A BAYESIAN APPROACH TO COVARIANCE ESTIMATION AND DATA FUSION

Zhiyuan Weng and Petar M. Djurić

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

ABSTRACT

In this paper, we address the fusion problem of two estimates, where the cross-correlation between the estimates is unknown. To solve the problem within the Bayesian framework, we assume that the covariance matrix has a prior distribution. We also assume that we know the covariance of each estimate, i.e., the diagonal block of the entire covariance matrix (of the random vector consisting of the two estimates). We then derive the conditional distribution of the off-diagonal blocks, which is the cross-correlation of our interest. The conditional distribution happens to be the inverted matrix variate t-distribution. We can readily sample from this distribution and use a Monte Carlo method to compute the minimum mean square error estimate for the fusion problem. Simulations show that the proposed method works better than the popular covariance intersection method.

Index Terms— Covariance Estimation, Data Fusion, Inverted Matrix Variate t-distribution, Monte Carlo Method, Wishart Distribution

1. INTRODUCTION

Distributed data fusion problems have been attracting researchers from many fields, especially in the sensor network community. The motivation for distributed systems is that they can provide a degree of scalability and robustness, which cannot be achieved with traditional centralized architectures. In many applications, the information propagated through a sensor network is transformed to a form that provides the estimated state of interest. In many cases, the information is converted into the first and second moment statistics, which can readily be exploited within the framework of Kalman-type filters [1, 2, 3, 4].

A serious problem in this setup is that the estimates provided by different nodes have unknown cross-correlations. This is particularly true for networks with unknown topological structure. Many approaches have been proposed to mitigate the problem. A popular one is known as the covariance intersection method [5]. It provides a general framework for information fusion with lack of knowledge about cross-correlation between noisy measurements, and it yields consistent estimates between the fused local estimates. Several authors have proposed various versions of this method [6, 7, 8, 9].

In this paper, we use a Bayesian approach to address the problem. We assume that the prior of the covariance matrix is the Wishart distribution. Since we know the covariance matrix of each estimate, which is just the diagonal submatrix of the entire covariance matrix, we can derive the conditional distribution of the off-diagonal submatrices. Furthermore, we show that this conditional distribution is the inverted matrix variate t-distribution. It is known that one can easily sample from this distribution, entailing that we can efficiently use the Monte Carlo method to compute the minimum mean square error (MMSE) estimate. Simulation results show that the proposed method works better than the traditional covariance intersection method.

The paper is organized as follows. We formulate the problem in Section 2 and describe our proposed algorithm in Section 3. In Section 4, we derive the conditional distribution of the off-diagonal block matrix. Simulation results of the proposed algorithm are presented in Section 5. Section 6 concludes our 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; $A > B$ means that $A - B$ is a positive definite matrix; $x \sim p(x)$ signifies that the random variable x is distributed according to $p(x)$; the symbol $\otimes$ denotes Kronecker product; I_k is the identity matrix with size $k \times k$; $\text{tr}(A)$ is the trace of the matrix A; O is a matrix with all entries equal to zero; $\Gamma(\cdot)$ is the standard gamma function, and $\Gamma_k(\cdot)$ is the multivariate gamma function [10] defined as

$$\Gamma_k(n) = \pi^{k(k-1)/4} \prod_{j=1}^{k} \Gamma\left(n - \frac{1}{2}(j-1)\right). \qquad (1)$$

2. PROBLEM STATEMENT

Consider the following problem. Given two estimates $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$ of the true state vector $\mathbf{x}_0$ with their covariance matrices P_{11} and P_{22} respectively, we seek a fusion scheme that combines the available information and provides an estimate $\hat{\mathbf{x}}_0$

with minimum mean square error. We use P_0 to denote the covariance of $\hat{\mathbf{x}}_0$.

