Multi-sensor Distributed Estimation Fusion Using Minimum Distance Sum

Zhansheng Duan
Center for Info Engr Science Research
Xi'an Jiaotong University
Xi'an, Shaanxi 710049, China
Email: zduan@uno.edu

X. Rong Li
Department of Electrical Engineering
University of New Orleans
New Orleans, LA 70148, USA
Email: xli@uno.edu

U. D. Hanebeck
Intelligent Sensor-Actuator-Systems Laboratory
Institute for Anthropomatics and Robotics
Karlsruhe Institute of Technology, Germany
Email: uwe.hanebeck@ieee.org

***Abstract*—In multi-sensor distributed estimation fusion, local estimation errors are correlated in general. Two extreme ways to handle this correlation is either to ignore them completely or to have them fully considered. There is another case in the middle: it admits the existence of the correlation, but does not know how large it is. A sensible way is to set up an optimality criterion and optimize it over all possible such correlations. This work is a new development in the third class. First, a new general objective function is introduced, which is the minimum sum of statistical distances between the fused density and the local posterior densities. Then it is shown that the new criterion leads to a convex optimization problem if the Kullback-Leibler (KL) divergence is used as the statistical distance between assumed Gaussian densities. It is found that although the analytically obtained fused estimate using the new criterion differs from the simple convex combination rule only in mean squared error (MSE) by a scaling factor N (the number of sensors used), it is pessimistic semi-definite in MSE. Numerical examples illustrate the effectiveness of the proposed distributed fuser by comparing with several widely used distributed fusers.**

Keywords: Estimation fusion, distributed fusion, statistical distance, KL divergence, Gaussian assumption, convex optimization.

I. Introduction

Estimation fusion, or data fusion for estimation, is the problem of how to best utilize useful information contained in multiple sets of data for the purpose of estimating a quantity—a parameter or a process [1]. It has wide application because of potentially improved estimation accuracy, enhanced reliability and survivability, extended coverage and observability, etc. Estimation fusion has two basic architectures: *centralized fusion* and *distributed fusion*. In centralized fusion, all raw measurements are sent to the fusion center, while in distributed fusion, each sensor only sends in processed data. They have pros and cons in terms of performance, communication requirements, reliability, survivability, information sharing, etc. Theoretically speaking, centralized fusion is nothing but traditional estimation with distributed data, which can be simply tackled through data stacking. Comparatively, distributed fusion is more challenging and has been the focus of most fusion related research.

Research supported in part by National 973 project of China through grant 2013CB329405, National Natural Science Foundation of China through grant 61174138, NASA/LEQSF(2013-15)-Phase3-06 through grant NNX13AD29A and the Fundamental Research Funds for the Central Universities of China.

Distributed fusion has been studied for decades and results are abundant. It can also be split into two classes: *standard* and *nonstandard* [1]. In standard distributed fusion, also known as track fusion, local sensors send in their local estimates (tracks). Any distributed fusion other than standard distributed fusion all belong to nonstandard distributed fusion. For example, two lossless linear transformations of the sensor data were proposed in [2]. It was proven that nonstandard distributed fusion using the transformed data at the fusion center can achieve the same estimation performance as the centralized fusion. Also this nonstandard distributed fuser can help save communication from local sensors to the fusion center and may even have increased numerical robustness. It was found that local estimation MSE matrices are all singular in a multi-sensor constrained distributed fusion problem. To still use the existing distributed fusion rules, it was suggested to send in dimension reduced local estimates to the fusion center in [3]. Obviously, this also belongs to the nonstandard distributed fusion class. In light of communication constraints, it is usually more desirable for local sensors to send in compressed data [4]–[7]. The distributed fusion rules developed therein all belong to the nonstandard distributed fusion family.

