CONSENSUS FOR CONTINUOUS BELIEF FUNCTIONS

Zhiyuan Weng and Petar M. Djurić

Department of Electrical and Computer Engineering
Stony Brook University, Stony Brook, NY 11790, USA
E-mails: {zhiyuan.weng, petar.djuric}@stonybrook.edu

ABSTRACT

We study the belief consensus problem in networks of agents. Unlike previous work in the literature, where agents try to reach consensus on a scalar or vector, here we investigate how agents can reach a consensus on a continuous probability distribution. In our setting, the agents fuse functions instead of point estimates. The objective is that every agent ends up with the belief being the global Bayesian posterior. We show that to achieve the objective, the agents need to know the number of total agents in the network. In many scenarios, this number is not available and therefore the global Bayesian posterior is not achievable. In such cases, we have to resort to approximation methods. We confine ourselves to Gaussian cases and formulate the optimization problem for them. Then we propose two methods for the selection of weighting coefficients used for combining information from neighbors in the fusion process. We also provide results of simulation that demonstrate the performance of the methods.

Index Terms— Agent networks, belief consensus, Covariance Intersection, fusion of probability distributions.

1. INTRODUCTION

Consensus problems have been studied for decades since they were raised in the 1970's. They have a wide range of applications in many areas, including social learning, wireless sensor networks and distributed computing. There are many articles on consensus problems in the literature, but most of them are focused on point estimates, i.e., the consensus on a scalar, a vector and/or a discrete distribution. In [1], the problem of consensus for a discrete distribution has been studied. It is one of the earliest papers on consensus problems. In [2], the authors have addressed the consensus of a likelihood ratio for target detection in sensor networks. A convex optimization technique for accelerating the convergence to consensus has been used in [3] and [4]. Consensus problems in dynamic network have also been studied [5]. In [6], the authors have investigated a broadcasting-based gossiping algorithm to compute the average of the initial measurements of the agents. An approach for increasing the convergence rate of the consensus based on an asymmetric interaction mechanism with time-varying weights has been introduced in [7]. In [8], the authors examine the problem of designing weights when the network is subject to random link failures and switching topology. An early work on consensus for belief functions is [9], but it only contains results on discrete functions.

In this paper, we study the belief consensus problem where the belief is a continuous probability distribution. The objective is that the agents in the network reach consensus on the posterior given all the observations in the network. We argue that this is impossible to achieve without knowing the number of agents in the network. We then propose two methods to approximate the posterior. Simulation results are provided to demonstrate the performance of the methods.

The paper is organized as follows. The problem is formulated in Section 2. In Section 3, we review the Covariance Intersection method. We propose the new approaches in Sections 4 and 5 where two criteria are used. In Section 6, we discuss how we use the proposed approach to perform the consensus in the network. Experimental results are provided in Section 7. Section 8 concludes the paper.

2. PROBLEM FORMULATION

Suppose there are N agents in a network, each of them interested in the state of $\mathbf{x}$. The agents have a prior belief about $\mathbf{x}$, denoted by $p(\mathbf{x})$. The agents also receive measurements $\mathbf{y}_i$ with information about $\mathbf{x}$. The belief of agent i is the posterior that it forms about $\mathbf{x}$ and which is given by $p_i(\mathbf{x}|\mathbf{y}_i) \propto p(\mathbf{y}_i|\mathbf{x})\, p(\mathbf{x})$, where $\propto$ signifies "proportional to" and $p(\mathbf{y}_i|\mathbf{x})$ is the likelihood of $\mathbf{x}$. We study the problem of reaching a consensus in the belief of the agents by iterative exchange of information among them.

According to Bayes' theorem and with the assumption of conditionally independent observations, the global optimal belief is the posterior

$$p(\mathbf{x}|\mathbf{y}_1, \cdots \mathbf{y}_N) = \frac{p(\mathbf{x}) \prod_{i=1}^{N} p(\mathbf{y}_i|\mathbf{x})}{\int p(\mathbf{x}) \prod_{i=1}^{N} p(\mathbf{y}_i|\mathbf{x})\, \mathrm{d}\mathbf{x}}. \quad (1)$$

This work was supported by NSF under Awards CCF-1320626 and ECCS-1346854.

