DATA FUSION BASED ON CONVEX OPTIMIZATION

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

A distributed fusion problem is addressed where cross-covariance matrices of estimated variables are unknown. We first try to estimate the cross-covariances, and then calculate the weighting coefficients to combine the estimates linearly. We consider two approaches, one where we do not use priors for the covariance matrices of the model and another, where we use priors and engage the Bayesian machinery. For the former, we exploit the maximum-entropy principle in finding the optimal cross-covariance estimate and for the latter, we employ Wishart distributions as priors and search for the maximum a posteriori estimate. Both problems turn out to require convex optimization which can be solved by existing techniques. When the cross-covariance estimates are obtained, the weighting coefficients can easily be calculated so that fusion can take place. Simulation results that demonstrate the performance of the proposed methods are provided.

Index Terms— Convex optimization, covariance estimation, data fusion, distributed estimation, maximum entropy

1. INTRODUCTION

Distributed data fusion problems have been drawing great attention from many fields, especially in the wireless sensor network community. The motivation for distributed processing is that it allows for gains in scalability and robustness, which are features that cannot be met with traditional centralized architectures. In many applications, the information propagated through a network is in the form of the estimated states of interest, along with the covariance of estimation error. A well-known difficulty in this setup is that the estimates, based on the observations collected from different places, have unknown cross-covariances. This is particularly true for networks with unknown topologies. Many approaches [1, 2, 3, 4] have been proposed to mitigate this problem. A popular one is known as the covariance intersection [5] method, which yields consistent estimates between the fused local estimates.

In this paper, we propose a convex optimization method to solve the problem. Our strategy is to estimate the cross-covariance first and then fuse the information from the various sources. We consider two cases, without and with priors: if we do not have any prior information about the covariance matrix, we can use the maximum-entropy (ME) principle [6] as a criterion in the search for the optimal cross-covariance; if we have priors of the covariance matrices, in estimating them, we maximize the a posteriori distributions of these matrices. The problems in both cases can be formulated as convex optimization problems and therefore, they can readily be solved by some well-known methods.

This work was supported by NSF under Award CCF-1018323.

The paper is organized as follows. Relation to prior work and contributions are briefly discussed in Section 2. The problem is introduced and formulated in Section 3. In Sections 4 and 5, the cases without and with priors are considered, respectively. The problems in both cases are formulated as convex optimization problems. We discuss the solutions in Section 6. In Section 7, numerical results are presented that show the performance of our methods. Section 8 concludes the paper.

The notation we use in this paper is as follows. Uppercase letters refer to matrices and lowercase letters to vectors or scalars; $|A|$ is the determinant of a matrix A; $A \succeq B$ means $A - B$ is a positive definite matrix; $x \sim p(x)$ signifies that the random variable x is distributed according to $p(x)$; $\mathrm{tr}(A)$ is the trace of the matrix A; I_n is the identity matrix with size $n \times n$; $\Gamma(\cdot)$ is the standard gamma function, and $\Gamma_n(\cdot)$ is the multivariate gamma function defined as [7]

$$\Gamma_n(l) = \pi^{n(n-1)/4} \prod_{j=1}^{n} \Gamma\left(l - \frac{1}{2}(j-1)\right). \tag{1}$$

With $\mathcal{N}(\mu, C)$ we denote the multivariate Gaussian distribution of an $n \times 1$ vector, i.e.,

$$f(x) = \frac{1}{(2\pi)^{n/2}|C|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^\top C^{-1}(x-\mu)\right), \tag{2}$$

where μ is the mean and C is the covariance matrix. The $n \times n$ random matrix A is said to have a Wishart distribution if its probability distribution function (pdf) is given by

$$p(A) = \frac{|A|^{\frac{l-n-1}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}A)\right)}{2^{\frac{nl}{2}}|\Sigma|^{\frac{l}{2}}\Gamma_n(\frac{l}{2})}, \tag{3}$$

in which Σ is a positive definite matrix, l is the degree of freedom and Γ_n is defined by (1). We use $\mathcal{W}_n(l, \Sigma)$ to denote the Wishart distribution.

