed description of the inverse Wishart distribution, see e.g. [13, Chapter 3]. A multiple extended target tracking framework, under association uncertainty and clutter, would inevitably face an increasing number of GIW mixture components. To the best of our knowledge, reduction of mixtures of GIW distributions has not been studied before.

In this paper, GIW mixture reduction via component merging is addressed. The GIW components are merged by analytically minimizing the Kullback-Leibler divergence (KL-div) [14] between the components and a single GIW distribution. In the presented top-down merging algorithm, a similarity measure based on the KL-div is used, similarly to [7]. However, here it is considered locally, rather than a local approximation of the global measure as in [7]. Note that, when it comes to approximating distributions in a maximum likelihood sense, the KL-div is considered the optimal difference measure [6], [7], [9].

The rest of the paper is organized as follows. Section II defines the problem at hand, and the main result of the paper is derived in Section III. In Section IV a merging criterion is presented, and the merging algorithm is given in Section V. Simulation results are presented in Section VI, and concluding remarks are given in Section VII.

<sup>1</sup>In this case, splitting may be a more appropriate name than merging.

## II. PROBLEM FORMULATION

The random matrix framework for extended target tracking, introduced by Koch [12], decomposes the extended target state  $\xi = (\mathbf{x}, X)$  into a kinematical state  $\mathbf{x} \in \mathbb{R}^{n_x}$  and an extension state  $X \in \mathbb{S}_{++}^d$ , where  $\mathbb{R}^{n_x}$  is the set of real  $n_x$ -vectors,  $\mathbb{S}_{++}^d$  is the set of symmetric positive definite  $d \times d$  matrices, and  $d$  is the dimension of the measurements. In [15], [16] the kinematical and extension state estimate at time step  $k$  is modeled as Gaussian inverse Wishart (GIW) distributed,

$$p(\xi_k) = \mathcal{N}(\mathbf{x}_k; m_{k|k}, P_{k|k}) \mathcal{IW}(X_k; v_{k|k}, V_{k|k}), \quad (1)$$

where  $\mathcal{N}(\cdot)$  denotes a multi-variate Gaussian distribution with mean vector  $m \in \mathbb{R}^{n_x}$  and covariance matrix  $P \in \mathbb{S}_{++}^{n_x}$  (set of symmetric positive semi-definite  $n_x \times n_x$  matrices), and  $\mathcal{IW}(\cdot)$  denotes an inverse Wishart distribution with degrees of freedom  $v > 2d$  and parameter matrix  $V \in \mathbb{S}_{++}^d$ . In this work, the inverse Wishart probability density function (pdf) from [13, Definition 3.4.1] is used<sup>2</sup>.

In multiple extended target tracking under clutter and association uncertainty, the target intensity can be described using a weighted sum of GIW distributions,

$$\begin{aligned} p(\xi_k) &= \sum_{i=1}^{J_{k|k}} w_i \mathcal{N}(\mathbf{x}_k; m_{k|k}^{(i)}, P_{k|k}^{(i)}) \mathcal{IW}(X_k; v_{k|k}^{(i)}, V_{k|k}^{(i)}) \\ &= \sum_{i=1}^N w_i p_i(\xi_k), \end{aligned} \quad (2)$$

where each distribution  $p_i(\cdot)$  is referred to as a GIW component. Note that in some target tracking frameworks<sup>3</sup> the weights do not necessarily sum to unity, and therefore  $p(\cdot)$  might not be a probability density. As time progresses, the number of GIW components grows larger, and approximations become necessary to keep  $J_{k|k}$  at a computationally tractable level. One such approximation, called pruning, is to discard components with weights  $w_i$  lower than some truncation threshold  $T$ . In this work, we explore merging of GIW components, i.e. approximating sums of components with just one component. The result of merging a sum of GIW components (2) is a sum

$$\begin{aligned} \tilde{p}(\xi_k) &= \sum_{i=1}^{\tilde{J}_{k|k}} \tilde{w}_i \mathcal{N}(\mathbf{x}_k; \tilde{m}_{k|k}^{(i)}, \tilde{P}_{k|k}^{(i)}) \mathcal{IW}(X_k; \tilde{v}_{k|k}^{(i)}, \tilde{V}_{k|k}^{(i)}) \\ &= \sum_{i=1}^{\tilde{J}_{k|k}} \tilde{w}_i \tilde{p}_i(\xi_k), \end{aligned} \quad (3)$$

where  $\tilde{J}_{k|k} < J_{k|k}$ .

Our approach to GIW mixture reduction takes the following steps. First we give a theorem which is used to find the GIW distribution  $q(\cdot)$  which minimizes the Kullback-Leibler divergence between  $\tilde{w}q(\cdot)$  and the sum  $p = \sum_{i \in L} w_i p_i$ , where  $\tilde{w} = \sum_{i \in L} w_i$  and  $L \subseteq \{1, \dots, J_{k|k}\}$ . Next we give a criterion

<sup>2</sup>The definition is also given in (34b).

<sup>3</sup>For instance PHD and CPHD filters, see e.g. [2]–[4], [17]–[19].

which is used to determine if two GIW components  $p_i(\cdot)$  and  $p_j(\cdot)$  should be merged or not, and we then give an algorithm which, given a threshold  $U$  for the merging criterion, reduces the number of GIW components in the mixture.

## III. APPROXIMATING A WEIGHTED SUM OF GIW-COMPONENTS WITH ONE GIW-COMPONENT

This section contains the main result of the paper – a theorem that describes how a sum of an arbitrary number of GIW components can be merged into just one GIW component. This is performed via analytical minimization of the KL-div,

$$\text{KL}(p||q) = \int p(x) \log \left( \frac{p(x)}{q(x)} \right) dx, \quad (4)$$

a measure of how similar two functions  $p$  and  $q$  are. The KL-div is well-known in the literature for its moment-matching characteristics, see e.g. [20], [21], and as mentioned above it is considered the optimal difference measure in a maximum likelihood sense [6], [7], [9]. Note that minimizing the KL-div between  $p$  and  $q$  w.r.t.  $q$  can be rewritten as a maximization problem,

$$\min_q \text{KL}(p||q) = \max_q \int p(x) \log(q(x)) dx. \quad (5)$$

*Theorem 1:* Let  $p(\cdot)$  be a weighted sum of GIW components,

$$\begin{aligned} p(\mathbf{x}, X) &= \sum_{i=1}^N w_i \mathcal{N}(\mathbf{x}; m_i, P_i) \mathcal{IW}(X; v_i, V_i) \\ &= \sum_{i=1}^N w_i p_i(\mathbf{x}, X), \end{aligned} \quad (6)$$

where  $\bar{w} = \sum_{i=1}^N w_i$ . Let

$$q(\mathbf{x}, X) = \bar{w} \mathcal{N}(\mathbf{x}; m, P) \mathcal{IW}(X; v, V) \quad (7)$$

be the minimizer of the KL-div between  $p(\mathbf{x}, X)$  and  $q(\mathbf{x}, X)$  among all GIW distributions, i.e.

