Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 910971, 6 pages
http://dx.doi.org/10.1155/2014/910971

Research Article

Minimax Robust Optimal Estimation Fusion for Distributed Multisensor Systems with a Relative Entropy Uncertainty

Hua Li and Jie Zhou

College of Mathematics, Sichuan University, Chengdu, Sichuan 610064, China

Correspondence should be addressed to Jie Zhou; jzhou@scu.edu.cn

Received 20 January 2014; Accepted 17 March 2014; Published 12 May 2014

Academic Editor: Ezzat G. Bakhoum

This paper considers the robust estimation fusion problem for distributed multisensor systems with uncertain correlations of local estimation errors. For an uncertain class characterized by the Kullback-Leibler (KL) divergence from the actual model to nominal model of local estimation error covariance, the robust estimation fusion problem is formulated to find a linear minimum variance unbiased estimator for the least favorable model. It is proved that the optimal fuser under nominal correlation model is robust while the estimation error has a relative entropy uncertainty.

1. Introduction

During the past decades, multisensor systems have received much attention in many applications, such as signal processing, communication, target tracking, and remote sensing [1–7]. In distributed estimation fusion frame, each sensor observes some local information of the physical environment, estimates the state or parameter of system in terms of some optimization criterion, and then transmits it to fusion center. The fusion center fuses the data from local sensors and obtains a better estimate. Generally, distributed fusion may have a less computational burden, lower communication rates, and higher survivability and be more flexible and reliable than centralized fusion which processes directly observation data from local sensors.

There is much work on developing distributed estimation fusion algorithms. In [8, 9], a distributed Kalman filtering fusion formula was presented which has the same estimation performance as centralized Kalman filtering fusion. For the general systems with known auto- and cross-correlations of estimation errors from local sensors, in [6, 10–12], the optimal linear estimation fusion formulas were proposed in the sense of linear minimum variance (LMV). In practice, the cross-correlations of estimation errors among the sensors may be completely or partially unknown. Some fusion methods have been developed for the systems with various uncertain correlations. The simple convex combination approach ignores the cross-correlations. The covariance intersection (CI) method linearly combines the local estimates and considers a conservative estimate of the estimation error covariance matrix (see [13–16]). An information-theoretic justification for CI was presented in [16] to minimize the Chernoff information. A robust estimation fuser was developed in [17] by finding the Chebyshev center, that is, center of the minimum radius ball enclosing the intersection of ellipsoid sets. For a class of estimation error covariance matrices with norm-bounded uncertainties, a robust fusion model to minimize the worst-case mean square error (MSE) was proposed in [18]. In [19, 20], a generalized convex combination method was discussed. Recently, by employing random matrix to describe the uncertainty of cross-correlation, a chance-constrained programming approach was presented in [21]. In [22], a robust estimation problem was addressed for a linear model in which the unknown parameter vector is norm-bounded and noise covariance matrix is uncertain with a special structure. It was to get a linear estimator which minimizes the worst-case MSE over all vectors and noise covariance matrices in a specified uncertain region.

This paper considers the robust estimation fusion approach for distributed multisensor systems with uncertain correlations among local estimation errors. Assuming the true joint probability density of local estimators belonging to a neighborhood of the nominal model, the robust estimation fusion can be modeled as a minimax problem that minimizes the variance for the least favorable model in an uncertain class. In this paper, the neighborhood is specified by the Kullback-Leibler (KL) divergence. It is worth noting that the use of KL divergence is rather natural as a metric for model mismatch in the statistics and information societies. From an information geometric viewpoint, it was argued in [23] that the KL divergence is a natural geometric "distance" between systems.

The remainder of the paper is organized as follows. In Section 2, we formulate the robust distributed estimation fusion as a min-max optimization problem. Then, we convert it equivalently into a min-min optimization problem by Lagrangian duality method in Section 3. Moreover, in Section 4, we give the analytical solution of original min-max problem and conclude that the linear minimum variance unbiased estimation (LMVUE) fusion is robust for the least favorable statistical model. Section 5 gives a conclusion.