2. RELATION TO PRIOR WORK

In the literature, there are various studies on the fusion problem where the cross-covariance matrices are unknown [2, 3, 5, 4]. A naïve but simple method is to use the weighted average, where the weighting coefficients are proportional to the degrees of the nodes (the numbers of neighbors of the nodes) [8]. The approach makes sense because if the node has a higher degree, it collects more information and therefore has better estimates. A more complicated and popular method is known as the covariance intersection method [5]. It yields consistent estimates between the fused local estimates. In order to reduce the computational complexity, several suboptimal non-iterative algorithms for fast covariance intersection have been

developed [9, 10]. In [11], we proposed to put the cross-covariance estimation problem into the Bayesian framework and employ the Monte Carlo method to provide a minimum mean square error estimate. All of the above approaches do not estimate the cross-covariance matrices. In this paper, we first aim at applying a convex optimization method to explicitly estimate the cross-covariance matrices, and then fuse the information from various sources accordingly. Our simulation experiments indicate that with this approach we obtain better performance.

3. PROBLEM FORMULATION

Consider that a node in a network has $k-1$ nodes in its neighborhood. By communication with its neighbors, it has k available estimates, including its own estimate. Each estimate x_i for $i \in \{1, \cdots, k\}$ is a $n \times 1$ vector, with the covariance matrices of the estimation error P_{ii}. We concatenate the k vectors and let

$$x = \begin{bmatrix} x_1^\top & x_2^\top & \cdots & x_k^\top \end{bmatrix}^\top, \tag{4}$$

where $x \in \mathbb{R}^{nk \times 1}$. We assume that the mean of x_i is the true state x_0. Therefore, the covariance matrix of x is also the covariance matrix of the estimation error of x. We denote with P_x the covariance matrix of x, where

$$P_x = \begin{bmatrix} P_{11} & \cdots & P_{1k} \\ \vdots & \ddots & \vdots \\ P_{1k}^\top & \cdots & P_{kk} \end{bmatrix}. \tag{5}$$

We start by considering a linear and unbiased estimator in the form

$$\hat{x}_0 = W^\top x, \tag{6}$$

where W is the weighting coefficient matrix

$$W = \begin{bmatrix} W_1 & W_2 & \cdots & W_k \end{bmatrix}^\top, \tag{7}$$

with $W_i \in \mathbb{R}^{n \times n}$. Since the estimator is unbiased, we require

$$W_1 + W_2 + \cdots + W_k = I. \tag{8}$$

Let $I_{(k)}$ be a $kn \times n$ matrix concatenated vertically by k identity matrices with size $n \times n$,

$$I_{(k)} = \begin{bmatrix} I_n & I_n & \cdots & I_n \end{bmatrix}^\top. \tag{9}$$

Then (8) becomes $W^\top I_{(k)} = I$. Let P_0 be the covariance matrix of $\hat{x}_0$, which can be expressed as

$$P_0 = W^\top E(xx^\top) W = W^\top P_x W. \tag{10}$$

The mean square error is just the trace of P_0. The minimization of the mean square error subject to the constraints (8) can be carried out by using the method of Lagrange multipliers. The solution becomes

$$W^\top = \left(I_{(k)}^\top P_x^{-1} I_{(k)} \right)^{-1} I_{(k)}^\top P_x^{-1} \tag{11}$$

$$P_0 = W^\top P_x W = \left(I_{(k)}^\top P_x^{-1} I_{(k)} \right)^{-1}. \tag{12}$$

By substituting (11) into (6), we have

$$\hat{x}_0 = \left(I_{(k)}^\top P_x^{-1} I_{(k)} \right)^{-1} I_{(k)}^\top P_x^{-1} x. \tag{13}$$

This estimate is optimal in the sense of being unbiased and with minimum mean square error. In many situations, however, we do not have information about P_{ij} for $i \neq j$, as is the case in the distributed Kalman filtering problem [8]. Therefore, we are unable to obtain W or fuse the information.