We combine $\mathbf{x}_1$ and $\mathbf{x}_2$ to form a single random vector, whose corresponding covariance matrix is

$$P_x = \begin{bmatrix} P_{11} & P_{12} \\ P_{12}^T & P_{22} \end{bmatrix}. \tag{2}$$

If we know P_{12}, things will be very easy. We can start by constructing an unbiased estimator in the form

$$\hat{\mathbf{x}}_0 = W_1\hat{\mathbf{x}}_1 + W_2\hat{\mathbf{x}}_2, \tag{3}$$

where $W_1 + W_2 = I$. The minimization of the mean square error is equivalent to the minimization of $\text{tr}(P_0)$, which can be expressed as

$$\begin{aligned}\text{tr}(P_0) = \text{tr}\left(W_1P_{11}W_1^T + W_1P_{12}W_2^T \right. \\ \left. + W_2P_{12}^TW_1^T + W_2P_{22}W_2^T\right).\end{aligned}$$

This can be carried out by using the method of Lagrange multipliers. Let Λ be the matrix of Lagrange multipliers. Define now L as

$$L = \text{tr}\left(P_0\right) + \Lambda\left(W_1 + W_2 - I\right).$$

Then, using the identity

$$\frac{\partial \text{tr}(XAX^T)}{\partial X} = XA + XA^T,$$

we obtain the stationary points by the following equations:

$$\begin{aligned}\frac{\partial L}{\partial W_1} &= 2W_1P_{11} + 2W_2P_{12}^T + \Lambda = 0 \\ \frac{\partial L}{\partial W_2} &= 2W_2P_{22} + 2W_1P_{12} + \Lambda = 0 \\ \frac{\partial L}{\partial \Lambda} &= W_1 + W_2 - I = 0.\end{aligned}$$

Combining all of the three equations, we find that

$$W_1 = (P_{22} - P_{12}^T)(P_{11} - P_{12} - P_{12}^T + P_{22})^{-1} \tag{4}$$

$$W_2 = (P_{11} - P_{12})(P_{11} - P_{12} - P_{12}^T + P_{22})^{-1}, \tag{5}$$

which are the weights for optimal fusion in the mean square error sense. When we substitute (4) back into (3), we obtain

$$\begin{aligned}\hat{\mathbf{x}}_{mmse} &= W_1\hat{\mathbf{x}}_1 + W_2\hat{\mathbf{x}}_2. \\ &= (P_{22} - P_{12}^T)(P_{11} - P_{12} - P_{12}^T + P_{22})^{-1}\hat{\mathbf{x}}_1 \\ &\quad + (P_{11} - P_{12})(P_{11} - P_{12} - P_{12}^T + P_{22})^{-1}\hat{\mathbf{x}}_2.\end{aligned} \tag{6}$$

However, in many situations we do not have information about P_{12}. For example, in a sensor network, when two nodes have their measurements and we want to fuse them, we often do not know their cross-covariance. In [5], the authors have proposed the covariance intersection method to minimize the upper bound for all possible P_{12} by a convex combination of the covariances, i.e.,

$$\begin{aligned}P_0^{-1} &= \omega P_{11}^{-1} + (1-\omega)P_{22}^{-1} \\ P_0^{-1}\hat{\mathbf{x}}_0 &= \omega P_{11}^{-1}\hat{\mathbf{x}}_1 + (1-\omega)P_{22}^{-1}\hat{\mathbf{x}}_2,\end{aligned}$$

where $\omega \in [0, 1]$. 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 [6, 7, 8, 9]. One of them sets ω according to [7]

$$\omega = \frac{|P_{22}|}{|P_{11}| + |P_{22}|}. \tag{7}$$

We will use it in the sequel for comparison with our algorithm.

3. THE MINIMUM MEAN SQUARE ERROR ESTIMATOR

Our strategy to solving the problem is to put it into a Bayesian framework. We assume that P_x has a prior and that the prior is the Wishart distribution. The Wishart distribution is any of a family of probability distributions defined over symmetric, nonnegative-definite matrix-valued random matrices. These distributions are of great importance in the estimation of covariance matrices in multivariate statistics [11]. The Wishart distribution is defined as follows.