Notations $\mathbb{R}^n$ and $\mathbb{R}^{m\times n}$ stand for the set of all n-dimensional real vectors and the set of all m-by-n real matrices, respectively. For a matrix A, A^T, $A^\dagger$, and $\|A\|$ denote the transpose, Moore-Penrose generalized inverse, and spectral norm of A, respectively. $A \succ 0$ means A is symmetric and positive definite . The symbol I represents the identity matrix with appropriate dimension. The expectation of a random variable (or random vector) is denoted by $\mathbb{E}[\cdot]$.

2. Problem Formulation

Consider an l-sensor distributed system in which $x \in \mathbb{R}^n$ is the state or parameter to be estimated, and $\widehat{x}^{(1)},\dots,\widehat{x}^{(l)}$ are local unbiased estimates of x available at fusion center. The relationship between the local unbiased estimates and the state or parameter can be formulated as follows:

$$y = Ax + \varepsilon, \tag{1}$$

where

$$y = \begin{pmatrix} \widehat{x}^{(1)} \\ \vdots \\ \widehat{x}^{(l)} \end{pmatrix} \in \mathbb{R}^{nl} \tag{2}$$

is augmented local unbiased estimates and can be viewed as the observation at fusion center, $A = [I,\dots,I]^T \in \mathbb{R}^{nl\times n}$, and $\varepsilon \in \mathbb{R}^{nl}$ is the augmented error vectors of all local estimates. It is clear that ε has mean zero and covariance matrix

$$V = \begin{pmatrix} V^{(11)} & \cdots & V^{(1l)} \\ \vdots & & \vdots \\ V^{(l1)} & \cdots & V^{(ll)} \end{pmatrix}, \tag{3}$$

where $V^{(ij)} = \mathbb{E}[(\widehat{x}^{(i)} - x)(\widehat{x}^{(j)} - x)^T]$ for $i, j = 1,\dots,l$.

In order to find an unbiased linear estimation fusion $\widehat{x} = B + W^T y$, where B and W are compatible vector and matrix, respectively, a necessary and sufficient condition is $B = 0$, and $A^T W = I$. Therefore, the fused estimate becomes $\widehat{x} = W^T y$.

When the error covariance matrix V is accurately given, a general linear fusion result is presented in [6, 11], which is optimal in the sense of LMV; that is, it has the LMV among all linear unbiased estimation fusion rules. In this case, the LMVUE fusion is uniquely given with probability one as follows:

$$\widehat{x} = \frac{1}{l} A^T \left(I - V(PVP)^\dagger\right) y, \tag{4}$$

where $P = I - AA^\dagger$ is an orthogonal projector (i.e., symmetric and idempotent matrix). Moreover, if V is nonsingular, then the optimal estimation fusion becomes

$$\widehat{x} = \left(A^T V^{-1} A\right)^{-1} A^T V^{-1} y. \tag{5}$$

However, the estimation error covariance matrix cannot be accurately obtained in practice. Denote the actual and nominal error covariance matrices by V and $\widehat{V}$, respectively. For measurement model (1), we assume that the noise ε has the nl-dimensional Gaussian distribution with zero mean. It is well known that the Gaussianity assumption is reasonable in many practical applications. The actual and nominal probability density functions of observation y are $N(Ax, V)$ and $N(Ax, \widehat{V})$, respectively, where the matrices V and $\widehat{V}$ are assumed to be invertible. Therefore, the probability density functions of the actual and nominal models are, respectively, given by

$$\begin{aligned} p(y) &= \frac{1}{(2\pi)^{nl/2}(\det V)^{1/2}} \\ &\quad \cdot \exp\left(-(y - Ax)^T V^{-1}(y - Ax)/2\right), \\ \widehat{p}(y) &= \frac{1}{(2\pi)^{nl/2}\left(\det \widehat{V}\right)^{1/2}} \\ &\quad \cdot \exp\left(-(y - Ax)^T \widehat{V}^{-1}(y - Ax)/2\right). \end{aligned} \tag{6}$$