Now suppose that the true state x_0 is a random vector with zero mean and covariance C_0, and that x_i is the estimate of x_0 corrupted by a zero mean noise with covariance C_i, for $i \in \{1, \cdots, k\}$. Also, we assume that the noises are independent of each other and independent of x_0. The covariance matrix of x, P_x, which is defined in (5), becomes

$$P_x = \begin{bmatrix} C_1 + C_0 & C_0 & \cdots & C_0 \\ C_0 & C_2 + C_0 & \cdots & C_0 \\ \vdots & \vdots & \ddots & \vdots \\ C_0 & C_0 & \cdots & C_k + C_0 \end{bmatrix}. \tag{14}$$

In the fusion problem, we know the diagonal blocks of the covariance matrix, i.e., P_{ii}. Note $P_{ii} = C_i + C_0$. But we do not know C_i or C_0. We wish to have an estimate of C_0 so that we can determine the weighting coefficients to combine those x_is.

There are some previous works on this problem. In [5], the authors have proposed the covariance intersection method to minimize the upper bound for all possible P_{ij} by a convex combination of the covariances, i.e., by using

$$P_0^{-1} = \sum_{j=1}^{k} \omega_j P_{jj}^{-1} \tag{15}$$

$$P_0^{-1} \hat{x}_0 = \sum_{j=1}^{k} \omega_j P_{jj}^{-1} x_j, \tag{16}$$

where the weighting coefficients satisfy $\omega_j \in [0, 1]$ and $\sum_{j=1}^{k} \omega_j = 1$. The minimization of the trace requires iterative minimization of the given nonlinear cost function with respect to the weight coefficient ω. In order to reduce the computational complexity, several suboptimal non-iterative algorithms for fast covariance intersection have been developed [9, 10]. One of them sets the ωj according to [9]

$$\omega_j = \frac{1/\text{tr}(P_{jj})}{\sum_{i=1}^{k} 1/\text{tr}(P_{ii})}. \tag{17}$$

In the sequel, we will use this method for comparison with our algorithms.

4. THE MAXIMUM ENTROPY APPROACH

In this section, for the sake of simplicity we make an additional assumption that the model is Gaussian, and we assume $x_0 \sim \mathcal{N}(0, C_0)$ and $x_i | x_0 \sim \mathcal{N}(x_0, C_i)$. Here we do not have the information about the priors of C_0 and C_i, and we propose to exploit the ME principle. The rationale for using the ME principle is discussed thoroughly in [6, 12].

The entropy is basically a functional, i.e., it maps a function f to a real number. It is defined as

$$H(f) = -\int f(y) \log f(y) \mathrm{d}y. \tag{18}$$

Plugging (2) in (18), we obtain the entropy of the multivariate Gaussian distribution,

$$H(f) = -\int_{y\in\mathbb{R}^k} f(y)\left(-\log((2\pi)^{k/2}|C|^{1/2}) - \frac{1}{2}y^\top C^{-1}y\right)\mathrm{d}y \tag{19}$$

$$= \log((2\pi)^{k/2}|C|^{1/2}) + \int_{y\in\mathbb{R}^k}\frac{1}{2}f(y)y^\top C^{-1}y\mathrm{d}y. \tag{20}$$

Let

$$y = [y_1, \cdots, y_n]^\top \tag{21}$$

$$D = C^{-1} = \begin{bmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{1n} & \cdots & c_{nn} \end{bmatrix}^{-1} = \begin{bmatrix} d_{11} & \cdots & d_{1n} \\ \vdots & \ddots & \vdots \\ d_{1n} & \cdots & d_{nn} \end{bmatrix}. \tag{22}$$

We have

$$y^\top C^{-1}y = \sum_{i=1}^n\sum_{j=1}^n d_{ij}y_iy_j. \tag{23}$$

Therefore the second term in (20) becomes

$$\int_{y\in\mathbb{R}^k} f(y)y^\top C^{-1}y\mathrm{d}y \tag{24}$$

$$= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(y)\sum_{i=1}^n\sum_{j=1}^n d_{ij}y_iy_j\,\mathrm{d}y_1\cdots\mathrm{d}y_n \tag{25}$$

$$= \sum_{i=1}^n\sum_{j=1}^n d_{ij}c_{ij} = \mathrm{tr}(DC) = n. \tag{26}$$

Thus the maximization of $H(f)$ reduces to the maximization of $\log(|C|)$.

