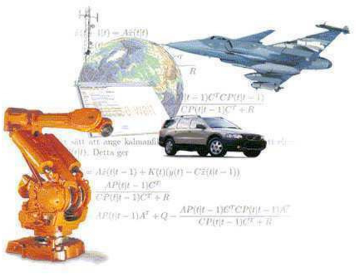
On the Reduction of Gaussian inverse Wishart Mixtures



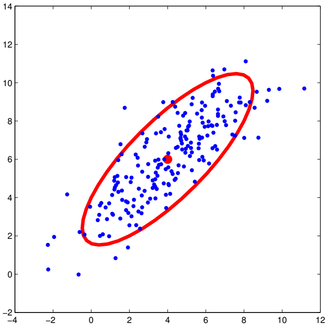
Karl Granström*, Umut Orguner**

*Division of Automatic Control
Department of Electrical Engineering
Linköping University, Sweden

**Department of Electrical
and Electronics Engineering
Middle East Technical University, Turkey

Introduction

- Extended targets ξ_k give rise to multiple meas. $\mathbf{z}_k^{(j)}$
- Random matrix framework by Koch (2008) decomposes ξ_k into
  - Kinematical state vector $\mathbf{x}_k \in \mathbb{R}^{n_x}$
  - Extension matrix $X_k \in \mathbb{S}_{++}^{d}$



- Feldmann *et al.* (2008) model the extended target state $\xi_k = (\mathbf{x}_k, X_k)$ as Gaussian inverse Wishart (GIW) distributed

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

- Multiple targets, clutter and association uncertainty lead to the use of distribution mixtures

$$p(\xi_k) = \sum_{i=1}^{J_{k|k}} w_i \mathcal{N}\left(\mathbf{x}_k\,;\, m_{k|k}^{(i)}, P_{k|k}^{(i)}\right) \mathcal{IW}\left(X_k\,;\, v_{k|k}^{(i)}, V_{k|k}^{(i)}\right)$$

- With time the number of components $J_{k|k}$ increases, and mixture reduction becomes necessary for computational tractability.

Mixture reduction

- Pruning and/or merging typically used for reduction.
- Merging well studied for Gaussian mixtures.
  - Top-down or bottom-up.
  - Local or global consideration of mixture information.
- No previous results for GIW mixtures.

Problem formulation

- Merging for GIW mixtures:
  1. Which single component best approximates a sum of N components?
  2. Which GIW components should be merged?

- Merging for GIW mixtures:
  1. Which single component best approximates a sum of N components?
  2. Which GIW components should be merged?
- Our approach: Use the Kullback-Leibler divergence (KL-div)
  1. Minimize KL-div of single component and sum.
  2. Merging criterion based on the KL *difference*.

Problem formulation

- Merging for GIW mixtures:
    1. Which single component best approximates a sum of N components?
    2. Which GIW components should be merged?
- Our approach: Use the Kullback-Leibler divergence (KL-div)
    1. Minimize KL-div of single component and sum.
    2. Merging criterion based on the KL *difference*.

Contributions:

- Theorem 1: Gives the parameters for the GIW component that minimizes the KL-div to the sum of N GIW components.
- Merging criterion for pairs of GIW components.

- KL-div is a measure of how similar $q(x)$ and $p(x)$ are

$$\mathrm{KL}\left(p||q\right) = \int p(x) \log\left(\frac{p(x)}{q(x)}\right) \mathrm{d}x,$$

- Positive, $\mathrm{KL}\left(p||q\right) \geq 0$.
- In general asymmetric, $\mathrm{KL}\left(p||q\right) \neq \mathrm{KL}\left(q||p\right)$.

Kullback-Leibler divergence

- KL-div is a measure of how similar $q(x)$ and $p(x)$ are

$$\mathrm{KL}\left(p||q\right) = \int p(x) \log\left(\frac{p(x)}{q(x)}\right) \mathrm{d}x,$$

- Positive, $\mathrm{KL}\left(p||q\right) \geq 0$.
- In general asymmetric, $\mathrm{KL}\left(p||q\right) \neq \mathrm{KL}\left(q||p\right)$.
- Well known moment matching characteristcs.
- Considered optimal difference measure in max likelihood sense.
- Minimizing KL-div can be rewritten as maximization problem

$$\min_q \mathrm{KL}\left(p||q\right) = \max_q \int p(x) \log\left(q(x)\right) \mathrm{d}x.$$

Let $p(\cdot)$ be a weighted sum of GIW components,

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

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

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

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

$$q(\mathbf{x}, X) \triangleq \underset{q(\mathbf{x}, X) \in \text{GIW}}{\arg\min}\ \text{KL}\left(p(\mathbf{x}, X)\, || q(\mathbf{x}, X)\right).$$