In order to measure the "distance" between the two models, we adopt the KL divergence from the actual model to nominal model:

$$\begin{aligned} D\left(V, \widehat{V}\right) &\triangleq 2\mathrm{KL}\left(p(y), \widehat{p}(y)\right) \\ &= 2\int p(y) \ln\left(\frac{p(y)}{\widehat{p}(y)}\right) dy \\ &= \operatorname{tr}\left(\widehat{V}^{-1}V - I\right) - \ln\det\left(\widehat{V}^{-1}V\right). \end{aligned} \tag{7}$$

Although it is not a conventional distance, since it is not symmetric and does not obey the triangle inequality, the KL divergence satisfies $D(V, \widehat{V}) \geq 0$ with equality if and only if $\widehat{V} = V$.

Let

$$\mathscr{C} = \left\{V \succ 0 : D\left(V, \widehat{V}\right) \leq c\right\} \tag{8}$$

denote the set of all possible estimation error covariance matrices such that the KL divergence from corresponding model to actual model is less than or equal to $c \geq 0$. Note that $\mathscr{C}$ is a convex set since $D(V,\widehat{V})$ is a convex function of V. The minimax robust estimation fusion problem is to seek for an optimal weighting matrix W^* to minimize the MSE in the worst case. It can be formulated as the following minimax problem:

$$W^* = \arg\min_{W\in\mathscr{W}}\max_{V\in\mathscr{C}}\mathbb{E}\|\hat{x}-x\|^2, \tag{9}$$

where $\mathscr{W} = \{W \in \mathbb{R}^{nl\times n} : A^T W = I\}$. From

$$\begin{aligned}\mathbb{E}\|\hat{x}-x\|^2 &= \mathbb{E}\left[\left(W^T y - x\right)^T\left(W^T y - x\right)\right]\\ &= \mathbb{E}\left[\left(W^T y - W^T Ax\right)^T\left(W^T y - W^T Ax\right)\right]\\ &= \mathbb{E}\left[(y-Ax)^T WW^T (y-Ax)\right]\\ &= \mathbb{E}\left[\operatorname{tr}\left(W^T (y-Ax)\ (y-Ax)^T W\right)\right]\\ &= \operatorname{tr}\left(W^T V W\right),\end{aligned} \tag{10}$$

the optimization problem (9) is equivalent to the following optimization problem:

$$W^* = \arg\min_{W\in\mathscr{W}}\max_{V\in\mathscr{C}}\operatorname{tr}\left(W^T V W\right). \tag{11}$$

3. Problem Conversion

Consider the inner maximization problem in (11); that is,

$$\begin{aligned}&\max_{V\succ 0} \quad \operatorname{tr}\left(W^T V W\right)\\ &\text{s.t.} \quad D\left(V,\widehat{V}\right) \le c.\end{aligned} \tag{12}$$

To find the optimal solution, we construct the Lagrangian function

$$\begin{aligned}L(W,V,\lambda) = {}& \operatorname{tr}\left(W^T V W\right)\\ &+ \lambda\left(c - \operatorname{tr}\left(\widehat{V}^{-1}V - I\right) + \ln\det\left(\widehat{V}^{-1}V\right)\right),\end{aligned} \tag{13}$$

where $\lambda \ge 0$ is the Lagrange multiplier associated with the inequality constraint. Let $G(W,\lambda)$ be the Lagrange dual function $\sup_{V\succ 0} L(W,V,\lambda)$ while the latter is finite.

Lemma 1. *For problem* (12), *the following strong Lagrangian duality identity holds:*

$$\max_{V\in\mathscr{C}}\operatorname{tr}\left(W^T V W\right) = \min_{\lambda > \|W^T\widehat{V}W\|} G(W,\lambda). \tag{14}$$

Proof. The function $L(W,V,\lambda)$ is the sum of a linear function and a logarithm of determinant which is concave, so $L(W,V,\lambda)$ is a concave function of V.