In order to estimate the cross-covariance, we try to maximize the entropy of the joint distribution of x_0 and x_i. Specifically, we try to maximize $H(p_{x_0,x_1,\cdots,x_k})$. We have

$$H(p_{x_0,x_1,\cdots,x_k}) = H(p_{x_0}) + \sum_{i=1}^k H(p_{x_i|x_0}) \tag{27}$$

$$\propto \log(|C_0|) + \sum_{i=1}^k \log(|C_i|) \tag{28}$$

$$= \log(|C_0|) + \sum_{i=1}^k \log(|P_{ii} - C_0|), \tag{29}$$

where the first equality can be found in [13]. Therefore, the entire optimization problem can be formulated as

$$\text{maximize}\quad \log(|C_0|) + \sum_{i=1}^k \log(|P_{ii} - C_0|) \tag{30}$$

$$\text{subject to}\quad P_{ii} - C_0 \succeq 0 \quad i \in \{1, \cdots, k\} \tag{31}$$

$$C_0 \succeq 0, \tag{32}$$

where the variable is the symmetric matrix C_0. The objective function (30) is a convex function on the positive semidefinite cone [14]; the constraints (31) and (32) are convex sets. Thus the optimization problem can be easily solved by some existing well-known methods, e.g., the interior point method [15].

Before ending this section, we wish to emphasize that the model does not have to be normal. In fact, it can be unknown as long as the first and the second moments are specified. Recall that the normal distribution has ME among all real-valued distributions with given mean and variance [16]. That is to say, even if the model is unknown, we still obtain the same solution if we employ the ME criterion.

5. THE BAYESIAN APPROACH

In this section, we consider the case where the priors of the unknown covariance matrices are available. We use the Wishart distribution as the prior since it is common for positive definite matrices. Suppose the priors of C_0 and C_i are $p_0(C_0) = \mathcal{W}_n(l_0, \Sigma_0)$ and $p_i(C_i) = \mathcal{W}_n(l_i, \Sigma_i)$, respectively. We use the maximum a posteriori (MAP) distribution as a criterion, and the optimal estimator can be written as

$$C_0 = \max_{C\succeq 0}\arg\, p_0(C)\prod_{i=1}^k p_i(P_{ii} - C). \tag{33}$$

If we substitute (3) into (33), we have

$$C_0 = \max_{C\succeq 0}\arg\, g(C), \tag{34}$$

where $g(C)$ is defined as

$$g(C) = \frac{l_0 - n - 1}{2}\log|C| - \frac{1}{2}\mathrm{tr}(\Sigma_0^{-1}C) + \sum_{i=1}^k \frac{l_i - n - 1}{2}\log|P_{ii} - C| - \sum_{i=1}^k \frac{1}{2}\mathrm{tr}(\Sigma_i^{-1}(P_{ii} - C)). \tag{35}$$

The optimization problem can be cast as

$$\text{maximize}\quad g(C) \tag{36}$$

$$\text{subject to}\quad P_{ii} - C \succeq 0 \quad i \in \{1, \cdots, k\} \tag{37}$$

$$C \succeq 0, \tag{38}$$

where the optimization variable is the symmetric matrix C. We know $\log|\cdot|$ is a concave function and $\mathrm{tr}(\cdot)$ is a convex function over the positive semidefinite cone [14]. Therefore $g(C)$ is a concave function with respect to C. Since the constraint also specifies a convex set, the problem is a convex optimization problem as well, which can be solved with no difficulty.

6. DISCUSSION

We point out that (30) and (35) are both with the same $\log(|\cdot|)$ terms. In fact, (30) is a special case of (35), where the hyperparameters Σ_is are infinitely large and make the terms $\mathrm{tr}(\Sigma_0^{-1}C)$ and $\mathrm{tr}(\Sigma_i^{-1}(P_{ii} - C))$ vanish for finite P_{ii} and C. To illustrate the connections between the ME and the Bayesian approaches further, we first associate an ellipsoid to each covariance matrix. The ellipsoid of A is defined as

$$\{x \in \mathbb{R}^n | x^\top A^{-1}x = 1\}. \tag{39}$$

For $n = 2$, the ellipsoid becomes an ellipse. Basically the major and minor axes of the ellipse show how large the variances are in