$$q(\mathbf{x}, X) \triangleq \arg \min_{q(\mathbf{x}, X) \in \text{GIW}} \text{KL}(p(\mathbf{x}, X) || q(\mathbf{x}, X)). \quad (8)$$

Then the parameters  $m$ ,  $P$ , and  $V$  are given by

$$m = \frac{1}{\bar{w}} \sum_{i=1}^N w_i m_i, \quad (9a)$$

$$P = \frac{1}{\bar{w}} \sum_{i=1}^N w_i (P_i + (m_i - m)(m_i - m)^T), \quad (9b)$$

$$V = \bar{w} (v - d - 1) \left( \sum_{i=1}^N w_i (v_i - d - 1) V_i^{-1} \right)^{-1}, \quad (9c)$$

and  $v$  is the solution to the equation

$$0 = \bar{w}d \log(v - d - 1) - \bar{w} \sum_{j=1}^d \psi_0\left(\frac{v - d - j}{2}\right) + \bar{w}d \log \bar{w} - \bar{w} \log \left| \sum_{i=1}^N w_i (v_i - d - 1) V_i^{-1} \right| + \sum_{i=1}^N \sum_{j=1}^d w_i \psi_0\left(\frac{v_i - d - j}{2}\right) - \sum_{i=1}^N w_i \log |V_i|, \quad (9d)$$

where  $|V|$  is the determinant of  $V$  and  $\psi_0(\cdot)$  is the digamma function (a.k.a. the polygamma function of order 0). ■

*Proof:* Given in Appendix A.

*Remarks:* The expressions for  $m$  in (9a) and  $P$  in (9b) are well known, see e.g. the textbook [22], and have been used earlier to merge Gaussians in a target tracking context, see e.g. [2]–[7], [9], [10]. To the best of the authors' knowledge the identities for the calculation of the parameters  $V$  and  $v$  have not been published before. The expressions for  $V$  and  $v$  in (9c) and (9d) correspond to matching the expected values of  $X^{-1}$  and  $\log |X|$  under both densities,

$$\bar{w} E_q [X^{-1}] = \sum_{i=1}^N w_i E_{p_i} [X^{-1}], \quad (10a)$$

$$\bar{w} E_q [\log |X|] = \sum_{i=1}^N w_i E_{p_i} [\log |X|]. \quad (10b)$$

There is a unique solution to (9d), and a value for the parameter  $v$  is easily obtained by applying a numerical root finding algorithm to (9d), e.g. Newton's algorithm, see e.g. [23].

#### IV. MERGING CRITERION

In this section we derive a criterion that is used to determine whether or not two GIW components should be merged. When reducing the number of components, it is preferred to preserve the overall modality of the mixture. Thus, if the initial mixture  $p(\mathbf{x}, X)$  has  $M$  modes, then the reduced mixture  $\tilde{p}(\mathbf{x}, X)$  should have  $M$  modes.

The optimal solution to this problem is to consider every possible way to reduce  $J_{k|k}$  components, compute the corresponding KL-div:s, and then find the best trade-off between low KL-div and reduction of  $J_{k|k}$ . For  $J_{k|k}$  components, there are  $B_{J_{k|k}}$  different ways to merge, where  $B_i$  is the  $i$ :th Bell number [24]. Because  $B_i$  increases rapidly with  $i$ , e.g.  $B_5 = 52$  and  $B_{10} = 115975$ , the optimal solution can not be used in practice.

Instead a merging criterion must be used to determine whether or not a pair of GIW components should be merged. In what follows we present a distance measure that can be thresholded to compare two GIW components, and we also elaborate on the Gaussian and inverse Wishart parts of this distance measure.

#### A. Distance measure

As distance measure the KL-div could be used, however because it is asymmetrical,  $\text{KL}(p||q) \neq \text{KL}(q||p)$ , it should not be used directly. Instead we use the Kullback-Leibler difference (KL-diff), defined for two distributions  $p(\mathbf{x}, X)$  and  $q(\mathbf{x}, X)$  as

$$\begin{aligned} D_{\text{KL}}(p(\mathbf{x}, X), q(\mathbf{x}, X)) &= \text{KL}(p(\mathbf{x}, X) || q(\mathbf{x}, X)) + \text{KL}(q(\mathbf{x}, X) || p(\mathbf{x}, X)) \\ &= \iint p(\mathbf{x}, X) \log \left( \frac{p(\mathbf{x}, X)}{q(\mathbf{x}, X)} \right) d\mathbf{x} dX \\ &\quad + \iint q(\mathbf{x}, X) \log \left( \frac{q(\mathbf{x}, X)}{p(\mathbf{x}, X)} \right) d\mathbf{x} dX. \end{aligned} \quad (11)$$

Let  $p(\mathbf{x}, X)$  and  $q(\mathbf{x}, X)$  be defined as

$$p(\mathbf{x}, X) = \mathcal{N}(\mathbf{x}; m_1, P_1) \mathcal{IW}(X; v_1, V_1), \quad (12a)$$

$$q(\mathbf{x}, X) = \mathcal{N}(\mathbf{x}; m_2, P_2) \mathcal{IW}(X; v_2, V_2). \quad (12b)$$