In this paper, we address the following problem: can (1) be obtained in a distributed way and be the belief at which the consensus is reached. It is not hard to see that the key for reaching this goal is to obtain $p(\mathbf{x})\prod_{i=1}^{N} p(\mathbf{y}_i|\mathbf{x})$. For simplicity, we assume that the prior $p(\mathbf{x})$ is proportional to a constant. Hence, (1) becomes

$$\begin{aligned} p(\mathbf{x}|\mathbf{y}_1, \cdots \mathbf{y}_N) = & \frac{\prod_{i=1}^{N} p(\mathbf{y}_i|\mathbf{x})}{\int \prod_{i=1}^{N} p(\mathbf{y}_i|\mathbf{x})\,\mathrm{d}\mathbf{x}} \\ = & \frac{1}{Z_c}\prod_{i=1}^{N} p(\mathbf{y}_i|\mathbf{x}). \end{aligned} \tag{2}$$

Therefore, it is crucial to obtain

$$\prod_{i=1}^{N} p(\mathbf{y}_i|\mathbf{x}). \tag{3}$$

Unfortunately, without knowing the number of agents in the network, it is not possible that each agent obtains (2). To see this, recall that a consensus usually ends up with an average value or a weighted average and that it is not possible to reach a sum using consensus-like iterations. Similarly, if we take the logarithm in (2), we obtain the log-posterior which becomes a sum of log-likelihoods, i.e.,

$$-\log Z_c + \sum_{i=1}^{N} \log p(\mathbf{y}_i|\mathbf{x}), \tag{4}$$

and therefore it is not possible to obtain it in a simple distributed way. Admittedly, the number of agents can be obtained in a distributed way, but for dynamic networks, it is not practical, if possible. Our solution to the problem is to use the weighted average

$$\sum_{i=1}^{N} w_i \log p(\mathbf{y}_i|\mathbf{x}) \tag{5}$$

to approximate (3) with the constraint $\sum_{i=1}^{N} w_i = 1, w_i \geq 0$. Since Z_c can be determined once (3) is obtained, we simply ignore it for now. If we remove the logarithm in (5), we have

$$\prod_{i=1}^{N} p^{w_i}(\mathbf{y}_i|\mathbf{x}). \tag{6}$$

For the approximation of (3) by (6), we use two metrics for measuring the distance between probability distributions: the Kullback-Leibler (KL) divergence and the χ^2 information metric [10]. The former is defined as

$$\mathrm{KL}(p||q) = \int_{-\infty}^{\infty} p(x)\ln\left(\frac{p(x)}{q(x)}\right)\mathrm{d}x, \tag{7}$$

and the latter by

$$\chi^2(p||q) = \int \frac{p^2(x)}{q(x)}\mathrm{d}x - 1. \tag{8}$$

Formally, we are trying to minimize

$$\mathrm{KL}\left(\prod_{i=1}^{N} p^{w_i}(\mathbf{y}_i|\mathbf{x}) \,\Big|\Big|\, \prod_{i=1}^{N} p(\mathbf{y}_i|\mathbf{x})\right) \tag{9}$$

and

$$\chi^2\left(\prod_{i=1}^{N} p^{w_i}(\mathbf{y}_i|\mathbf{x}) \,\Big|\Big|\, \prod_{i=1}^{N} p(\mathbf{y}_i|\mathbf{x})\right) \tag{10}$$

with respect to w_i under the constraint $\sum_{i=1}^{N} w_i = 1, w_i \geq 0$.

The proposed criteria are widely applicable, but for general distributions the $\mathrm{KL}(p||q)$ and $\chi^2(p||q)$ metrics do not have closed-form expressions. Therefore, they are hard to analyze.

In this paper, we study the case when the likelihood is Gaussian. The observation model of agent i is