The $k \times k$ random matrix A is said to have a Wishart distribution if its probability distribution function (pdf) is given by

$$p(A) = \frac{|A|^{\frac{n-k-1}{2}}\exp\left(-\frac{1}{2}\text{tr}(\Sigma^{-1}A)\right)}{2^{\frac{kn}{2}}|\Sigma|^{\frac{n}{2}}\Gamma_k(\frac{n}{2})},$$

where Σ is a positive definite matrix, $n \geq k$ is the degree of freedom, and Γ_k is defined by (1). We use $\mathcal{W}_k(n, \Sigma)$ to denote the Wishart distribution. We will omit k and write simply $\mathcal{W}(n, \Sigma)$ if the size of the matrix is obvious from the context.

The Wishart distribution is strongly related to the multivariate normal distribution. Suppose X is an $n \times k$ matrix, the rows of which have k-variate normal distribution with zero mean and covariance matrix Σ, denoted as $\mathcal{N}(0, \Sigma)$. Then the $k \times k$ random matrix $A = X^TX$ has a Wishart distribution, i.e., $\mathcal{W}(n, \Sigma)$. This property makes the generation of Wishart random matrices easy.

In our problem, we know P_{11} and P_{22}. To fuse the data, we would like to have information of P_{12} conditioned on P_{11} and P_{22}. We express this by the conditional

$$p(P_{12}|P_{11}, P_{22}) = \frac{p(P_x)}{p(P_{11}, P_{22})}.$$

Since P_{11} and P_{22} are known, our weight matrices W_1 and W_2, and therefore the MMSE estimator $\hat{\mathbf{x}}_{mmse}$, are uniquely determined by P_{12} as in (6). We think of it as a function of the matrix variable P_{12} and use $f(P_{12})$ to denote it. Note that P_{12} cannot be an arbitrary matrix. According to Theorem 7.7.6 in [12], $P_{22} - P_{12}^T P_{11}^{-1} P_{12}$ must be positive definite because P_x is a positive definite matrix. Therefore, we express our MMSE estimator by

$$\hat{\mathbf{x}}_{mmse} = \int_{P_{22} > P_{12}^T P_{11}^{-1} P_{12}} f(P_{12}) p(P_{12}|P_{11}, P_{22}) dP_{12}.$$

Unfortunately, the above integral is computationally intractable.

In order to approximate the integral, we have to resort to the Monte Carlo method. We sample M independent random matrices, $P_{12}^{(m)} \sim p(P_{12}|P_{11}, P_{22})$ for $m = 1, \cdots, M$. Then the Monte Carlo method approximates $\hat{\mathbf{x}}_{mmse}$ by the following expression:

$$\hat{\mathbf{x}}_{mmse} \approx \frac{1}{M} \sum_{m=1}^{M} f(P_{12}^{(m)}).$$

An immediate question is how we can sample from the conditional distribution $p(P_{12}|P_{11}, P_{22})$. We answer the question in the next section.

4. CONDITIONAL DISTRIBUTION OF THE OFF-DIAGONAL BLOCK SUBMATRICES

Suppose that the random matrix A is distributed according to $\mathcal{W}(n, \Sigma)$. Let the partitions of the two positive definite matrices A and Σ be denoted by

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^T & A_{22} \end{bmatrix} \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22} \end{bmatrix}. \tag{8}$$

Here we assume $\Sigma_{12} = O$. Recall that our objective is to derive the expression for $p(A_{12}|A_{11}, A_{22})$. We will need two properties of the Wishart distribution in our derivation [11].

Lemma 1. *Let A and Σ be partitioned into l and $k-l$ rows and columns as shown in (8). If A is distributed according to $\mathcal{W}_l(n, \Sigma)$, then A_{11} is distributed according to $\mathcal{W}_{k-l}(n, \Sigma_{11})$.*

Lemma 2. *If $\Sigma_{12} = O$ and A is distributed according to $\mathcal{W}(n, \Sigma)$, then A_{11} and A_{22} are independently distributed.*

Lemma 1 provides the marginal distributions of $p(A_{11})$ and $p(A_{22})$ (they are $\mathcal{W}(n, \Sigma_{11})$ and $\mathcal{W}(n, \Sigma_{22})$, respectively). Lemma 2 maintains that A_{11} and A_{22} are independent. Therefore, $p(A_{12}|A_{11}, A_{22})$ becomes

$$\begin{aligned} p(A_{12}|A_{11}, A_{22}) &= \frac{p(A)}{p(A_{11}, A_{22})} \\ &= \frac{p(A)}{p(A_{11})p(A_{22})}. \end{aligned}$$