The KL-div between  $p(\cdot)$  and  $q(\cdot)$  is

$$\begin{aligned} \text{KL}(p(\mathbf{x}, X) || q(\mathbf{x}, X)) &= \int \mathcal{N}(\mathbf{x}; m_1, P_1) \log \left( \frac{\mathcal{N}(\mathbf{x}; m_1, P_1)}{\mathcal{N}(\mathbf{x}; m_2, P_2)} \right) d\mathbf{x} \\ &\quad + \int \mathcal{IW}(X; v_1, V_1) \log \left( \frac{\mathcal{IW}(X; v_1, V_1)}{\mathcal{IW}(X; v_2, V_2)} \right) dX \\ &= \text{KL}(\mathcal{N}(\mathbf{x}; m_1, P_1) || \mathcal{N}(\mathbf{x}; m_2, P_2)) \\ &\quad + \text{KL}(\mathcal{IW}(X; v_1, V_1) || \mathcal{IW}(X; v_2, V_2)), \end{aligned} \quad (13)$$

where

$$\begin{aligned} \text{KL}(\mathcal{N}(\mathbf{x}; m_1, P_1) || \mathcal{N}(\mathbf{x}; m_2, P_2)) &= \frac{1}{2} \left[ \log |P_2| - \log |P_1| - n_x + \text{Tr}(P_2^{-1} P_1) \right. \\ &\quad \left. + (m_1 - m_2)^T P_2^{-1} (m_1 - m_2) \right], \end{aligned} \quad (14)$$

and

$$\begin{aligned} \text{KL}(\mathcal{IW}(X; v_1, V_1) || \mathcal{IW}(X; v_2, V_2)) &= \frac{v_1 - d - 1}{2} \log |V_1| - \frac{v_2 - d - 1}{2} \log |V_2| \\ &\quad + \sum_{j=1}^d \left( \log \Gamma \left( \frac{v_2 - d - j}{2} \right) - \log \Gamma \left( \frac{v_1 - d - j}{2} \right) \right) \\ &\quad + \frac{v_2 - v_1}{2} \left( \log |V_1| - \sum_{j=1}^d \psi_0 \left( \frac{v_1 - d - 1}{2} \right) \right) \\ &\quad + \text{Tr} \left( -\frac{1}{2} (v_1 - d - 1) V_1^{-1} (V_1 - V_2) \right). \end{aligned} \quad (15)$$

Showing (14) and (15) is straightforward, the tedious details are omitted. The KL-div between  $q(\cdot)$  and  $p(\cdot)$  is defined analogously.

Note that the decomposition of  $\text{KL}(p(\cdot) || q(\cdot))$  into a sum (13) is inherited from the separability of the Gaussian and

inverse Wishart distributions in (12). From (13) it follows that the KL-diff is separable,

$$\begin{aligned} D_{\text{KL}}(p(\mathbf{x}, X), q(\mathbf{x}, X)) &= D_{\text{KL}}^{\mathcal{N}} + D_{\text{KL}}^{\mathcal{IW}} \\ &= D_{\text{KL}}(\mathcal{N}(\mathbf{x}; m_1, P_1), \mathcal{N}(\mathbf{x}; m_2, P_2)) \\ &\quad + D_{\text{KL}}(\mathcal{IW}(X; v_1, V_1), \mathcal{IW}(X; v_2, V_2)), \end{aligned} \quad (16)$$

where

$$\begin{aligned} D_{\text{KL}}(\mathcal{N}(\mathbf{x}; m_1, P_1), \mathcal{N}(\mathbf{x}; m_2, P_2)) \\ = \frac{1}{2} (m_1 - m_2)^\top (P_1^{-1} + P_2^{-1}) (m_1 - m_2) \\ - n_x + \frac{1}{2} \text{Tr}(P_2^{-1} P_1 + P_1^{-1} P_2), \end{aligned} \quad (17)$$

and

$$\begin{aligned} D_{\text{KL}}(\mathcal{IW}(X; v_1, V_1), \mathcal{IW}(X; v_2, V_2)) \\ = \frac{1}{2} \text{Tr} \left( \left[ (v_1 - d - 1) V_1^{-1} - (v_2 - d - 1) V_2^{-1} \right] (V_2 - V_1) \right) \\ + \frac{v_2 - v_1}{2} \left( \log |V_1| - \sum_{j=1}^d \psi_0 \left( \frac{v_1 - d - j}{2} \right) \right. \\ \left. - \log |V_2| + \sum_{j=1}^d \psi_0 \left( \frac{v_2 - d - j}{2} \right) \right). \end{aligned} \quad (18)$$

Note that the Gaussian KL-diff (17) has similarities to the merging criterion

$$(m_i - m_j)^\top P_i^{-1} (m_i - m_j), \quad w_i > w_j, \quad (19)$$

which is used to merge sums of Gaussians in e.g. [2], [3], [5].

Thresholding the KL-diff

$$D_{\text{KL}}(p(\mathbf{x}, X), q(\mathbf{x}, X)) < U \quad (20)$$

is a straightforward way to determine whether or not two Gaussian inverse Wishart distributions should be merged. Alternatively, the Gaussian and inverse Wishart KL-diff:s can be thresholded separately,

$$(D_{\text{KL}}^{\mathcal{N}} < U_{\mathcal{N}}) \& (D_{\text{KL}}^{\mathcal{IW}} < U_{\mathcal{IW}}), \quad (21)$$

where  $\&$  is the logical and operator. In the following two subsections we will elaborate on the Gaussian and inverse Wishart KL-diff:s to gain a better understanding of how the merging criterion works.

### B. A closer look at the Gaussian KL-diff

Under the assumption that  $P_2 = \alpha P_1$ ,  $\alpha > 0$ , and  $m_2 = m_1 + P_1^{1/2} m_e$ ,  $P_1^{1/2} P_1^{1/2} = P_1$ , the KL-diff is independent of the specific values of  $m_1$  and  $P_1$ ,

$$D_{\text{KL}}^{\mathcal{N}} = -n_x + \frac{1 + \frac{1}{\alpha} m_e^\top m_e}{2} + \frac{\alpha + \frac{1}{\alpha} n_x}{2}. \quad (22)$$

If  $m_e = \mathbf{0}$  the KL-diff is  $D_{\text{KL}}^{\mathcal{N}} = \frac{1}{2} (\alpha + \frac{1}{\alpha}) n_x$ . With a threshold  $U_{\mathcal{N}}$ ,  $D_{\text{KL}}^{\mathcal{N}} < U_{\mathcal{N}}$  is equivalent to  $\alpha_1 < \alpha < \alpha_2$ , where

$$\alpha_i = 1 + \frac{U_{\mathcal{N}}}{n_x} + (-1)^i \sqrt{\left(1 + \frac{U_{\mathcal{N}}}{n_x}\right)^2 - 1}. \quad (23)$$

Thus, the upper and lower limit of  $\alpha$  is dependent on both the threshold, and on the dimension of the kinematical state  $n_x$ . For a given threshold  $U_{\mathcal{N}}$ , a larger  $n_x$  means that  $\alpha$  must be closer to 1 for  $D_{\text{KL}}^{\mathcal{N}} < U_{\mathcal{N}}$  to be fulfilled.