The standard distributed fusion is the other focus of distributed fusion and there are also many results available for this class. The simple convex combination (SCC) method [8], [9] is probably the earliest such algorithm. It is also probably the simplest one in which the fused estimate is a matrix-valued weighted sum of local estimates. Due to the ignorance of the cross-correlation among local estimation errors, the SCC method is computationally simple. It was shown explicitly in [10] that due to common process noise and the common initialization, local estimation errors across sensors are correlated. By treating local estimates as data in an estimation problem, a two-sensor track-to-track fusion algorithm based on the maximum likelihood principle for the Gaussian case was proposed in [11] and [12]. For more than two sensors, a track-to-track fusion algorithm based on the maximum likelihood (for the Gaussian case) and optimal weighted least squares (OWLS) was proposed in [1], [13] and [14]. A standard distributed fusion algorithm that is optimal in the sense of maximum a posteriori principles was proposed in [15]. Furthermore, by additional use of local predicted estimates when compared with [15], another fusion algorithm was proposed in [1], [16], which is optimal in the sense of minimum mean-squared error (MMSE) for the Gaussian case or linear MMSE. Performance of distributed fusion with

LMMSE and OWLS estimation and measures of their relative efficiency compared with centralized fusion were discussed in [17]. Relationships among various LMMSE and OWLS fusion rules with complete, incomplete, and no prior information were clarified in [18]. Also, the effect of prior information and data on fusion performance was quantified in [18]. A general data model for discretized asynchronous multi-sensor systems was presented in [19] and serveral theoretically important issues unique to fusion for dynamic systems were also discussed in [19].

In all these standard distributed fusion algorithms, the cross-correlation among local estimation errors is fully considered. There is another class of standard distributed fusion algorithms, in which the cross-correlation among local estimation errors is circumvented. For example, for the case with uncorrelated measurement noise across sensors, by using both predicted and updated estimates from local sensors, the information matrix fusion method [20] was proposed in [21] and [22], which can completely reconstruct the optimal centralized estimate. An extension to the case with correlated measurement noises was proposed in [23]. Another well known standard distributed fusion algorithm in this class is the federated Kalman filter [24], [25]. It uses the upper bounding approach for the state and process noise covariance matrices to circumvent the cross-correlation among local estimation errors. It also uses the information-sharing principle and a master filter to combine the sensor-dedicated local filters. As a result, it can also achieve the same performance as the optimal centralized fusion. In many cases, however, it is not necessarily easy or feasible to obtain the cross-correlation among local estimation errors. For example, when local estimators are nonlinear filters or multiple-model estimators, it is usually difficult to obtain the cross-correlation exactly, sometimes even just approximately. To tackle this, other types of standard distributed fusion algorithms have been studied. For example, the covariance intersection (CI) method [26], [27] tries to reserve the simple form of the SCC method while considering the cross-correlation among local estimation errors to some extent. To make the fused estimate unique, CI uses some optimality criterion, e.g., minimization of the determinant or trace of the fused MSE. Modifications of the original CI method are abundant [28]–[31]. A main difference between the improved ones and the original one is in the optimality criterion.

In this paper, it is also assumed that cross-correlation among local estimation errors are unavailable/unknown to the fusion center and standard distributed fusion is considered. Unlike existing work on this problem, a new optimality criterion to obtain the fused estimate is introduced. Inspired by the least squares fitting for parameter estimation and the fact that statistical distance is a measure of similarity between densities, we propose to use the minimum sum of statistical "distances" between the fused density and local posterior densities as a new objective function. Intuitively, the fused density should be most similar to all local posterior densities. The KL divergence is chosen as the statistical closeness measure between densities in the new criterion. Although it is not a true distance, it does measure similarity between densities. To simplify computation, Gaussian distributions are assumed for all densities. It is found that the new criterion leads to a convex optimization problem, which is then solved analytically. The fused estimate using the new criterion is the same as the one given by the SCC method. But the fused MSE is different. For the two-sensor case, it is proven that the new fuser is pessimistic semi-definite in MSE. Numerical examples show that the new distributed fuser is effective when compared with several other widely used fusion rules.

This paper is organized as follows. Sec. II formulates the problem. Sec. III gives a brief summary and classification of the existing standard distributed fusion algorithms. Sec. IV presents the minimum sum of statistical distances based distributed fusion algorithm. Sec. V provides two numerical examples to illustrate the effectiveness of the proposed fuser. Sec. VI gives conclusions.

II. Problem Formulation

For simplicity, assume we want to estimate a parameter x. Suppose that altogether N sensors are used to observe x simultaneously and the local estimates of x are $\hat{x}^i = f^i(z^i)$, where z^i is the measurement of x at local sensor i and f^i is the corresponding local estimator, $i = 1, 2, \cdots, N$. The associated MSE of $\hat{x}^i$, $P^i = \text{MSE}(\hat{x}^i)$, is known.