If $0 \le \lambda \le \|W^T\widehat{V}W\|$, then, noticing the invertibility of $W^T\widehat{V}W$ because W is full column rank from $A^T W = I$, we have the following eigenvalue decomposition:

$$\widehat{V}^{1/2}WW^T\widehat{V}^{1/2} = Q^T\Lambda Q, \tag{15}$$

where Q is an orthogonal matrix and

$$\Lambda = \operatorname{diag}\left(\eta_1,\dots,\eta_n,0,\dots,0\right) \in \mathbb{R}^{nl\times nl} \tag{16}$$

with $\eta_1 \ge \cdots \ge \eta_n > 0$ and $\eta_1 \ge \lambda$. Let

$$\begin{aligned}\Sigma &= \operatorname{diag}\left(\mu,1,\dots,1\right) \in \mathbb{R}^{nl\times nl}, \quad \mu > 0,\\ V &= \widehat{V}^{1/2}Q^T\Sigma Q\widehat{V}^{1/2} \succ 0.\end{aligned} \tag{17}$$

Then,

$$\begin{aligned}L(W,V,\lambda) = {}& \operatorname{tr}\left(\left(\widehat{V}^{1/2}WW^T\widehat{V}^{1/2} - \lambda I\right)\widehat{V}^{-1/2}V\widehat{V}^{-1/2}\right)\\ &+ \lambda\left(c + \operatorname{tr} I + \ln\det\left(\widehat{V}^{-1}V\right)\right)\\ = {}& \operatorname{tr}\left((\Lambda - \lambda I)\Sigma\right) + \lambda\left(c + nl + \ln\det(\Sigma)\right)\\ = {}& (\eta_1-\lambda)\mu + \sum_{i=2}^{n}(\eta_i - \lambda) + \lambda\left(c + nl + \ln\mu\right)\end{aligned} \tag{18}$$

can approach $+\infty$ while $\mu \to +\infty$.

If $\lambda > \|W^T\widehat{V}W\|$, then all eigenvalues of matrix $W^T\widehat{V}W$ or $\widehat{V}^{1/2}WW^T\widehat{V}^{1/2}$ are smaller than λ. So, $\lambda I - \widehat{V}^{1/2}WW^T\widehat{V}^{1/2} \succ 0$; that is, $\widehat{V}^{-1} - (1/\lambda)WW^T \succ 0$. Using the formulas for derivatives of traces and determinants of matrices (see, e.g., [24, pp. 178, 182]), we have

$$\begin{aligned}\frac{\partial\operatorname{tr}\left(W^T V W\right)}{\partial V} &= WW^T,\\ \frac{\partial\operatorname{tr}\left(\widehat{V}^{-1}V\right)}{\partial V} &= \widehat{V}^{-1},\\ \frac{\partial\ln\det\left(\widehat{V}^{-1}V\right)}{\partial V} &= \widehat{V}^{-1}\left(\left(\widehat{V}^{-1}V\right)^T\right)^{-1} = V^{-1}.\end{aligned} \tag{19}$$

Then,

$$\begin{aligned}\frac{\partial L(W,V,\lambda)}{\partial V} = {}& \frac{\partial\operatorname{tr}\left(W^T V W\right)}{\partial V}\\ &- \lambda\left(\frac{\partial\operatorname{tr}\left(\widehat{V}^{-1}V\right)}{\partial V} - \frac{\partial\ln\det\left(\widehat{V}^{-1}V\right)}{\partial V}\right)\\ = {}& WW^T - \lambda\left(\widehat{V}^{-1} - V^{-1}\right).\end{aligned} \tag{20}$$

Taking $\partial L(W,V,\lambda)/\partial V$ as zero, we can obtain the positive definite matrix maximizing $L(W,V,\lambda)$ as follows:

$$V^*(\lambda) = \left(\widehat{V}^{-1} - \frac{1}{\lambda}WW^T\right)^{-1}, \tag{21}$$

and $G(W,\lambda) = L(W,V^*(\lambda),\lambda) < +\infty$.