With a little algebraic manipulation, we have

$$\begin{aligned} & p(A_{12}|A_{11}, A_{22}) \\ =& Z \cdot |A|^{\frac{n-k-1}{2}} \\ =& Z \cdot \left(|A_{11}||A_{22} - A_{12}^T A_{11}^{-1} A_{12}|\right)^{\frac{n-k-1}{2}} \\ =& Z \cdot \left(|A_{11}A_{22}||I - A_{22}^{-1} A_{12}^T A_{11}^{-1} A_{12}|\right)^{\frac{n-k-1}{2}}, \end{aligned} \tag{9}$$

where $n > k-1$, and the constant Z equals

$$\begin{aligned} Z &= \frac{\left(\prod_{i=1}^{\frac{k}{2}} \Gamma(\frac{1}{2}(n+1-i))\right)^2}{\prod_{j=1}^{k} \Gamma(\frac{1}{2}(n+1-j))} \cdot \frac{1}{\pi^{\frac{k^2}{8}} |A_{11}A_{22}|^{\frac{n-\frac{k}{2}-1}{2}}} \\ &= \frac{\prod_{i=1}^{\frac{k}{2}} \Gamma(\frac{1}{2}(n+1-i))}{\prod_{j=1+\frac{k}{2}}^{k} \Gamma(\frac{1}{2}(n+1-j))} \cdot \frac{1}{\pi^{\frac{k^2}{8}} |A_{11}A_{22}|^{\frac{n-\frac{k}{2}-1}{2}}} \\ &= \frac{\prod_{i=1}^{\frac{k}{2}} \Gamma(\frac{1}{2}(n+1-i))}{\prod_{j=1}^{\frac{k}{2}} \Gamma(\frac{1}{2}(n-\frac{k}{2}+1-j))} \cdot \frac{1}{\pi^{\frac{k^2}{8}} |A_{11}A_{22}|^{\frac{n-\frac{k}{2}-1}{2}}} \\ &= \frac{\Gamma_{\frac{k}{2}}(\frac{n}{2})}{\Gamma_{\frac{k}{2}}(\frac{1}{2}(n-\frac{k}{2}))} \cdot \frac{1}{\pi^{\frac{k^2}{8}} |A_{11}A_{22}|^{\frac{n-\frac{k}{2}-1}{2}}}. \end{aligned}$$

The above distribution is the inverted matrix variate t-distribution whose definition is as follows [13]:

Definition 1. *The random matrix $T \in \mathbb{R}^{k \times m}$ is said to have an inverted matrix variate t-distribution with parameters $M \in \mathbb{R}^{k \times m}$, $\Sigma \in \mathbb{R}^{k \times k}$, $\Omega \in \mathbb{R}^{m \times m}$ and n if its pdf is given by*

$$\begin{aligned} p(T) = & \frac{\Gamma_k(\frac{1}{2}(n+m+k-1))}{\pi^{\frac{mk}{2}} \Gamma_k(\frac{1}{2}(n+k-1))} |\Sigma|^{-\frac{m}{2}} |\Omega|^{-\frac{k}{2}} \\ & |I - \Sigma^{-1}(T-M)\Omega^{-1}(T-M)^T|^{\frac{n-2}{2}}, \end{aligned}$$

where $\Omega > 0$, $\Sigma > 0$, $n > 0$ and $I - \Sigma^{-1}(T-M)\Omega^{-1}(T-M)^T > 0$. We denote this by $T \sim \mathcal{IT}_{k,m}(n, M, \Sigma, \Omega)$.