In distributed fusion, we are trying to figure out a fused estimate $\hat{x} = f(\{\hat{x}^i, P^i\}_{i=1}^N)$ through some fusion rule $f(\cdot)$. It is expected that the fused estimate is better than any local estimate.

III. Existing classes of fusion algorithms

In this section, we give a summary of existing fusion rules in three basic classes. They differ in how the cross-correlation among local estimation errors is used. For each class, a representative fusion rule is provided.

A. Completely Ignored

In this class, the cross-correlation among local estimation errors is completely ignored although it is known to exist. This is so mainly for computational consideration. A representative of this class is the SCC rule, which is probably the earliest fusion rule in history:

$$\hat{x}^{\text{SCC}} = P^{\text{SCC}} \sum\nolimits_{i=1}^{N} (P^i)^{-1} \hat{x}^i,$$
$$P^{\text{SCC}} = (\sum\nolimits_{i=1}^{N} (P^i)^{-1})^{-1},$$

where P^{SCC} is the fuser calculated MSE matrix of $\hat{x}^{\text{SCC}}$.

It can be easily shown that the true MSE matrix of $\hat{x}^{\text{SCC}}$ is

$$\begin{aligned} P^{\text{SCC}}_{\text{true}} &= \text{MSE}(\hat{x}^{\text{SCC}}) = E[(x - \hat{x}^{\text{SCC}})(x - \hat{x}^{\text{SCC}})'] \\ &= P^{\text{SCC}}(\sum\nolimits_{i=1}^{N} (P^i)^{-1} \\ &\quad + \sum\nolimits_{i,j=1, i\neq j}^{N} (P^i)^{-1} P^{i,j} (P^j)^{-1}) P^{\text{SCC}} \\ &= P^{\text{SCC}} + P^{\text{SCC}} \sum\nolimits_{i,j=1,\ i\neq j}^{N} (P^i)^{-1} P^{i,j} (P^j)^{-1}) P^{\text{SCC}}, \end{aligned}$$

where $P^{i,j} = E[\tilde{x}^i (\tilde{x}^j)']$, $i \neq j$. However, we do not know whether $P^{\text{SCC}} \geq P^{\text{SCC}}_{\text{true}}$ or $P^{\text{SCC}} \leq P^{\text{SCC}}_{\text{true}}$ since the definiteness of the difference term $P^{\text{SCC}} \sum_{i,j=1, i\neq j}^{N} (P^i)^{-1} P^{i,j} (P^j)^{-1}) P^{\text{SCC}}$ is not clear. This can be easily seen from the numerical example provided later.

B. Fully Considered

The second class takes full account of the cross-correlation among local estimation errors and no information is lost. A representative of this class is the OWLS fusion rule:

$$\hat{x}^{\text{OWLS}} = P^{\text{OWLS}}\mathbf{I}_N'\mathbf{P}_N^{-1}\hat{\mathbf{x}}_N,$$
$$P^{\text{OWLS}} = (\mathbf{I}_N'\mathbf{P}_N^{-1}\mathbf{I}_N)^{-1},$$

where

$$\mathbf{I}_N = [\ I_n \quad I_n \quad \cdots \quad I_n\ ]',$$
$$\hat{\mathbf{x}}_N = [\ (\hat{x}^1)' \quad (\hat{x}^2)' \quad \cdots \quad (\hat{x}^N)'\ ]',$$
$$\mathbf{P}_N = \begin{bmatrix} P^1 & P^{1,2} & \dots & P^{1,N} \\ P^{2,1} & P^2 & \dots & P^{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ P^{N,1} & P^{N,2} & \dots & P^N \end{bmatrix},$$

and I_n is the $n \times n$ identity matrix.

Since the cross-correlation is fully accounted for, the fuser calculated MSE is the true MSE:

$$P^{\text{OWLS}} = P^{\text{OWLS}}_{\text{true}} = E[(x - \hat{x}^{\text{OWLS}})(x - \hat{x}^{\text{OWLS}})'].$$

Moreover, the MSE of the OWLS rule is the minimum one of all linear distributed fusion rules [18].