In summary, we obtain the Lagrange dual function

$$\sup_{V\succ 0} L(W,V,\lambda) = \begin{cases} G(W,\lambda), & \lambda > \left\|W^T\widehat{V}W\right\|,\\ +\infty, & \text{otherwise.}\end{cases} \tag{22}$$

Therefore, for the dual problem of problem (12), it is sufficient and necessary to consider only the case of $\lambda > \|W^T \widehat{V} W\|$.

Since the primal problem (12) involves the maximization of a linear function $\operatorname{tr}(W^T V W)$ over a bounded convex set $\mathscr{C}$, the optimal objective value is finite. In addition, there exists $V = \widehat{V} \succ 0$ with $D(V, \widehat{V}) = 0 < c$. Therefore, the Slater conditions are satisfied. From Proposition 6.4.3 in [25], strong Lagrangian duality holds. □

From Lemma 1, the original min-max problem (9) or (11) is converted equivalently into the following min-min problem:

$$\min_{W \in \mathscr{W}} \min_{\lambda > \|W^T \widehat{V} W\|} G(W, \lambda). \tag{23}$$

Furthermore, from (21), we have

$$\begin{gathered} W W^T = \lambda \left(\widehat{V}^{-1} - (V^*(\lambda))^{-1}\right), \\ \widehat{V}^{-1} V^*(\lambda) = \left(I - \frac{1}{\lambda} W W^T \widehat{V}\right)^{-1}. \end{gathered} \tag{24}$$

Then,

$$\begin{aligned} G(W, \lambda) = & \operatorname{tr}\left(W^T V^*(\lambda) W\right) \\ & + \lambda \left(c - \operatorname{tr}\left(\widehat{V}^{-1} V^*(\lambda) - I\right)\right) \\ & + \lambda \ln \det\left(\widehat{V}^{-1} V^*(\lambda)\right) \\ = & \lambda c + \lambda \ln \det\left(\widehat{V}^{-1} V^*(\lambda)\right) \\ = & \lambda c - \lambda \ln \det\left(I - \frac{1}{\lambda} W^T \widehat{V} W\right). \end{aligned} \tag{25}$$

4. Solution of the Robust Optimization Problem

The objective function $G(W, \lambda)$ of problem (23) is convex in (W, λ) since $L(W, V, \lambda)$ is convex in (W, λ) and $G(W, \lambda)$ is obtained by maximizing $L(W, V, \lambda)$ over V (see [25, Proposition 1.2.4c]). In addition, the set

$$\mathscr{D} = \left\{(W, \lambda) : A^T W = I, \ \left\|W^T \widehat{V} W\right\| < \lambda\right\} \tag{26}$$

is convex because it is the intersection of a linear constraint and a semidefinite programming (SDP) constraint. Consequently, the problem (23) is a convex optimization problem. Next, we pursue the analytical solution of problem (23).

Lemma 2. *If $M \succ 0$ and Q is an orthogonal projector with compatible dimensions, then*

$$(QMQ)^{\dagger} (QMQ) = Q, \tag{27}$$

$$\left(I - (QMQ)^{\dagger} M\right)(I - Q) = I - (QMQ)^{\dagger} M. \tag{28}$$

Proof. Using the properties of Moore-Penrose generalized inverse (see, e.g., [26]), we have

$$Q(QMQ)^{\dagger} = (QMQ)^{\dagger} Q = (QMQ)^{\dagger}. \tag{29}$$

Therefore, $(QMQ)^{\dagger}(QMQ)$ and $Q - (QMQ)^{\dagger}(QMQ)$ are orthogonal projectors. From the properties of orthogonal projector, we have

$$\begin{aligned} & \operatorname{rank}\left(Q - (QMQ)^{\dagger} (QMQ)\right) \\ & \quad = \operatorname{tr}\left(Q - (QMQ)^{\dagger} (QMQ)\right) \\ & \quad = \operatorname{tr}(Q) - \operatorname{tr}\left((QMQ)^{\dagger} (QMQ)\right) \\ & \quad = \operatorname{rank}(Q) - \operatorname{rank}\left((QMQ)^{\dagger} (QMQ)\right) \\ & \quad = \operatorname{rank}(Q) - \operatorname{rank}(QMQ) = 0. \end{aligned} \tag{30}$$