For our case in (9), it is not difficult to obtain that

$$A_{12}^T | A_{11}, A_{22} \sim \mathcal{IT}_{\frac{k}{2}, \frac{k}{2}}(n-k+1, O, A_{22}, A_{11}). \tag{10}$$

For sampling from the inverted matrix variate t-distribution, we use the following lemma [13]:

Lemma 3. *Let $S \sim \mathcal{W}_k(n+k-1, I_k)$ and $X \sim \mathcal{N}_{k,m}(0, I_k \otimes I_m)$ be independently distributed. For $M \in \mathbb{R}^{k \times m}$, define*

$$T = \Sigma^{\frac{1}{2}} (S + XX^T)^{-\frac{1}{2}} X \Omega^{\frac{1}{2}} + M,$$

where $S + XX^T = (S + XX^T)^{\frac{1}{2}}((S + XX^T)^{\frac{1}{2}})^T$ *and* $\Sigma^{\frac{1}{2}}$ *and* $\Omega^{\frac{1}{2}}$ *are the symmetric square roots of the positive definite matrices* Σ *and* Ω*, respectively. Then,* $T \sim \mathcal{IT}_{k,m}(n, M, \Sigma, \Omega)$.

According to Lemma 3, the following theorem follows immediately.

Theorem 1. *Let the random matrices* S *and* X *be distributed according to* $S \sim \mathcal{W}_{\frac{k}{2}}(n, I_{\frac{k}{2}})$ *and* $X \sim \mathcal{N}_{\frac{k}{2},\frac{k}{2}}(0, I_{\frac{k}{2}} \otimes I_{\frac{k}{2}})$. *If*

$$A_{12}^T = (A_{22})^{\frac{1}{2}}(S + XX^T)^{-\frac{1}{2}}X(A_{11})^{\frac{1}{2}},$$

then $A_{12}^T \sim p\,(A_{12}|A_{11}, A_{22})$.

5. NUMERICAL EXPERIMENTS

In this section, we construct a model to test our algorithm. Suppose the variable to be estimated is $\mathbf{x}_0$ and that it has the normal distribution $\mathcal{N}(\boldsymbol{\mu}_0, \Sigma_0)$. We have two available measurements $\mathbf{x}_1$ and $\mathbf{x}_2$, both of which have the condition distributions $\mathcal{N}(\mathbf{x}_0, \Sigma_1)$ and $\mathcal{N}(\mathbf{x}_0, \Sigma_2)$, respectively. We can consider $\mathbf{x}_1$ and $\mathbf{x}_2$ to be measurements as well as estimates since we shall let $\hat{\mathbf{x}}_1 = \mathbf{x}_1$ if we make estimation only based on $\mathbf{x}_1$. If we concatenate $\mathbf{x}_1$ and $\mathbf{x}_2$ into one vector, the distribution of the vector conditioned on $\mathbf{x}_0$ is

$$\begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{bmatrix}\Bigg|\,\mathbf{x}_0 \sim \mathcal{N}\left(\begin{bmatrix}\mathbf{x}_0\\ \mathbf{x}_0\end{bmatrix}, \begin{bmatrix}\Sigma_1, O\\ O, \Sigma_2\end{bmatrix}\right).$$

Furthermore, we can easily obtain its marginal distribution, or

$$\begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}\boldsymbol{\mu}_0\\ \boldsymbol{\mu}_0\end{bmatrix}, \begin{bmatrix}\Sigma_1 + \Sigma_0, & \Sigma_0\\ \Sigma_0, & \Sigma_2 + \Sigma_0\end{bmatrix}\right). \tag{11}$$

Note that the covariance matrix in (11) is just the one in (2), which is of our interest. So $P_{11} = \Sigma_1 + \Sigma_0$ and $P_{22} = \Sigma_2 + \Sigma_0$, and they are both known exactly. On the other hand, $P_{12} = \Sigma_0$ is unknown.