C. Selecting One Out of Many

The OWLS rule does perform the best theoretically, but its success depends heavily on correct knowledge of the cross-correlation. Unfortunately, such knowledge is not available in many cases. For a linear dynamic system with linear measurements, recursive formulas to obtain the cross-correlation were given in [10], [32]. However, if the dynamic system is nonlinear, it is usually hard to obtain the cross-correlation exactly. The same difficulty exists if local filtering is a multiple-model estimation problem. Also, even for a linear system, the limited communication bandwidth may not allow the transmission of all information needed to recursively calculate the cross-correlation exactly at the fusion center. In these cases, cross-correlation among local estimation errors is known to exist but not available precisely to the fusion center. Then we can get fused estimates that are worse than when the cross-correlation is fully considered, but not worse than when it is completely ignored. The third class of fusion rules is exactly for this purpose. A typical idea here is to consider all feasible cross-correlations. As a result, such fusion with unavailable cross-correlation has infinitely many possible fused estimators. The common practice in this class is to set up some objective function first and then optimize it over the feasible set of cross-correlations to get a meaningful one. This idea is widely used in signal processing. For example, an underdetermined system of linear equations also has infinitely many possible solutions. To get a unique meaningful solution, a criterion such as norm minimization or sparseness maximization [33], [34] is imposed.

A representative of this class is the CI algorithm:

$$\hat{x}^{\text{CI}} = P^{\text{CI}}\sum\nolimits_{i=1}^{N}\omega_i(P^i)^{-1}\hat{x}^i,$$
$$P^{\text{CI}} = (\sum\nolimits_{i=1}^{N}\omega_i(P^i)^{-1})^{-1},$$

where

$$\sum\nolimits_{i=1}^{N}\omega_i = 1 \text{ and } \omega_i \geq 0,$$

and ω^i, $i = 1, 2, \cdots, N$, are determined by some optimality criterion, e.g., minimization of the determinant or trace of P^{CI}.

It can be easily shown that the true MSE matrix of $\hat{x}^{\text{CI}}$ is

$$\begin{aligned} P^{\text{CI}}_{\text{true}} &= \text{MSE}(\hat{x}^{\text{CI}}) = E[(x - \hat{x}^{\text{CI}})(x - \hat{x}^{\text{CI}})'] \\ &= P^{\text{CI}}(\sum\nolimits_{i=1}^{N}\omega_i^2(P^i)^{-1} \\ &\quad + \sum\nolimits_{i,j=1,\ i\neq j}^{N}\omega_i\omega_j(P^i)^{-1}P^{i,j}(P^j)^{-1})P^{\text{CI}}. \end{aligned}$$

It was proven in [26], [27] that for the two-sensor case (i.e., $N = 2$), we have

$$P^{\text{CI}} \geq P^{\text{CI}}_{\text{true}}.$$

That is why this fuser is sometimes said to be conservative.

The new fusion rule we will propose also falls into the third class and is presented next.

IV. Fusion based on minimum distance sum

A. Optimization Problem Formulation

Suppose that the local posterior densities are $p(x|z^i)$, $i = 1, 2, \cdots, N$. A potential fusion rule is to minimize distance sum:

$$\hat{p}(x|z) = \arg\min_{p(x|z)}\sum\nolimits_{i=1}^{N} d(p(x|z), p(x|z^i)),$$

where $d(\cdot,\cdot)$ is some statistical "distance" between two distributions, $p(x|z)$ is an arbitrary posterior density function of x, $\hat{p}(x|z)$ is the fused posterior density of x, and $z = \{z^i\}_{i=1}^N$.

The use of the minimum distance sum criterion for distributed fusion can be justified as follows. The statistical "distance" $d(p(x), p(x|z^i))$ measures the similarity between $p(x)$ and $p(x|z^i)$. The smaller $d(p(x), p(x|z^i))$ is, the more similar $p(x)$ and $p(x|z^i)$ are. Thus of all possible distributions $\hat{p}(x)$ obtained is most similar to the local posterior densities.