Equation (27) thus holds, and then (28) is direct. □

Theorem 3. *The optimal weighting matrix of the robust fusion problem* (9) *or* (11) *is given by*

$$W^* = \widehat{V}^{-1} A \left(A^T \widehat{V}^{-1} A\right)^{-1}. \tag{31}$$

Proof. The derivative of $G(W, \lambda)$ given by (25) with respect to λ is

$$\begin{aligned} \frac{\partial G(W, \lambda)}{\partial \lambda} = & \ c - \ln \det\left(I - \frac{1}{\lambda} W^T \widehat{V} W\right) \\ & - \lambda \cdot \frac{\partial}{\partial \lambda} \ln \det\left(I - \frac{1}{\lambda} W^T \widehat{V} W\right) \\ = & \ c - \ln \det\left(I - \frac{1}{\lambda} W^T \widehat{V} W\right) \\ & - \operatorname{tr}\left(\left(I - \frac{1}{\lambda} W^T \widehat{V} W\right)^{-1} \frac{1}{\lambda} W^T \widehat{V} W\right). \end{aligned} \tag{32}$$

Let $\eta_1 \geq \cdots \geq \eta_n$ be the eigenvalues of $W^T \widehat{V} W$. It is easy to show that

$$\frac{\partial G(W, \lambda)}{\partial \lambda} = 0 \tag{33}$$

is equivalent to

$$c - \sum_{i=1}^{n} \left(\frac{\eta_i}{\lambda - \eta_i} + \ln\left(1 - \frac{\eta_i}{\lambda}\right)\right) = 0. \tag{34}$$

The above equation has only one solution $\lambda^* > \eta_1 = \|W^T \widehat{V} W\|$ for any nonzero matrix W since

$$h(x) = \sum_{i=1}^{n} \left(\frac{\eta_i}{x - \eta_i} + \ln\left(1 - \frac{\eta_i}{x}\right)\right) \tag{35}$$

is a monotonically decreasing continuous function in $(\eta_1, +\infty)$ and

$$\lim_{x \to \eta_1^+} h(x) = +\infty, \qquad \lim_{x \to +\infty} h(x) = 0. \tag{36}$$

From $A^TW = I$ and the result on derivatives of determinants of matrices (see, e.g., [24, p. 182]), the derivative of function $G(W,\lambda)$ with respect to W is

$$\begin{aligned}\frac{\partial G(W,\lambda)}{\partial W} &= -\lambda\frac{\partial}{\partial W}\ln\det\left(W^T\left(AA^T-\frac{1}{\lambda}\widehat{V}\right)W\right)\\ &= -2\lambda\left(AA^T-\frac{1}{\lambda}\widehat{V}\right)W\left(W^T\left(AA^T-\frac{1}{\lambda}\widehat{V}\right)W\right)^{-1}\\ &= -2\lambda\left(A-\frac{1}{\lambda}\widehat{V}W\right)\left(I-\frac{1}{\lambda}W^T\widehat{V}W\right)^{-1}.\end{aligned} \quad (37)$$

The solution of the matrix equation $A^TW = I$ given by equality constraint in the optimization problem (23) can be expressed as

$$W = \frac{1}{l}A + PZ, \quad (38)$$

where $P = I - AA^{\dagger}$ and $Z \in \mathbb{R}^{nl\times n}$ is any matrix. Next, we seek the matrix Z to optimize the weighting matrix W. Redenote the objective function $G(W,\lambda)$ given by (25) as $G(Z,\lambda)$. From $PA = 0$, the derivative of $G(Z,\lambda)$ with respect to Z is

$$\begin{aligned}\frac{\partial G(Z,\lambda)}{\partial Z} &= P\frac{\partial G(W,\lambda)}{\partial W}\\ &= 2P\widehat{V}W\left(I-\frac{1}{\lambda}W^T\widehat{V}W\right)^{-1}.\end{aligned} \quad (39)$$