If  $\alpha = 1$  the KL-diff is  $D_{\text{KL}}^{\mathcal{N}} = m_e^\top m_e$ , i.e. the length of  $m_e$  squared. For a given threshold  $U_{\mathcal{N}}$  the difference between  $m_1$  and  $m_2$  can at most be  $\sqrt{U_{\mathcal{N}}}$  standard deviations. Thus, given  $\alpha = 1$ , the KL-diff can be defined in terms of the standard deviation  $P_1^{1/2}$ , and is independent of the size of the kinematical state  $\mathbf{x}$ .

### C. A closer look at the inverse Wishart KL-diff

Under the assumption that  $V_2 = \beta V_1$ , the KL-diff becomes independent of the specific value of  $V_1$ . If  $v_2 = v_1$  the KL-diff is

$$D_{\text{KL}}^{\mathcal{IW}} = \frac{(v_1 - d - 1)d(\beta - 1)^2}{2\beta}. \quad (24)$$

With a threshold  $U_{\mathcal{IW}}$ ,  $D_{\text{KL}}^{\mathcal{IW}} < U_{\mathcal{IW}}$  is equivalent to  $\beta_1 < \beta < \beta_2$  where

$$\beta_i = 1 + \frac{U_{\mathcal{IW}}}{(v_1 - d - 1)d} + (-1)^i \sqrt{\left(1 + \frac{U_{\mathcal{IW}}}{(v_1 - d - 1)d}\right)^2 - 1}. \quad (25)$$

The upper and lower limit of  $\beta$  is dependent on the threshold  $U_{\mathcal{IW}}$ , the dimension of the measurements  $d$ , and on the inverse Wishart degrees of freedom  $v_1$ . A higher threshold gives larger  $\beta_2$  and smaller  $\beta_1$ , while a higher  $d$  and/or  $v_1$  forces both limits closer to one.

Unfortunately there is no obvious way to choose  $v_2$  as a function of  $v_1$  to make the KL-diff independent of the specific value of  $v_1$ , making it difficult to make a similar examination of how the inverse Wishart degrees of freedom affect the KL-diff.

### D. Discussion

The subsections above give some intuition as to how  $U$  (or  $U_{\mathcal{N}}$  and  $U_{\mathcal{IW}}$ ) affects the merging criterion, however it is difficult to give specific hints for choosing a numerical value of  $U$ . Such a value is likely best determined empirically. In the results section below we will examine all four GIW parameters, and how they affect the KL-diff, in numerical examples.

## V. MERGING ALGORITHM

In this section we present a merging algorithm that uses the merging method and criterion defined above, see Table I. In the algorithm a choice is made regarding how aggressively the components are bundled for merging, i.e. how aggressively  $J_{k|k}$  is reduced. There are many possible ways to do this, two are given in Table I. Both alternatives start by picking out the GIW component with highest weight, say the  $j$ :th. The first alternative,  $L_1$  in Table I, then merges component  $j$  with all other components  $i$  for which it holds

$$D_{\text{KL}}(p_j(\mathbf{x}, X), p_i(\mathbf{x}, X)) < U. \quad (26)$$

TABLE I  
GAUSSIAN INVERSE WISHART REDUCTION

---

1: **require:**  $p(\mathbf{x}_k, X_k)$  as in (2), a merging threshold  $U$ , and  $\theta \in \{1, 2\}$ .  
2: **initialize:** Set  $\ell = 0$  and  $I = \{1, \dots, J_{k|k}\}$ .  
3: **repeat**  
4:   Set  $\ell = \ell + 1$  and  $j = \arg \max_{i \in I} w_{k|k}^{(i)}$   
5:   Set  $L = L_\theta$ , where  $L_1 = \{i \in I \mid D_j^i < U\}$ ,  

$$L_2 = \left\{ i \in I \mid \exists \{i_1 = i, \dots, i_N = j\} \right.$$

$$\left. \ni D_{i_k}^{i_{k+1}} < U, k = 1, \dots, N - 1 \right\},$$
and  $D_j^i = D_{\text{KL}}(p_j(\mathbf{x}, X), p_i(\mathbf{x}, X))$ .  
6:   Use Theorem 1 to compute  

$$\tilde{w}_{k|k}^{(\ell)}, \tilde{m}_{k|k}^{(\ell)}, \tilde{P}_{k|k}^{(\ell)}, \tilde{v}_{k|k}^{(\ell)}, \tilde{V}_{k|k}^{(\ell)} \quad (28)$$
for the components  $i \in L$ .  
7:    $I = I \setminus L$   
8: **until**  $I = \emptyset$   
9: **output:**  $\tilde{p}(\mathbf{x}_k, X_k) = \sum_{i=1}^{\tilde{J}_{k|k}} \tilde{w}_i \mathcal{N}(\mathbf{x}_k; \tilde{m}_i, \tilde{P}_i) \mathcal{IW}(X_k; \tilde{v}_i, \tilde{V}_i)$ ,  
where the number of components is  $J_{k|k} = \ell$ .

---

The second alternative,  $L_2$  in Table I, finds all other components such that for each component  $i \in L_2$ , there exists a sequence of indices  $\{i_1 = i, \dots, i_N = j\}$  such that

$$D_{\text{KL}}(p_{i_k}(\mathbf{x}, X), p_{i_{k+1}}(\mathbf{x}, X)) < U, k = 1, \dots, N - 1. \quad (27)$$

$L_1$  is a special case of  $L_2$ , where  $\{i_1 = i, i_2 = j\}$ , and it immediately follows that  $|L_1| \leq |L_2|$ , where  $|L|$  is the cardinality of the set  $L$ . Thus  $L_2$  merges more components than  $L_1$ , resulting in a higher reduction of  $J_{k|k}$ , but also a cruder approximation of  $p(\mathbf{x}, X)$ .

## VI. SIMULATION RESULTS

This section presents results from numerical simulations. Simulations of the Gaussian and inverse Wishart parts of the KL-diff are presented in Section VI-A, and merging of GIW components in  $n_x = d = 1$  and  $n_x = d = 2$  dimensions are presented in Sections VI-B and VI-C. In Section VI-D we compare the two merging choices  $L_1$  and  $L_2$  in  $n_x = d = 1$  dimension.