Remark 1: The above criterion is inspired by the widely used least squares criterion. As is well known, the sum of fitting error squared is minimized in the least squares method where the fitting error measures similarity between the actual data and the fitted data. Since the fitting error can be either positive or negative, it is squared to guarantee non-negativeness and mathematical tractability. For the minimum distance sum criterion, there is no need to do so because the statistical "distance" is always non-negative.

To make this criterion work for estimation fusion, a key issue is to determine the statistical distance used. A widely used closeness measure between two arbitrary density functions $p(x)$ and $q(x)$ is the KL divergence [35] (also referred to as information divergence, information gain, relative entropy, KL distance, etc.):

$$\begin{aligned} D_{KL}(p(\cdot), q(\cdot)) &= \int \log(\frac{p(x)}{q(x)}) \cdot p(x)dx \\ &= E_{p(x)}[\log(p(x)) - \log(q(x))]. \end{aligned}$$

The KL divergence measures the average logarithmic difference between $p(\cdot)$ and $q(\cdot)$, so it can be used as a candidate distance in our criterion. However, the KL divergence is not a true distance: it is not symmetric, nor does it satisfy the triangle inequality.

Since the KL divergence is asymmetric: $D_{KL}(p(\cdot), q(\cdot)) \neq D_{KL}(q(\cdot), p(\cdot))$, to make our criterion work for estimation fusion, it matters whether $p(x|z)$ or $p(x|z^i)$ is the first argument of the KL divergence. It is more preferable to use $p(x|z)$ as the first argument, that is, to use the following criterion

$$\hat{p}(x|z) = \arg\min_{p(x|z)} \sum_{i=1}^{N} D_{KL}(p(x|z), p(x|z^i)).$$

This is because $D_{KL}(p(\cdot), q(\cdot))$ is the divergence of $q(x)$ from $p(x)$. In other words, $p(x)$ is considered the underlying "true" or "best guess" distribution, so expectations are evaluated with reference to it, while $q(x)$ is another distribution.

As formulated above in estimation fusion, most often we know only the first two moments of each local estimation instead of the local posterior densities. To make our criterion using the KL divergence work, the following Gaussian assumption is made

$$\begin{aligned} p(x|z^i) &\approx \mathcal{N}(x; \hat{x}^i, P^i),\ i = 1, 2, \cdots, N \\ p(x) &\approx \mathcal{N}(x; \hat{x}, P). \end{aligned}$$

Remark 2: Such a Gaussian approximation to each local posterior density is widely used in filtering, especially in nonlinear filtering. Nonlinear filters based on such an approximation include Gaussian-Hermite filtering [36], [37], etc. Although some other nonlinear filters, e.g., the unscented filter [38] and the divided difference filter [39] using Stirling's interpolation do not depend on Gaussian approximation directly, they often use a Gaussian assumption indirectly when designing deterministic sampling points.

When both $p(x)$ and $q(x)$ are Gaussian as

$$p(x) = \mathcal{N}(x; \mu_p, \Sigma_p),\ q(x) = \mathcal{N}(x; \mu_q, \Sigma_q),$$

their KL divergence is simplified to

$$\begin{aligned} D_{KL}(p(\cdot), q(\cdot)) = \frac{1}{2}(&\text{tr}(\Sigma_q^{-1}\Sigma_p) + (\mu_p - \mu_q)'\Sigma_q^{-1}(\mu_p - \mu_q) \\ &+ \ln(\frac{|\Sigma_q|}{|\Sigma_p|}) - n), \end{aligned}$$

where n is the dimension of x.

With the above Gaussian assumptions for both local posterior densities and the fused global density, our criterion using the KL divergence becomes

$$\begin{aligned} (\hat{x}, P) = \arg\min_{x,\Sigma} \sum_{i=1}^{N} \frac{1}{2}(&\text{tr}((P^i)^{-1}\Sigma) \\ &+ (\hat{x}^i - x)'(P^i)^{-1}(\hat{x}^i - x) + \ln(\frac{|P^i|}{|\Sigma|}) - n). \end{aligned}$$

Since x and Σ are uncoupled, this minimization problem can be equivalently decomposed into the following two minimization sub-problems

$$\hat{x} = \arg\min_{x} \sum_{i=1}^{N} \frac{1}{2}(\hat{x}^i - x)'(P^i)^{-1}(\hat{x}^i - x), \quad (1)$$

$$P = \arg\min_{\Sigma>0} \sum_{i=1}^{N} \frac{1}{2}(\text{tr}((P^i)^{-1}\Sigma) + \ln(\frac{|P^i|}{|\Sigma|}) - n). \quad (2)$$