$$\mathbf{y}_i = \mathbf{x} + \mathbf{w}_i, \tag{11}$$

where $\mathbf{w}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_i)$. We assume that $\mathbf{C}_i$ is known to agent i. Thus, for agent i, with observation $\mathbf{y}_i$, the likelihood is $\mathcal{N}(\mathbf{y}_i|\mathbf{x}, \mathbf{C}_i)$. Due to the symmetry of the Gaussian densities, we can exchange $\mathbf{x}$ with $\mathbf{y}_i$ without changing the function, i.e., express $\mathcal{N}(\mathbf{y}_i|\mathbf{x}, \mathbf{C}_i)$ as $\mathcal{N}(\mathbf{x}|\mathbf{y}_i, \mathbf{C}_i)$. Moreover, since $\mathbf{y}_i$ is the only measurement from agent i, this agent takes $\mathbf{y}_i$ as the mean of the initial distribution (belief). Therefore, we substitute $\mathbf{y}_i$ with notation $\mathbf{m}_i$ for the sake of readability. We note that given M Gaussian densities $\mathcal{N}(\mathbf{x}|\mathbf{m}_i, \mathbf{C}_i)$ for $i \in \{1, \cdots N\}$, the product $\prod_{i=1}^{N} \mathcal{N}(\mathbf{x}|\mathbf{m}_i, \mathbf{C}_i)$ is still a Gaussian. Let

$$\mathcal{N}(\mathbf{x}|\mathbf{m}_c, \mathbf{C}_c) = \frac{1}{Z_c}\prod_{i=1}^{N} \mathcal{N}(\mathbf{x}|\mathbf{m}_i, \mathbf{C}_i). \tag{12}$$

The mean $\mathbf{m}_c$ and the covariance matrix $\mathbf{C}_c$ can easily be derived [11]. They are given by

$$\mathbf{C}_c = \left(\sum_{i=1}^{N} \mathbf{C}_i^{-1}\right)^{-1}, \tag{13}$$

$$\mathbf{m}_c = \mathbf{C}_c\left(\sum_{i=1}^{N} \mathbf{C}_i^{-1}\mathbf{m}_i\right). \tag{14}$$

The product is not a density because it is yet to be normalized. Likewise, the product (6) is also in Gaussian form. If

$$\mathcal{N}(\mathbf{x}|\mathbf{m}_d, \mathbf{C}_d) = \frac{1}{Z_c}\prod_{i=1}^{N} \mathcal{N}^{w_i}(\mathbf{x}|\mathbf{m}_i, \mathbf{C}_i) \tag{15}$$

with $\sum_{i=1}^{N} w_i = 1$, then

$$\mathbf{C}_d = \left(\sum_{i=1}^{N} w_i\mathbf{C}_i^{-1}\right)^{-1}, \tag{16}$$

$$\mathbf{m}_d = \left(\sum_{i=1}^{N} w_i\mathbf{C}_i^{-1}\right)^{-1}\left(\sum_{i=1}^{N} w_i\mathbf{C}_i^{-1}\mathbf{m}_i\right). \tag{17}$$

3. COVARIANCE INTERSECTION

The Covariance Intersection (CI) was first proposed in [12, 13]. The objective of it was to obtain a consistent estimate of the covariance matrix when multiple random variables were linearly combined without knowing the correlation. Here "consistent" means that the resulting covariance is an upper-bound of the true covariance. CI selects the values of w_i such that the determinant or trace of $\mathbf{C}_d$ is minimized. In [14], it is suggested that we can choose the values of the weighting coefficients according to

$$w_i = \frac{1/\text{tr}(\mathbf{C}_i)}{\sum_{j=1}^{K} 1/\text{tr}(\mathbf{C}_j)} \tag{18}$$

as a fast approximation algorithm. As pointed out in [15], the criterion used in CI is equivalent to minimizing the Shannon information of the fused function with the assumption that the fusion functions are Gaussian.

4. KULLBACK-LEIBLER DIVERGENCE

The KL divergence is a non-symmetric measure of the difference between two probability distributions, say $p(x)$ and $q(x)$. The definition is given in (7). Suppose we have two normal distributions, $\mathcal{N}_0(\mathbf{m}_0, \mathbf{C}_0)$ and $\mathcal{N}_1(\mathbf{m}_1, \mathbf{C}_1)$. The K-L divergence between $\mathcal{N}_0$ and $\mathcal{N}_1$ is

$$\begin{aligned}\text{KL}(\mathcal{N}_0 || \mathcal{N}_1) &= \frac{1}{2}\Big(\text{tr}\left(\mathbf{C}_1^{-1}\mathbf{C}_0\right) \\ &+ (\mathbf{m}_1 - \mathbf{m}_0)^T \mathbf{C}_1^T (\mathbf{m}_1 - \mathbf{m}_0) - K - \ln\frac{|\mathbf{C}_0|}{|\mathbf{C}_1|}\Big),\end{aligned} \tag{19}$$

where $|\mathbf{C}_0|$ is the determinant of $\mathbf{C}_0$; tr$(\mathbf{C}_0)$ is the trace of $\mathbf{C}_0$; K is the dimension of the variable.