### A. Merging criterion

This section presents results that evaluate the merging criterion in Section IV. Let  $p_1(\mathbf{x}, X)$  and  $p_2(\mathbf{x}, X)$  be defined as

$$p_1(\mathbf{x}, X) = \mathcal{N}(\mathbf{x}; m_1, P_1) \mathcal{IW}(X; v_1, V_1), \quad (29a)$$

$$p_2(\mathbf{x}, X) = \mathcal{N}(\mathbf{x}; m_2, P_2) \mathcal{IW}(X; v_2, V_2). \quad (29b)$$

The evaluation is performed *ceteris paribus*, i.e. by changing the parameters of the Gaussian while holding the parameters of the inverse Wishart equal, and vice versa.

1) *Different Gaussian parameters:* Let  $P_2 = \alpha P_1$ , and  $m_2 = m_1 + P_1^{1/2} m_e$ . A contour plot of the KL-diff for two uni-variate Gaussians ( $n_x = 1$ ) is shown in Figure 1a. In accordance with the discussion in Section IV, the KL-diff

increases with the length of  $m_e$ , and it increases when  $\alpha < 1$  or  $\alpha > 1$ .

2) *Different inverse Wishart parameters:* Let  $V_2 = \beta V_1$  to make the KL-diff independent of the specific value of  $V_1$ . For a given  $\beta$ , setting  $v_2 = 2d + 2 + \beta(v_1 - 2d - 2)$  will give correct expected value of  $X$ . We make changes to this value by multiplying with a factor  $\eta$ , i.e.  $v_2 = \eta(2d + 2 + \beta(v_1 - 2d - 2))$ . A contour plot of the KL-diff for one dimensional inverse Wisharts is shown in Figure 1b, in this figure  $v_1 = 20$ . The contours  $D_{\text{KL}} = 3$  are shown for  $v_1 = 20, 40, 60, 80, 100$  in Figure 1c, where it shows how the area enclosed by  $D_{\text{KL}} = 3$  decreases when  $v_1$  increases.

### B. Merging of one dimensional components

An intensity  $p(\mathbf{x}, X)$  with four GIW components,  $n_x = d = 1$ , was reduced to two components using a KL-diff threshold of  $U = 3$ . The GIW components and sums are shown before and after merging in Figure 2.

### C. Merging of two dimensional components

An intensity  $p(\mathbf{x}, X)$  with two GIW components,  $n_x = d = 2$ , was reduced to one component using a KL-diff threshold of  $U = 12$ . The GIW components are shown before and after merging in Figure 3.

### D. Comparison of merging algorithms

An intensity  $p(\mathbf{x}, X)$  with 50 GIW components,  $n_x = d = 1$ , was reduced using both  $L_1$  and  $L_2$  in Table I. The GIW mixture parameters were sampled uniformly from the following intervals,

$$w_i \in [0.05 \ 0.95], m_i \in [0 \ 10], P_i \in [0.25^2 \ 0.75^2], \quad (30)$$

$$v_i \in [50 \ 250], \frac{V_i}{v_i - 2d - 2} \in [15 \ 50], \quad (31)$$

i.e.  $V_i$  was sampled such that, given a sampled  $v_i$ , the expected value of  $X$  belongs to  $[15 \ 50]$ . The original mixture, and the two approximations, are shown in Figure 4. Using  $L_1$  the reduced mixture has 29 components, using  $L_2$  gives only 23 components, but also a cruder approximation.

## VII. CONCLUDING REMARKS

This paper presented a reduction algorithm for mixtures of Gaussian inverse Wishart distributions. A theorem was given, which is used to reduce an arbitrary number of GIW components to just one component by analytically minimizing the Kullback-Leibler divergence, in a maximum likelihood sense the optimal difference measure. Using the Kullback-Leibler difference, a merging criterion for pairs of GIW components was given. The criterion has the benefit of decomposing easily into separate criteria for the Gaussians and inverse Wisharts, respectively. A simple algorithm for GIW mixture reduction was also given, and tested in simulation examples in both one and two dimensions.

The outlook on future work includes considering a global difference measure between the original and reduced mixture, instead of just a local measure. The reduction algorithm will

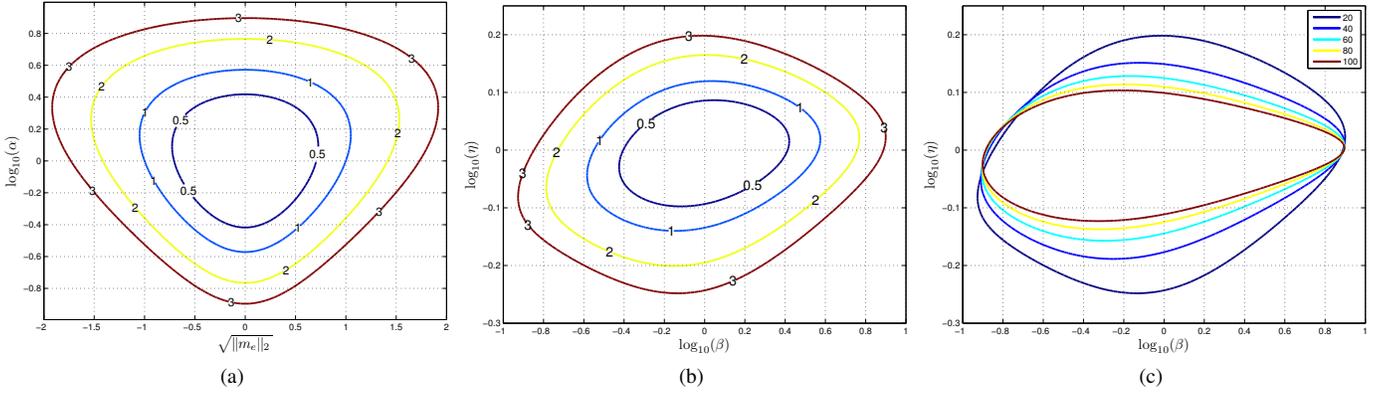


Fig. 1. Contour plots showing the KL-diff when the Gaussian inverse Wishart parameters are changed. (a) KL-diff for two uni-variate Gaussian distributions. (b) KL-diff for two one dimensional inverse Wishart distributions, here  $v_1 = 20$ . (c) KL-diff for pairs of one dimensional inverse Wishart distributions, the outlines show  $D_{KL} = 3$  and the legend shows the value of  $v_1$ .