Thus, from the invertibility of $I - \frac{1}{\lambda}W^T\widehat{V}W$,

$$\frac{\partial G(Z,\lambda)}{\partial Z} = 0 \quad (40)$$

if and only if

$$P\widehat{V}\left(\frac{1}{l}A + PZ\right) = 0. \quad (41)$$

That is,

$$P\widehat{V}PZ = -\frac{1}{l}P\widehat{V}A. \quad (42)$$

Because $\widehat{V}$ is invertible, from (27), the solution of matrix equation (42) is

$$Z^* = -\frac{1}{l}\left(P\widehat{V}P\right)^{\dagger}P\widehat{V}A + (I-P)\xi, \quad (43)$$

where $\xi \in \mathbb{R}^{nl\times n}$ is any matrix. Substituting Z^* into (38) and using $P(I-P) = 0$, we have

$$W^* = \frac{1}{l}\left(I-\left(P\widehat{V}P\right)^{\dagger}\widehat{V}\right)A. \quad (44)$$

From the convexity of problem (23) and Proposition 4.7.1 in [25], the pair $(W^*,\lambda^*) \in \mathscr{D}$ is the global minimum of (23). Moreover, from the definition of P and (28), we can obtain

$$\begin{aligned}&\frac{1}{l}\left(I-\left(P\widehat{V}P\right)^{\dagger}\widehat{V}\right)AA^T\widehat{V}^{-1}A\\ &= \left(I-\left(P\widehat{V}P\right)^{\dagger}\widehat{V}\right)(I-P)\widehat{V}^{-1}A\\ &= \left(I-P-\left(P\widehat{V}P\right)^{\dagger}\widehat{V}+\left(P\widehat{V}P\right)^{\dagger}\widehat{V}P\right)\widehat{V}^{-1}A\\ &= \left(I-\left(P\widehat{V}P\right)^{\dagger}\widehat{V}\right)\widehat{V}^{-1}A\\ &= \widehat{V}^{-1}A-\left(P\widehat{V}P\right)^{\dagger}PA = \widehat{V}^{-1}A.\end{aligned} \quad (45)$$

As a result , by the invertibility of $A^T\widehat{V}^{-1}A$, we have

$$\frac{1}{l}\left(I-\left(P\widehat{V}P\right)^{\dagger}\widehat{V}\right)A = \widehat{V}^{-1}A\left(A^T\widehat{V}^{-1}A\right)^{-1}. \quad (46)$$

The proof of this theorem is completed. □

Remark 4. The optimal weighting matrix W^* given in Theorem 3 does not depend on the parameter c. However, it is worth noting that the optimum value of the objective function is increasing as c increases. In fact, the optimal Lagrange multiplier λ^* depends on c.

From Theorem 3, although the least favorable distribution given by covariance V^* is not the nominal distribution, the robust estimation fusion formula $(W^*)^T y$ is the same as the fusion formulas given by (4) and (5) which are derived from the nominal distribution. It concludes that the usual LMVUE fusion for a Gaussian nominal model is robust.

5. Conclusion

A minimax optimization problem is formulated to find the robust estimation fusion for multisensor systems with uncertain correlations of local estimation errors characterized by the relative entropy. By rigorous mathematical deduction, it has been proved that the usual LMVUE fusion for a Gaussian nominal model is robust no matter how far the KL divergence is from the actual model to nominal model. It is an interesting and significant discovery.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported in part by the National Key Basic Research Program of China under Grant 2013CB329405, the National Natural Science Foundation of China under Grants 61374027 and 61032001, and the Specialized Research Fund for the Doctoral Program of Higher Education under Grant 20130181110042.

References

[1] Y. Bar-Shalom, Ed., *Multitarget-Multisensor Tracking: Advanced Applications*, vol. 1, Artech House, Norwood, Mass, USA, 1990.

[2] Y. Bar-Shalom, *Multitarget-Multisensor Tracking: Applications and Advances*, vol. 2, Artech House, Norwood, Mass, USA, 1992.