Suppose that the centralized Bayesian posterior is

$$p_c(\mathbf{x}) = \frac{1}{Z_c}\prod_{i=1}^{N} p(\mathbf{x}|\mathbf{m}_i, \mathbf{C}_i), \tag{20}$$

and the distributed consensus belief is

$$p_d(\mathbf{x}) = \frac{1}{Z_d}\prod_{i=1}^{N} p^{w_i}(\mathbf{x}|\mathbf{m}_i, \mathbf{C}_i). \tag{21}$$

We would like to find the weighting coefficients w_i such that the KL divergence KL $(p_c||p_d)$ is minimized. Using (19), we have

$$\begin{aligned}\text{KL}(\mathcal{N}_d || \mathcal{N}_c) &= \frac{1}{2}\Big(\text{tr}\left(\mathbf{C}_c^{-1}\mathbf{C}_d\right) \\ &+ (\mathbf{m}_c - \mathbf{m}_d)^T \mathbf{C}_c^{-1} (\mathbf{m}_c - \mathbf{m}_d) - K - \ln\frac{|\mathbf{C}_d|}{|\mathbf{C}_c|}\Big),\end{aligned} \tag{22}$$

where $\mathbf{m}_c$, $\mathbf{m}_d$, $\mathbf{C}_c$ and $\mathbf{C}_d$ are defined by (14), (17), (13) and (16), respectively. We would like to solve the following optimization problem:

$$\text{minimize KL}(\mathcal{N}_d || \mathcal{N}_c) \tag{23}$$

$$\text{subject to} \sum_{i=1}^{N} w_i = 1, \tag{24}$$

$$w_i \geq 0, \tag{25}$$

where w_i are the variables. In fact, with additional mild assumptions, (22) can be proved to be convex. Due to limited space, the proof is not presented here.

5. χ^2 INFORMATION METRIC

The χ^2 information is another metric for measuring the difference between two continuous probability distributions. The definition is given in (8). Suppose $p(x)$ and $q(x)$ are Gaussian densities, say $\mathcal{N}_0(\mathbf{m}_0, \mathbf{C}_0)$ and $\mathcal{N}_1(\mathbf{m}_1, \mathbf{C}_1)$. Then $p^2(x)/q(x)$ is still in Gaussian a form as long as the covariance matrix $\left(2\mathbf{C}_0^{-1} - \mathbf{C}_1^{-1}\right)^{-1}$ is positive definite. In order to make the Gaussian form valid, i.e., the covariance matrix be positive definite, we let $p(x)$ be (3) and $q(x)$ be (6) so that $2\mathbf{C}_0^{-1} - \mathbf{C}_1^{-1}$ can be positive definite. This can easily be seen from their definitions in (16) and (13).

For $\mathcal{N}_0(\mathbf{m}_c, \mathbf{C}_c)$ and $\mathcal{N}_1(\mathbf{m}_d, \mathbf{C}_d)$, the χ^2 information can be derived as

$$\chi^2(\mathcal{N}_0 || \mathcal{N}_1) = A_2 \exp\left(-\frac{1}{2}A_1\right) - 1,$$

where

$$\begin{aligned}A_1 &= (\mathbf{m}_c - \mathbf{m}_d)^T \left(\frac{\mathbf{C}_c}{2} - \mathbf{C}_d\right)^{-1} (\mathbf{m}_c - \mathbf{m}_d), \\ A_2 &= \frac{\sqrt{|\mathbf{C}_d|}}{|\mathbf{C}_c|}\sqrt{\left|\left(2\mathbf{C}_c^{-1} - \mathbf{C}_d^{-1}\right)^{-1}\right|}.\end{aligned}$$

The optimization problem becomes

$$\text{minimize } \chi^2(\mathcal{N}_d || \mathcal{N}_c) \tag{26}$$

$$\text{subject to} \sum_{i=1}^{N} w_i = 1, \tag{27}$$

$$w_i \geq 0, \tag{28}$$

where w_i are the variables. In most cases, the objective function is convex. We will discuss the convexity elsewhere.