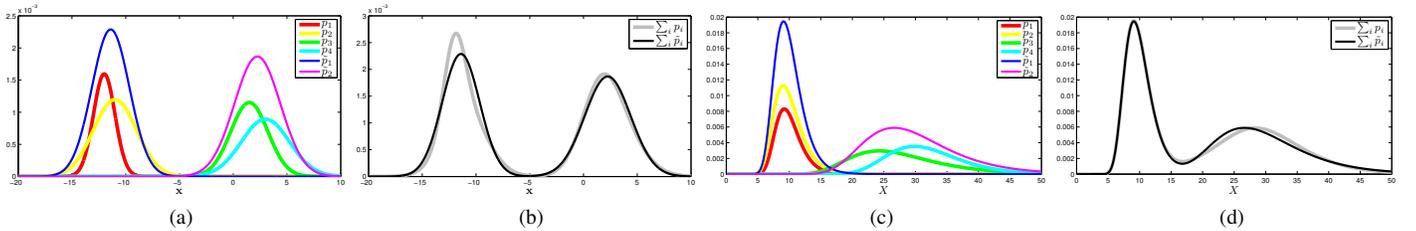


Fig. 2. Four GIW components,  $n_x = d = 1$ , merged into two components using a threshold  $U = 3$ . (a) shows the Gaussian parts of the components before and after merging, and (b) shows the sums of the Gaussians before and after merging. (c) shows the inverse Wishart parts of the components before and after merging, and (d) shows the sums of the inverse Wisharts before and after merging.

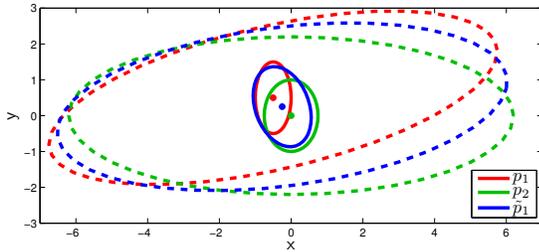


Fig. 3. Two GIW components,  $n_x = d = 2$ , merged into one component using a threshold  $U = 12$ . Shown are the kinematical state means  $m$  (dots), the corresponding covariances  $P$  (solid ellipses), and the estimated expected values of the extension state  $X$  (dashed ellipses).

be used in the Gaussian inverse Wishart PHD-filter for multiple extended target tracking under association uncertainty and clutter [19].

## APPENDIX A PROOF OF THEOREM 1

### A. Expected value of inverse extension

Let  $X$  be inverse Wishart distributed  $\mathcal{IW}(X; v, V)$ . Then  $X^{-1}$  is Wishart distributed  $\mathcal{W}(X^{-1}; v - d - 1, V^{-1})$  [13, Theorem 3.4.1]. The expected value of  $X^{-1}$  is [13, Theorem 3.3.15]

$$\mathbb{E}[X^{-1}] = (v - d - 1)V^{-1}. \quad (32)$$

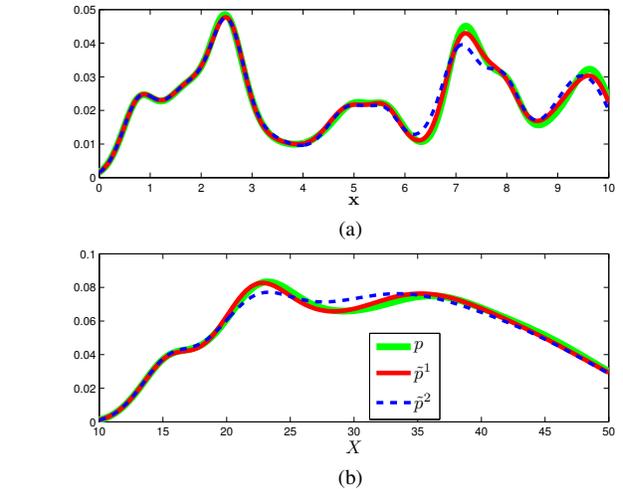


Fig. 4. Merging of 50 GIW components. (a) shows the Gaussian sum, before merging (green) as well as after merging using  $L_1$  (red) and  $L_2$  (blue). (b) shows the corresponding inverse Wishart sum. Using  $L_1$  results in 28 GIW components, using  $L_2$  results in 21 components, but also a cruder approximation.

### B. Expected value of log determinant of extension

Let  $y$  be a uni-variate random variable. The moment generating function for  $y$  is defined as  $\mu_y(s) \triangleq \mathbb{E}_y[e^{sy}]$ , and the

expected value of  $y$  is given in terms of  $\mu_y(s)$  as

$$\mathbb{E}[y] = \left. \frac{d\mu_y(s)}{ds} \right|_{s=0}. \quad (33)$$

Let  $y = \log |X|$ , where  $X \sim \mathcal{IW}(X; v, V)$ . The moment generating function of  $y$  is

$$\mu_y(s) = \mathbb{E}[|X|^s] = \int |X|^s p(X) dX \quad (34a)$$

$$= \int |X|^s \frac{2^{-\frac{(v-d-1)d}{2}} |V|^{\frac{v-d-1}{2}}}{\Gamma_d\left(\frac{v-d-1}{2}\right) |X|^{\frac{v}{2}}} \text{etr}\left(-\frac{1}{2}X^{-1}V\right) dX \quad (34b)$$

$$= \int \frac{2^{-\frac{(v-d-1)d}{2}} |V|^{\frac{v-d-1}{2}}}{\Gamma_d\left(\frac{v-d-1}{2}\right) |X|^{\frac{v-2s}{2}}} \text{etr}\left(-\frac{1}{2}X^{-1}V\right) dX \quad (34c)$$

$$= \frac{\Gamma_d\left(\frac{v-2s-d-1}{2}\right)}{\Gamma_d\left(\frac{v-d-1}{2}\right)} \left(\frac{|V|}{2^d}\right)^s \int \mathcal{IW}(X; v-2s, V) dX \quad (34d)$$

$$= \frac{\Gamma_d\left(\frac{v-d-1}{2} - s\right)}{\Gamma_d\left(\frac{v-d-1}{2}\right)} \left(\frac{|V|}{2^d}\right)^s, \quad (34e)$$

where  $\Gamma_d(\cdot)$  is the multivariate gamma function. By [13, Theorem 1.4.1], the logarithm of  $\Gamma_d(\cdot)$  can be expressed as

$$\log \Gamma_d(a) = d(d-1) \log \pi + \sum_{i=1}^d \log \Gamma\left(a - \frac{i-1}{2}\right). \quad (35)$$