To generate the data for our numerical experiment, we first draw Σ_0, Σ_1, and Σ_2 from $\mathcal{W}_{\frac{k}{2}}(n, \sigma_0^2 I_{\frac{k}{2}})$, $\mathcal{W}_{\frac{k}{2}}(n, \sigma^2 I_{\frac{k}{2}})$ and $\mathcal{W}_{\frac{k}{2}}(n, \sigma^2 I_{\frac{k}{2}})$, respectively. We set the degree of freedom $n = 4$ and use $k = 4$. Then we generate the true value $\mathbf{x}_0$ by sampling from $\mathcal{N}(\boldsymbol{\mu}_0, \Sigma_0)$, where $\boldsymbol{\mu}_0 = \mathbf{0}$. Similarly we generate the measurements $\mathbf{x}_1$ and $\mathbf{x}_2$ from $\mathcal{N}(\mathbf{x}_0, \Sigma_1)$ and $\mathcal{N}(\mathbf{x}_0, \Sigma_2)$, respectively. As stated above, the marginal covariance matrix (11) of the combined measurements becomes

$$P_x = \begin{bmatrix}\Sigma_1 + \Sigma_0, & \Sigma_0\\ \Sigma_0, & \Sigma_2 + \Sigma_0\end{bmatrix}. \tag{12}$$

Now we have all the data we need for testing and comparing the estimators. For comparison, we use two other estimators, the optimal estimator (6) with all the available information and the fast covariance intersection method from [7]. For each configuration, we ran 200 tests. In the proposed algorithm, for approximating $\hat{\mathbf{x}}_{mmse}$ we generated 500 samples. In the legend, we use *optimal, Bayesian* and *CI* to indicate the optimal method, the proposed method, and the fast covariance intersection method, respectively.

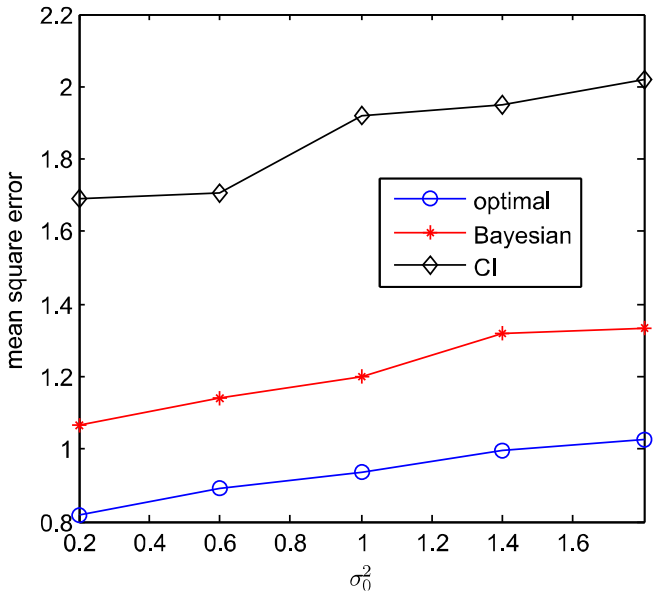

Fig. 1. The mean square error of the three estimators for different σ_0^2 ($\sigma^2 = 1$).

Figure 1 shows the mean square error of the three estimators for different σ_0^2. From (12), we can see that the cross-covariance is determined by σ_0^2. Roughly speaking, the 'larger' the matrix values are, the 'more' the two estimates $\mathbf{x}_1$ and $\mathbf{x}_2$ relate to each other. From Fig. 1, we see that the optimal method works best as expected. The proposed Bayesian algorithm is about 20 percent worse than the optimal one, but much better than the covariance intersection estimator. Figure 2 shows the mean square error for different values of σ^2. Unlike σ_0^2, σ^2 has no effect on the cross-covariance. We have similar performance as shown in the first figure. Again, the optimal estimator is the best, and the proposed estimator has performance that is close to that of the optimal estimator and much better than the performance of the covariance intersection method.

We would like to point out that, strictly speaking, the covariance matrix P_x in the simulations does not have the Wishart distribution, as the off-diagonal block matrix is always symmetric. Nevertheless, our estimator still performs well.