6. APPLICATION OF THE PROPOSED METHOD

We reiterate that the objective is that the agents in a network achieve the global posterior. However, to compute the KL divergence or the χ^2 information, the agents need to know (20), which is not available (we also note that if it was available, there would be no need for these computations). Even if (20) is known, and the agents have the optimal values w_i, the consensus is not always achievable due to the constraints of the network topology.

The agents can only exchange information with their neighbors and not with every agent in the network. To adapt the proposed methods to a distributed consensus problem, we can apply the algorithm locally. We can rewrite (20) and (21) as

$$p_{c,i}(\mathbf{x}) = \frac{1}{Z_c} \prod_{j \in N_i} p(\mathbf{x}|\mathbf{m}_j, \mathbf{C}_j)$$
$$p_{d,i}(\mathbf{x}) = \frac{1}{Z_d} \prod_{j \in N_i} w_j p(\mathbf{x}|\mathbf{m}_j, \mathbf{C}_j).$$

At each iteration, an agent collects information from its neighbors and solves an optimization problem to choose the values of w_j for the fusion. We assume that the consensus process is synchronous in that at each iteration, all the agents are involved in the computation of their next beliefs.

7. NUMERICAL EXPERIMENT

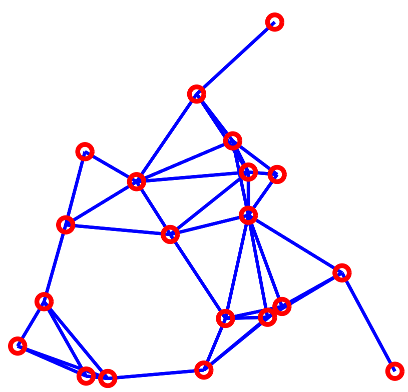

Fig. 1. The topology of the network in the experiment.

In this section, we provide numerical experiments to show the performance of the proposed methods. The topology of the network is shown in Fig. 1. At the beginning, each agent has a belief which is represented as a Gaussian distribution. The mean and the covariance matrix are generated according to a Gaussian distribution and a Wishart distribution, respectively. At each iteration, an agent decides the weighting coefficients using the proposed methods. For the optimization problems in (23) and (26), the gradient method is used to search for the optimal weighting coefficients. We use the KL divergence and the χ^2 information to measure the distance between the optimal local posterior and the belief of the agents as expressed in (9) and (10), respectively.

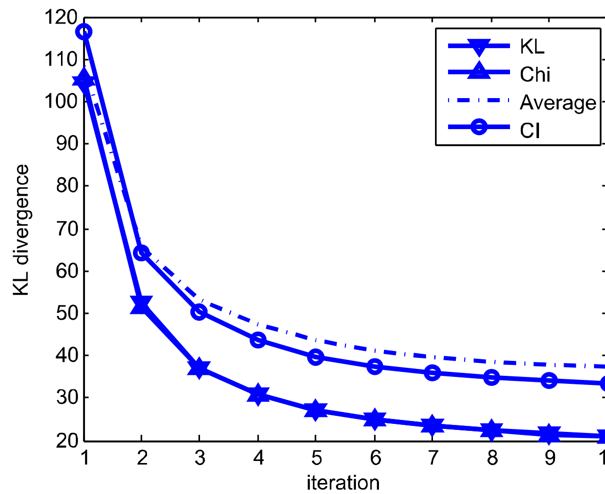

Fig. 2. The KL divergence of the belief consensus using different methods.

In Fig. 2, the KL divergence at each iteration is shown for the four different methods. 'KL' stands for selecting the weighting coefficients using the KL divergence, 'Chi' for using the χ^2 information, 'Average' for using equal weighting coefficients and 'CI' for using the Covariance Intersection. Figure 3 shows the χ^2 information metric for the four methods. All the results are averaged over 1000 runs. We can see that the methods based on the KL divergence and χ^2 information metric perform the same and better than the other two methods.

8. CONCLUSION

In this paper, we proposed a new approach for reaching belief consensus. Unlike traditional consensus, where networks reach consensus at point estimates, the objective in this work was to reach consensus at probability distributions. The ideal probability distribution we hoped to reach was the Bayesian posterior given all the information available over the network. We addressed the cases where the agents do not know the size of the network and where the beliefs are represented by Gaussians. We adopted two criteria for forming beliefs, one based on the KL divergence, and the other, on the χ^2 information metric. They were used for choosing values of the weighting coefficients while forming the approximations. In the numerical experiments, we showed that the two proposed criteria work almost equally well and are better than the Covariance Intersection method and the method based on averaging.