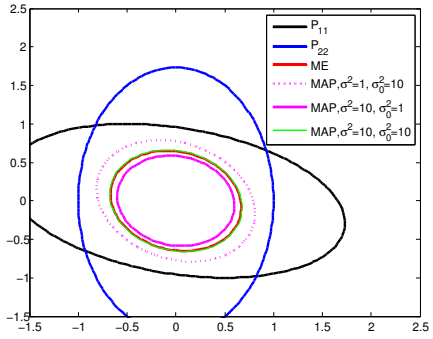

Fig. 1. Illustration of the ellipses defined by P_{11}, P_{22} and the estimated cross-covariances by different methods.

the directions of the axes. The angle between the x-axis and the major axis of the ellipse indicates how much the data from the two dimensions correlate with each other. Figure 1 shows the ellipses of P_{11}, P_{22} and the estimated cross-covariances. We set $k = 2$ for simplicity and let

$$P_{11} = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 3 \end{bmatrix}, \quad P_{22} = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}. \tag{40}$$

The priors are $C_0 \sim \mathcal{W}_2(4, \sigma_0^2 I_2)$, $C_i \sim \mathcal{W}_2(4, \sigma^2 I_2)$. The ellipses for different matrices are shown in Fig. 1. We can see that the ellipses of C_0 are inside those of C_1 and C_2, which makes sense since any point in the feasible set shall make its associated ellipse in the intersection of those of the P_{ii}s. Also, for larger σ_0^2/σ^2 the solution ellipse becomes larger; for smaller σ_0^2/σ^2, the ellipse becomes smaller. When both σ_0^2 and σ^2 are large, in this case $\sigma_0^2 = \sigma^2 = 10$, the ellipse (green) is very close to the ME solution (red).

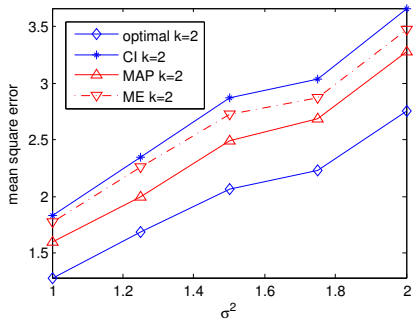

Fig. 2. Performance comparison of fusion for two nodes under different values of σ^2.

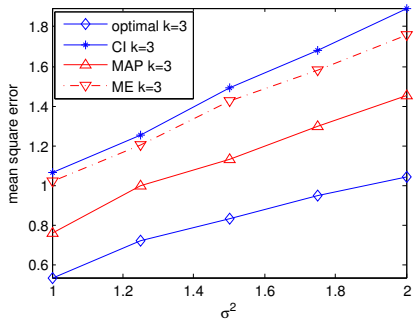

Fig. 3. Performance comparison of fusion for three nodes under different values of σ^2.

7. SIMULATION RESULTS

We use the Gaussian model in the numerical experiment. Suppose that the variable to be estimated is x_0 and that it has distribution $\mathcal{N}(\mu_0, C_0)$. We let $\mu_0 = 0$ for the sake of simplicity. The estimates x_i have the conditional distributions $\mathcal{N}(x_0, C_i)$ for $i \in \{1, \cdots, k\}$. The noise of the measurements is assumed to be independent of each other. We can consider x_i to be measurements as well as estimates, since we shall let $\hat{x}_i = x_i$ if we make estimation only based on x_i. If we concatenate k estimates into one vector x as before, the distribution of the vector conditioned on x_0 is

$$x|\, x_0 \sim \mathcal{N}\left(\begin{bmatrix} x_0 \\ \vdots \\ x_0 \end{bmatrix}, \begin{bmatrix} C_1 & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & C_k \end{bmatrix} \right). \tag{41}$$

The marginal distribution of x becomes $p(x) = \mathcal{N}(x|0, P_x)$, where P_x is defined in (14). The diagonal blocks $C_i + C_0$ are known exactly. On the other hand, neither C_0 nor C_i is known. To generate the data for our numerical experiment, we let x_0 be two-dimensional. First, we drew C_0 from its prior, assumed to be $\mathcal{W}_2(2, \Lambda_1)$ and C_1, $\cdots$, C_k from $\mathcal{W}_2(2, \sigma^2 \Lambda_2)$ independently, where

$$\Lambda_1 = \begin{bmatrix} 4 & -1 \\ -1 & 3 \end{bmatrix}, \quad \text{and} \quad \Lambda_2 = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 2 \end{bmatrix}. \tag{42}$$