[3] Y. Bar-Shalom and X. R. Li, *Multitarget-Multisensor Tracking: Principles and Techniques*, YBS Publishing, Storrs, Conn, USA, 1995.

[4] Y. Bar-Shalom and W. D. Blair, Eds., *Multitarget-Multisensor Tracking: Applications and Advances*, vol. 3, Artech House, Norwood, Mass, USA, 2000.

[5] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation: Theory, Algorithm and Software*, John Wiley & Sons, New York, NY, USA, 2001.

[6] Y. Zhu, *Multisensor Decision and Estimation Fusion*, Kluwer Academic Publishers, Boston, Mass, USA, 2003.

[7] M. E. Liggins, D. L. Hall, and J. Llinas, Eds., *Handbook of Multisensor Data Fusion: Theory and Practice*, CRC Press, 2nd edition, 2009.

[8] C. Y. Chong, K. C. Chang, and S. Mori, "Distributed tracking in distributed sensor networks," in *Proceedings of the American Control Conference*, pp. 1863–1868, Seattle, Wash, USA, June 1986.

[9] H. R. Hashemipour, S. Roy, and A. J. Laub, "Decentralized structures for parallel Kalman filtering," *IEEE Transactions on Automatic Control*, vol. 33, no. 1, pp. 88–94, 1988.

[10] Y. Bar-Shalom and L. Campo, "The effect of the common process noise on the two-sensor fused-track covariance," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 22, no. 6, pp. 803–805, 1986.

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

[12] S. Sun and Z. Deng, "Multi-sensor optimal information fusion Kalman filter," *Automatica*, vol. 40, no. 6, pp. 1017–1023, 2004.

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

[14] C. Y. Chong and S. Mori, "Convex combination and covariance intersection algorithms in distributed fusion," in *Proceedings of the 4th International Conference on Information Fusion*, vol. 1, pp. WeA2-11–WeA2-18, Montreal, Canada, August 2001.

[15] 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.

[16] M. Hurley, "An information theoretic justification for covariance intersection and its generalization," in *Proceedings of the 5th International Conference on Information Fusion*, pp. 505–511, Annapolis, Md, USA, 2002.

[17] D. Wu, J. Zhou, and X. Qu, "A robust estimation fusion with unknown cross-covariance in distributed systems," in *Proceedings of the 48th IEEE Conference on Decision and Control*, pp. 7603–7607, Shanghai, China, December 2009.

[18] X. Qu, J. Zhou, E. Song, and Y. Zhu, "Minimax robust optimal estimation fusion in distributed multisensor systems with uncertainties," *IEEE Signal Processing Letters*, vol. 17, no. 9, pp. 811–814, 2010.

[19] Y. Wang and X. R. Li, "Distributed estimation fusion under unknown cross-correlation: an analytic center approach," in *Proceedings of the 13th Conference on Information Fusion*, pp. 1–8, Edinburgh, Scotland, July 2010.

[20] 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.

[21] D. Wu, J. Zhou, A. Hu, and F. Li, "Robust distributed fusion for system with randomly uncertain sensor estimation error cross-covariance," *International Journal of Systems Science*, vol. 44, no. 7, pp. 1245–1252, 2013.

[22] Y. C. Eldar, "Minimax MSE estimation of deterministic parameters with noise covariance uncertainties," *IEEE Transactions on Signal Processing*, vol. 54, no. 1, pp. 138–145, 2006.

[23] S. Amari and H. Nagaoka, *Methods of Information Geometry*, American Mathematical Society, 2000.

[24] H. Lütkepohl, *Handbook of Matrices*, John Wiley & Sons, New York, NY, USA, 1996.

[25] D. P. Bertsekas, A. Nedić, and A. E. Ozdarglar, *Convex Analysis and Optimization*, Athena Scientific, Belmont, Mass, USA, 2003.

[26] A. Ben-Israel and T. N. E. Greville, *Generalized Inverses: Theory and Applications*, John Wiley & Sons, New York, NY, USA, 2nd edition, 2002.