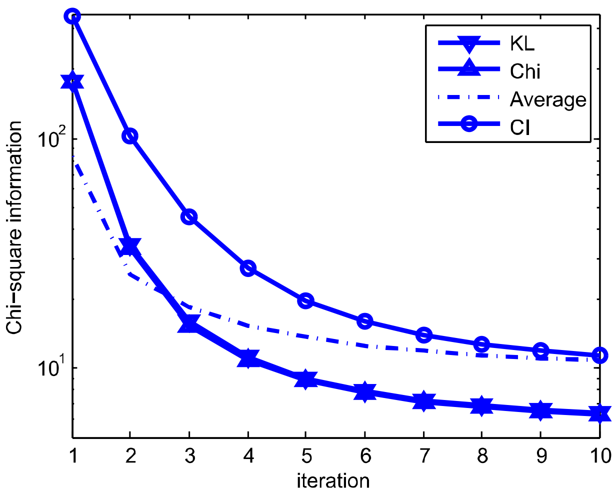

Fig. 3. χ^2 information of the belief consensus using different methods.

9. REFERENCES

[1] M. H. DeGroot, "Reaching a consensus," *Journal of the American Statistical Association*, vol. 69, no. 345, pp. 118–121, 1974.

[2] R. Olfati-Saber, E. Franco, E. Frazzoli, and J. S. Shamma, "Belief consensus and distributed hypothesis testing in sensor networks," in *Networked Embedded Sensing and Control*, pp. 169–182. Springer, 2006.

[3] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems & Control Letters*, vol. 53, no. 1, pp. 65–78, 2004.

[4] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *Information Theory, IEEE Transactions on*, vol. 52, no. 6, pp. 2508–2530, 2006.

[5] R. Olfati-Saber and R. M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *Automatic Control, IEEE Transactions on*, vol. 49, no. 9, pp. 1520–1533, 2004.

[6] T. C. Aysal, M. E. Yildiz, A. D. Sarwate, and A. Scaglione, "Broadcast gossip algorithms for consensus," *Signal Processing, IEEE Transactions on*, vol. 57, no. 7, pp. 2748–2761, 2009.

[7] S. Sardellitti, M. Giona, and S. Barbarossa, "Fast distributed average consensus algorithms based on advection-diffusion processes," *Signal Processing, IEEE Transactions on*, vol. 58, no. 2, pp. 826–842, 2010.

[8] D. Jakovetic, J. Xavier, and J. M. F. Moura, "Weight optimization for consensus algorithms with correlated switching topology," *Signal Processing, IEEE Transactions on*, vol. 58, no. 7, pp. 3788–3801, 2010.

[9] C. G. Wagner, "Consensus for belief functions and related uncertainty measures," *Theory and Decision*, vol. 26, no. 3, pp. 295–304, 1989.

[10] M. Vemula, M. F. Bugallo, and P. M. Djurić, "Performance comparison of Gaussian-based filters using information measures," *Signal Processing Letters, IEEE*, vol. 14, no. 12, pp. 1020–1023, 2007.

[11] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook*, Technical University of Denmark, 2006.

[12] S. J. Julier and J. K. Uhlmann, "A non-divergent estimation algorithm in the presence of unknown correlations," in *American Control Conference, 1997. Proceedings of the 1997*. IEEE, 1997, vol. 4, pp. 2369–2373.

[13] J. K. Uhlmann, "General data fusion for estimates with unknown cross covariances," in *Aerospace/Defense Sensing and Controls*. International Society for Optics and Photonics, 1996, pp. 536–547.

[14] D. Franken and A. Hupper, "Improved fast covariance intersection for distributed data fusion," in *Information Fusion, 2005 8th International Conference on*. IEEE, 2005, vol. 1, pp. 7–pp.

[15] M. B. Hurley, "An information theoretic justification for covariance intersection and its generalization," in *Information Fusion, 2002. Proceedings of the Fifth International Conference on*. IEEE, 2002, vol. 1, pp. 505–511.