AUTOMATIC CONTROL
REGLERTEKNIK
LINKÖPINGS UNIVERSITET

The Gaussian parameters m, P are given by

$$m = \frac{1}{\bar{w}} \sum_{i=1}^{N} w_i m_i,$$

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

Corresponds to matching the first and second order moments,

$$\bar{w}\, \mathrm{E}_q \left[\mathbf{x} \right] = \sum_{i=1}^{N} w_i\, \mathrm{E}_{p_i} \left[\mathbf{x} \right],$$

$$\bar{w}\, \mathrm{E}_q \left[\mathbf{x}\mathbf{x}^{\mathrm{T}} \right] = \sum_{i=1}^{N} w_i\, \mathrm{E}_{p_i} \left[\mathbf{x}\mathbf{x}^{\mathrm{T}} \right].$$

Theorem 1 – inverse Wishart parameters

The inverse Wishart parameter V is given by

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

and the parameter v is the solution to the equation

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

AUTOMATIC CONTROL
REGLERTEKNIK
LINKÖPINGS UNIVERSITET

Corresponds to matching the expected values of X_k^{-1} and $\log |X_k|$,

$$\bar{w}\, \mathrm{E}_q \left[X^{-1} \right] = \sum_{i=1}^{N} w_i\, \mathrm{E}_{p_i} \left[X^{-1} \right],$$

$$\bar{w}\, \mathrm{E}_q \left[\log |X| \right] = \sum_{i=1}^{N} w_i\, \mathrm{E}_{p_i} \left[\log |X| \right].$$

Conjecture: there is a unique solution v.

It can be obtained using numerical root finding, e.g. Newton's algorithm. Quick convergence in practice.

AUTOMATIC CONTROL
REGLERTEKNIK
LINKÖPINGS UNIVERSITET

Theorem 1 – proof

- The proof is simple.
- Requires basic knowledge of calculus.
- Full details given in the paper.

AUTOMATIC CONTROL
REGLERTEKNIK
LINKÖPINGS UNIVERSITET

- Kullback-Leibler *difference* for components i and j defined as

$$D_{\mathrm{KL}}\left(p_i, p_j\right) = \mathrm{KL}\left(p_i || p_j\right) + \mathrm{KL}\left(p_j || p_i\right)$$

- Due to assumed conditional independence it becomes

$$\begin{aligned} D_{\mathrm{KL}}\left(p_i, p_j\right) =& D_{\mathrm{KL}}\left(\mathcal{N}\left(\mathbf{x}\,;\, m_i, P_i\right), \mathcal{N}\left(\mathbf{x}\,;\, m_j, P_j\right)\right) \\ &+ D_{\mathrm{KL}}\left(\mathcal{IW}\left(X\,;\, v_i, V_i\right), \mathcal{IW}\left(X\,;\, v_j, V_j\right)\right) \\ =& D_{\mathrm{KL}}^{\mathcal{N}} + D_{\mathrm{KL}}^{\mathcal{IW}}, \end{aligned}$$

AUTOMATIC CONTROL
REGLERTEKNIK
LINKÖPINGS UNIVERSITET

Merging criterion

- Kullback-Leibler *difference* for components i and j defined as

$$D_{\mathrm{KL}}\left(p_i,p_j\right) = \mathrm{KL}\left(p_i||p_j\right) + \mathrm{KL}\left(p_j||p_i\right)$$

- Due to assumed conditional independence it becomes

$$\begin{aligned} D_{\mathrm{KL}}\left(p_i,p_j\right) =& D_{\mathrm{KL}}\left(\mathcal{N}\left(\mathbf{x}\,;\, m_i, P_i\right), \mathcal{N}\left(\mathbf{x}\,;\, m_j, P_j\right)\right) \\ &+ D_{\mathrm{KL}}\left(\mathcal{IW}\left(X\,;\, v_i, V_i\right), \mathcal{IW}\left(X\,;\, v_j, V_j\right)\right) \\ =& D_{\mathrm{KL}}^{\mathcal{N}} + D_{\mathrm{KL}}^{\mathcal{IW}}, \end{aligned}$$

- Merge components i and j if

$$D_{\mathrm{KL}} < U \quad \text{or if} \quad \left(D_{\mathrm{KL}}^{\mathcal{N}} < U_{\mathcal{N}}\right) \,\&\, \left(D_{\mathrm{KL}}^{\mathcal{IW}} < U_{\mathcal{IW}}\right)$$

- Further analysis of criterion in the paper.