Then we generated x_0 by sampling from $\mathcal{N}(0, C_0)$. Similarly, we generated the measurements x_i from $\mathcal{N}(x_0, C_i)$. We set $k = 2, 3$. Then we had all the data we needed for testing and comparing the estimators. For comparison, we used two other estimators, the optimal estimator (13) with all the information (including C_0), and the fast covariance intersection method (17) from [9]. For each configuration, we ran 2000 tests. In the legend, we use *optimal, ME, MAP*, and *CI* to indicate the optimal method, the proposed ME method, the proposed MAP method (with the prior available), and the fast covariance intersection method, respectively.

Figures 2 and 3 show the mean square error performance for $k = 2$ and $k = 3$, respectively. We can see that the proposed methods are better than the CI method in both situations. Meanwhile, as the hyperparameters Σ_0 and Σ_i are much different, the MAP estimator outperforms the ME estimator thanks to its priors.

8. CONCLUSION

In this article, we propose to use the convex optimization techniques to solve the fusion of correlated estimates with unknown correlations. Specifically, given the diagonal block of the error covariance matrix, we cast the problem of estimating cross-covariance as a convex optimization problem which can readily be solved by well-known methods. Two cases are considered: for the non-Bayesian case, we employ the maximum entropy criterion in the search for optimal cross-covariance; for the Bayesian case, we assume that the priors of the unknown covariance matrices follow a Wishart distribution. We then maximize the posterior probability of the cross-covariances. As soon as the cross-covariances are obtained, the weighting coefficients can be determined and the distributed estimates can be combined by a simple calculation. We demonstrated the performance of our methods with numerical experiments.

9. REFERENCES

[1] X. R. Li, Y. Zhu, J. Wang, and C. Han, "Optimal linear estimation fusion. I. Unified fusion rules," *IEEE Transactions on Information Theory*, vol. 49, no. 9, pp. 2192–2208, 2003.

[2] U. D. Hanebeck, K. Briechle, and J. Horn, "A tight bound for the joint covariance of two random vectors with unknown but constrained cross-correlation," in *the Proceedings of the International Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2001, pp. 85–90.

[3] L. Chen, P. O. Arambel, and R. K. Mehra, "Estimation under unknown correlation: Covariance intersection revisited," *IEEE Transactions on Automatic Control*, vol. 47, no. 11, pp. 1879–1882, 2002.

[4] C. Y. Chong and S. Mori, "Convex combination and covariance intersection algorithms in distributed fusion," in *the Proceedings of the 4th International Conference on Information Fusion*, 2001.

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

[6] E. T. Jaynes, "On the rationale of maximum-entropy methods," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 939–952, 1982.

[7] A. T. James, "Distributions of matrix variates and latent roots derived from normal samples," *The Annals of Mathematical Statistics*, pp. 475–501, 1964.

[8] F.S. Cattivelli and A.H. Sayed, "Diffusion strategies for distributed Kalman filtering and smoothing," *IEEE Transactions on Automatic Control*, vol. 55, no. 9, pp. 2069–2084, 2010.

[9] W. Niehsen, "Information fusion based on fast covariance intersection filtering," in *the Proceedings of the IEEE Fifth International Conference on Information Fusion*, 2002, vol. 2, pp. 901–904.

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

[11] Z. Weng and P. M. Djurić, "A Bayesian approach to covariance estimation and data fusion," in *the Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, August 2012, pp. 2352 –2356.

[12] E. T. Jaynes, *Probability Theory: The Logic of Science*, Cambridge University Press, 2003.

[13] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, 2001.

[14] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[15] Y. Nesterov and A. S. Nemirovskii, *Interior-point Polynomial Algorithms in Convex Programming*, vol. 13, SIAM, 1994.

[16] S. Y. Park and A. K. Bera, "Maximum entropy autoregressive conditional heteroskedasticity model," *Journal of Econometrics*, vol. 150, no. 2, pp. 219–230, 2009.