The expected value of  $y$  is

$$\mathbb{E}[y] = \mathbb{E}[\log |X|] \quad (36a)$$

$$= \left. \frac{d}{ds} \left( \frac{\Gamma_d\left(\frac{v-d-1}{2} - s\right)}{\Gamma_d\left(\frac{v-d-1}{2}\right)} \left(\frac{|V|}{2^d}\right)^s \right) \right|_{s=0} \quad (36b)$$

$$= \left(\frac{|V|}{2^d}\right)^s \left. \frac{d}{ds} \frac{\Gamma_d\left(\frac{v-d-1}{2} - s\right)}{\Gamma_d\left(\frac{v-d-1}{2}\right)} \right|_{s=0} + \left. \frac{\Gamma_d\left(\frac{v-d-1}{2} - s\right)}{\Gamma_d\left(\frac{v-d-1}{2}\right)} \frac{d}{ds} \left(\frac{|V|}{2^d}\right)^s \right|_{s=0} \quad (36c)$$

$$= \left(\frac{|V|}{2^d}\right)^s \left. \frac{d}{ds} \log \Gamma_d\left(\frac{v-d-1}{2} - s\right) \right|_{s=0} + \frac{\Gamma_d\left(\frac{v-d-1}{2} - s\right)}{\Gamma_d\left(\frac{v-d-1}{2}\right)} \left(\frac{|V|}{2^d}\right)^s \log\left(\frac{|V|}{2^d}\right) \Big|_{s=0} \quad (36d)$$

$$= -\sum_{j=1}^d \psi_0\left(\frac{v-d-1}{2} - \frac{j-1}{2}\right) + \log\left(\frac{|V|}{2^d}\right) \quad (36e)$$

$$= \log |V| - d \log 2 - \sum_{j=1}^d \psi_0\left(\frac{v-d-j}{2}\right). \quad (36f)$$

### C. Proof of Theorem 1

The density  $q(\mathbf{x}, X)$  is

$$\begin{aligned} q(\mathbf{x}, X) &\triangleq \arg \min_{q(\mathbf{x}, X)} \text{KL}(p(\mathbf{x}, X) || q(\mathbf{x}, X)) \\ &= \arg \max_{q(\mathbf{x}, X)} \sum_{i=1}^N w_i \iint \mathcal{N}(\mathbf{x}; m_i, P_i) \\ &\quad \times \mathcal{IW}(X; v_i, V_i) \log(q(\mathbf{x}, X)) d\mathbf{x} dX, \end{aligned} \quad (37)$$

where the  $i$ :th double integral over  $\mathbf{x}$  and  $X$  can be rewritten as

$$\begin{aligned} &\iint \mathcal{N}(\mathbf{x}; m_i, P_i) \mathcal{IW}(X; v_i, V_i) \log(q(\mathbf{x}, X)) d\mathbf{x} dX \\ &= \log \bar{w} + \int \mathcal{N}(\mathbf{x}; m_i, P_i) \log \mathcal{N}(\mathbf{x}; m, P) d\mathbf{x} \\ &\quad + \int \mathcal{IW}(X; v_i, V_i) \log \mathcal{IW}(X; v, V) dX. \end{aligned} \quad (38)$$

The integral over  $\mathbf{x}$  simplifies to

$$\int \mathcal{N}(\mathbf{x}; m_i, P_i) \log \mathcal{N}(\mathbf{x}; m, P) d\mathbf{x} \quad (39a)$$

$$= \int \mathcal{N}(\mathbf{x}; m_i, P_i) \left[ -\frac{d}{2} \log(2\pi) - \frac{1}{2} \log |P| - \frac{1}{2} \text{Tr}((\mathbf{x} - m)(\mathbf{x} - m)^T P^{-1}) \right] d\mathbf{x} \quad (39b)$$

$$= -\frac{d}{2} \log(2\pi) - \frac{1}{2} \log |P| - \frac{1}{2} \text{Tr}(\mathbb{E}_{P_i}[(\mathbf{x} - m)(\mathbf{x} - m)^T] P^{-1}) \quad (39c)$$

$$= -\frac{d}{2} \log(2\pi) - \frac{1}{2} \log |P| - \frac{1}{2} \text{Tr}((P_i + (m_i - m)(m_i - m)^T) P^{-1}) \quad (39d)$$

$$\triangleq f_i(m, P) \quad (39e)$$

and the integral over  $X$  simplifies to

$$\int \mathcal{IW}(X; v_i, V_i) \log \mathcal{IW}(X; v, V) dX \quad (40a)$$

$$= \int \mathcal{IW}(X; v_i, V_i) \left[ -\frac{(v-d-1)d}{2} \log 2 + \frac{v-d-1}{2} \log |V| - \log \Gamma_d\left(\frac{v-d-1}{2}\right) - \frac{v}{2} \log |X| + \text{Tr}\left(-\frac{1}{2}X^{-1}V\right) \right] dX \quad (40b)$$

$$= -\frac{(v-d-1)d}{2} \log 2 + \frac{v-d-1}{2} \log |V| - \log \Gamma_d\left(\frac{v-d-1}{2}\right) - \frac{v}{2} \mathbb{E}_{P_i}[\log |X|] - \frac{1}{2} \text{Tr}(\mathbb{E}_{P_i}[X^{-1}]V) \quad (40c)$$

$$= -\frac{(v-d-1)d}{2} \log 2 + \frac{v-d-1}{2} \log |V| - \log \Gamma_d\left(\frac{v-d-1}{2}\right)$$

$$-\frac{v}{2} \left( \log |V_i| - d \log 2 - \sum_{j=1}^d \psi_0 \left( \frac{v_i - d - j}{2} \right) \right) - \frac{1}{2} \text{Tr} \left( (v_i - d - 1) V_i^{-1} V \right) \quad (40d)$$

$$\triangleq g_i(v, V) \quad (40e)$$

where the expected values are derived above. We thus have

$$q(\mathbf{x}, X) \triangleq \arg \min_{q(\mathbf{x}, X)} \text{KL}(p(\mathbf{x}, X) || q(\mathbf{x}, X)) \quad (41a)$$

$$= \arg \max_{q(\mathbf{x}, X)} \sum_{i=1}^N w_i (\log \bar{w} + f_i(m, P) + g_i(v, V)) \quad (41b)$$

$$= \arg \max_{q(\mathbf{x}, X)} h(m, P, v, V). \quad (41c)$$

Differentiating the objective function  $h(\cdot)$  w.r.t.  $m$ , setting equal to zero and solving for  $m$  gives

$$m = \frac{1}{\bar{w}} \sum_{i=1}^N w_i m_i. \quad (42)$$

Differentiating the objective function  $h(\cdot)$  w.r.t.  $P$ , setting equal to zero and solving for  $P$  gives

$$P = \frac{1}{\bar{w}} \sum_{i=1}^N w_i (P_i + (m_i - m)(m_i - m)^T). \quad (43)$$