Components are bundled for merging as follows:

1. Let I be set of all components.
2. Pick component with highest weight, $j = \arg\max_{i \in I} w_{k|k}^{(i)}$
3. Merge with components in L,

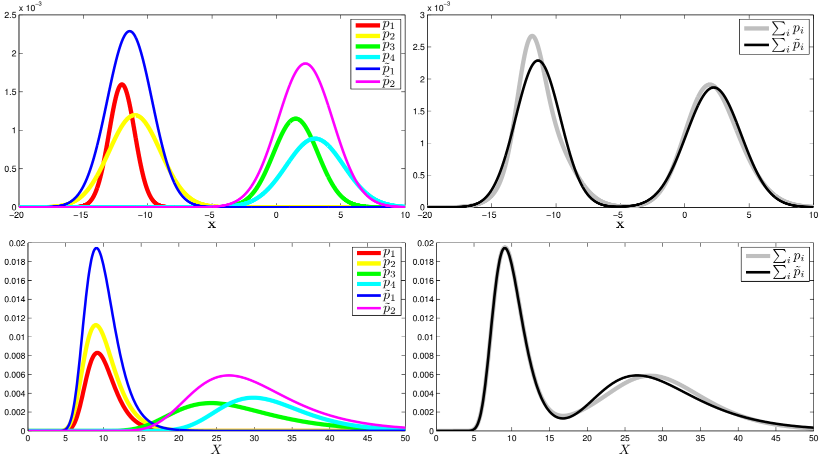
$$L_1 = \left\{ i \in I \middle| D_j^i < U \right\}$$

$$L_2 = \left\{ i \in I \middle| \exists \left\{ i_1 = i, \ldots, i_N = j \right\} \ni D_{i_k}^{i_{k+1}} < U, \, k = 1, \ldots, N-1 \right\}$$

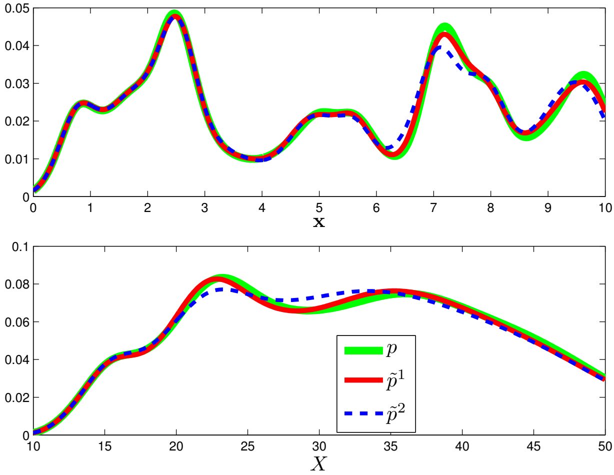
where $D_j^i = D_{\mathrm{KL}}\left(p_i, p_j\right)$

L_2 results in fewer components, but cruder approximation of mixture.

Four GIW components, $n_x = d = 1$, use L_1.

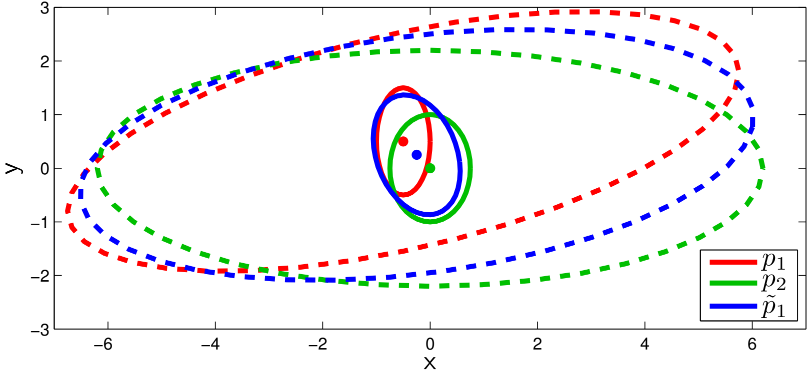


50 GIW components, $n_x = d = 1$.



Results

2 GIW components, $n_x = d = 2$.



Works in higher dimensions too.

Contributions:

- Theorem 1: Gives the parameters for the GIW component that minimizes the KL-div to the sum of N GIW components.
- Merging criterion for pairs of GIW components.

Summary and future work

Contributions:

- Theorem 1: Gives the parameters for the GIW component that minimizes the KL-div to the sum of N GIW components.
- Merging criterion for pairs of GIW components.

Future work:

- Use in GIW-PHD filter for multiple extended target tracking.

Thank you for listening!

Any questions?