6. DISCUSSION

In this article, we propose a Bayesian approach to solve the data fusion problem when the cross-covariance between two

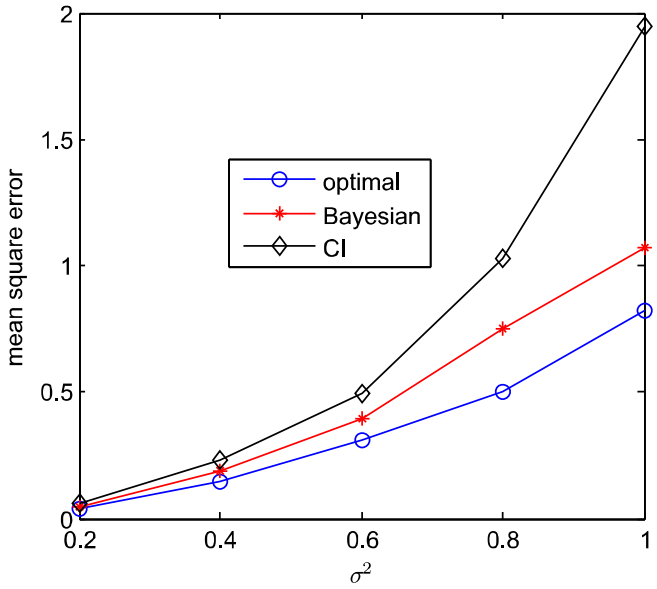

Fig. 2. The mean square error of three estimators for different σ^2 ($\sigma_0^2 = 0.2$).

estimates is not available. We first assume that the prior of the covariance matrix is the Wishart distribution. Because we know the covariance of each estimate, which is the diagonal block of the covariance matrix, we can obtain the conditional distribution of the off-diagonal block. The distribution of this block is the inverted matrix variate t-distribution. We also show how to sample from this distribution. As a result, we can use the Monte Carlo method to compute the MMSE estimator. Numerical experiments show that the performance of our method is much better than that of the covariance intersection method. Another advantage of our algorithm is that under the Bayesian framework, we are able to adjust the hyperparameter of the prior according to the information available, making the algorithm more robust in some special cases.

The curious reader may wonder why we assume the parameter Σ of the prior Wishart distribution $\mathcal{W}(n, \Sigma)$ to be a block diagonal matrix. The reason is that by doing so, the diagonal blocks of the resulting covariance matrix are independent from each other. Otherwise, the joint distribution of the diagonal blocks are very complicated making the derivation of the conditional distribution of the off-diagonal blocks very difficult, if not impossible. We can see that the Wishart distribution with block diagonal parameter matrix Σ is still general enough to allow for good performance.

In this work, we only consider the fusion problem with two nodes. Admittedly the application of the proposed algorithm is confined to the situations with two nodes, while we probably need to deal with the fusion problem with information from several nodes in usual sensor network application. This is going to be the direction of our future efforts.

7. ACKNOWLEDGEMENTS

This work has been supported by the National Science Foundation under Award CCF-1018323 and the Office of Naval Research under Award N00014-09-1-1154.

8. REFERENCES

[1] J. Hu, L. Xie, and C. Zhang, "Diffusion Kalman filtering based on covariance intersection," *IEEE Transactions on Signal Processing*, vol. 60, no. 2, pp. 891–902, 2012.

[2] 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.

[3] R. Olfati-Saber, "Distributed Kalman filtering for sensor networks," in *the Proceedings of the IEEE Decision and Control Conference*, 2007, pp. 5492–5498.

[4] R. Olfati-Saber, "Distributed Kalman filter with embedded consensus filters," in *the Proceedings of the IEEE Conference on Decision and Control and the European Control Conference*, 2005, pp. 8179–8184.

[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] W. Niehsen, "Information fusion based on fast covariance intersection filtering," in *the Proceedings of the 5th IEEE International Conference on Information Fusion*, 2002, vol. 2, pp. 901–904.

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

[8] Y. Wang and X. R. Li, "A fast and fault-tolerant convex combination fusion algorithm under unknown cross-correlation," in *the Proceedings of the 12th IEEE International Conference on Information Fusion*, 2009, pp. 571–578.

[9] Y. Wang and X. R. Li, "Distributed estimation fusion with unavailable cross-correlation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 1, pp. 259–278, 2012.

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

[11] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, Wiley Series in Probability and Statistics, 2003.

[12] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1990.

[13] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions*, vol. 104, Chapman & Hall/CRC, 2000.