Differentiating the objective function  $h(\cdot)$  w.r.t.  $V$ , setting equal to zero and solving for  $V$  gives

$$V = \bar{w} (v - d - 1) \left( \sum_{i=1}^N w_i (v_i - d - 1) V_i^{-1} \right)^{-1}. \quad (44)$$

Differentiating the objective function  $h(\cdot)$  w.r.t.  $v$ , inserting  $V$  (44), and setting equal to zero gives

$$0 = \bar{w} d \log(v - d - 1) - \bar{w} \sum_{j=1}^d \psi_0 \left( \frac{v - d - j}{2} \right) + \bar{w} d \log \bar{w} - \bar{w} \log \left| \sum_{i=1}^N w_i (v_i - d - 1) V_i^{-1} \right| + \sum_{i=1}^N \sum_{j=1}^d w_i \psi_0 \left( \frac{v_i - d - j}{2} \right) - \sum_{i=1}^N w_i \log |V_i|. \quad (45)$$

#### ACKNOWLEDGMENT

The authors would like to thank the Linnaeus research environment CADICS and the frame project grant Extended Target Tracking (621-2010-4301), both funded by the Swedish Research Council, and the project Collaborative Unmanned Aircraft Systems (CUAS), funded by the Swedish Foundation for Strategic Research (SSF), for financial support.

#### REFERENCES

- [1] Y. Bar-Shalom and X. Rong Li, *Multitarget-Multisensor Tracking: Principles and Techniques*. YBS, 1995.
- [2] B.-N. Vo and W.-K. Ma, "The Gaussian mixture probability hypothesis density filter," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4091–4104, Nov. 2006.
- [3] K. Granström, C. Lundquist, and U. Orguner, "A Gaussian mixture PHD filter for extended target tracking," in *Proceedings of the International Conference on Information Fusion*, Edinburgh, UK, Jul. 2010.
- [4] —, "Extended Target Tracking using a Gaussian Mixture PHD filter," *IEEE Transactions on Aerospace and Electronic Systems*, 2012.
- [5] D. J. Salmond, "Mixture reduction algorithms for target tracking in clutter," in *Proceedings of SPIE Signal and Data Processing of Small Targets*, Orlando, FL, USA, Jan. 1990, pp. 434–445.
- [6] J. L. Williams and P. S. Maybeck, "Cost-Function-Based Gaussian Mixture Reduction for Target Tracking," in *Proceedings of the International Conference on Information Fusion*, Cairns, Queensland, Australia, Jul. 2003.
- [7] A. R. Runnalls, "Kullback-Leibler approach to Gaussian mixture reduction," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 3, pp. 989–999, Jul. 2007.
- [8] M. Huber and U. Hanebeck, "Progressive gaussian mixture reduction," in *Proceedings of the International Conference on Information Fusion*, Cologne, Germany, Jul. 2008.
- [9] D. Schieferdecker and M. F. Huber, "Gaussian Mixture Reduction via Clustering," in *Proceedings of the International Conference on Information Fusion*, Seattle, WA, USA, Jul. 2009.
- [10] D. F. Crouse, P. Willett, K. Pattipati, and L. Svensson, "A Look At Gaussian Mixture Reduction Algorithms," in *Proceedings of the International Conference on Information Fusion*, Chicago, IL, USA, Jul. 2011.
- [11] K. Granström and U. Orguner, "Estimation and Maintenance of Measurement Rates for Multiple Extended Target Tracking," in *Proceedings of the International Conference on Information Fusion*, Singapore, Jul. 2012.
- [12] J. W. Koch, "Bayesian approach to extended object and cluster tracking using random matrices," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 44, no. 3, pp. 1042–1059, Jul. 2008.
- [13] A. K. Gupta and D. K. Nagar, *Matrix variate distributions*, ser. Chapman & Hall/CRC monographs and surveys in pure and applied mathematics. Chapman & Hall, 2000.
- [14] S. Kullback and R. A. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, Mar. 1951.
- [15] M. Feldmann and D. Fränken, "Tracking of Extended Objects and Group Targets using Random Matrices - A New Approach," in *Proceedings of the International Conference on Information Fusion*, Cologne, Germany, Jul. 2008.
- [16] M. Feldmann, D. Fränken, and J. W. Koch, "Tracking of extended objects and group targets using random matrices," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1409–1420, Apr. 2011.
- [17] M. Ulmke, O. Erdinc, and P. Willett, "Gaussian Mixture Cardinalized PHD Filter for Ground Moving Target Tracking," in *Proceedings of the International Conference on Information Fusion*, Quebec City, Canada, Jul. 2007, pp. 1–8.
- [18] U. Orguner, C. Lundquist, and K. Granström, "Extended Target Tracking with a Cardinalized Probability Hypothesis Density Filter," in *Proceedings of the International Conference on Information Fusion*, Chicago, IL, USA, Jul. 2011, pp. 65–72.
- [19] K. Granström and U. Orguner, "A PHD filter for tracking multiple extended targets using random matrices," *IEEE Transactions on Signal Processing*.
- [20] C. M. Bishop, *Pattern recognition and machine learning*. New York, USA: Springer, 2006.
- [21] T. Minka, "A family of algorithms for approximate Bayesian inference," Ph.D. dissertation, Massachusetts Institute of Technology, Jan. 2001.
- [22] F. Gustafsson, *Statistical Sensor Fusion*. Studentlitteratur, 2010.
- [23] J. Stoer and R. Bulirsch, *Introduction to Numerical Analysis*, 2nd ed. New York: Springer-Verlag, 1993.
- [24] G.-C. Rota, "The number of partitions of a set," *The American Mathematical Monthly*, vol. 71, no. 5, pp. 498–504, May 1964.