B. Solution of the Formulated Optimization Problem

Sub-problem (1) is a weighted least squares problem if we treat $\hat{x}^i$ as sensor observed data and P^i as the corresponding measurement noise covariance matrix. That is, we can construct the following pseudo measurement equations [1] about x: $\hat{x}^i = x - \tilde{x}^i,\ i = 1, 2, \cdots, N$, where the estimation error $\tilde{x}^i$ is treated as the pseudo measurement noise with the first two moments

$$\begin{aligned} E[\tilde{x}^i] = 0,\ \text{cov}(\tilde{x}^i) = P^i, \\ \text{cov}(\tilde{x}^i, \tilde{x}^j) = 0,\ i, j = 1, 2, \cdots, N,\ i \neq j. \end{aligned}$$

Note that the cross-correlations between $\tilde{x}^i$ and $\tilde{x}^j$, $i, j = 1, 2, \cdots, N,\ i \neq j$, are all completely ignored so that the fitting error under the least squares criterion can be written in the summation form of (1) although they do exist. Thus the solution of (1) is

$$\hat{x} = (\sum_{i=1}^{N} (P^i)^{-1})^{-1} \sum_{i=1}^{N} (P^i)^{-1}\hat{x}^i.$$

The second sub-problem (2) for the MSE matrix is much harder than problem (1) in two aspects. First, problem (2) is a matrix optimization problem. Second, the minimizer P is required to be symmetric and positive definite since it is an MSE matrix and (2) involves $\ln|\Sigma|$.

Due to the constraint $\Sigma > 0$, problem (2) is a semi-definite programming problem in essence. Next we discuss how to solve this minimization problem analytically.

Define

$$f(\Sigma) = \sum_{i=1}^{N} \frac{1}{2}(\text{tr}((P^i)^{-1}\Sigma) + \ln(\frac{|P^i|}{|\Sigma|})).$$

For any $X_1,\ X_2 \in \mathbb{S}_{++}^n$, where $\mathbb{S}_{++}^n$ is the set of all $n \times n$ positive definite matrices, and any θ with $0 \leq \theta \leq 1$, it can be easily seen that

$$0 < \theta X_1 + (1 - \theta)X_2 \in \mathbb{S}_{++}^n.$$

So the domain $\mathbb{S}_{++}^n$ of $f(\Sigma)$ is a convex set.

It is well known that affine functions are both convex and concave [40]. Since $g(X) = \text{tr}(A'X) + b = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}X_{ij} + b$, where $X \in \mathbb{R}^{m\times n}$, is an affine function, all terms $\text{tr}((P^i)^{-1}\Sigma)$, $i = 1, 2, \cdots, N$, within $f(\Sigma)$ are convex and concave. Also, it is known that the function $\ln|X|$ with $X \in \mathbb{S}_{++}^n$ is a concave function [40]. So $-\ln|\Sigma|$ is a convex function. Furthermore, since a nonnegative weighted sum of convex functions is still convex [40], $f(\Sigma)$ is clearly a convex function over $\mathbb{S}_{++}^n$. Therefore, problem (2) is a convex programming problem.

[1] This is exactly the same idea for unified linear model used in [1] for distributed fusion so that the centralized fusion and distributed fusion can be treated in a unified manner.

Using Properties 1 and 2 in the appendix, we have

$$\frac{\partial f}{\partial \Sigma} = \frac{1}{2}(2C - \text{diag}(C)),$$

where

$$C = \sum\nolimits_{i=1}^{N}((P^i)^{-1} - \Sigma^{-1}).$$

Set $\frac{\partial f}{\partial \Sigma} = 0$, that is,

$$2C - \text{diag}(C) = 0.$$

Clearly this equation holds if and only if $C = 0$. As a result, the unique stationary point of $f(\Sigma)$ is

$$\Sigma^* = N \cdot (\sum\nolimits_{i=1}^{N}(P^i)^{-1})^{-1}.$$

Since $f(\Sigma)$ is a convex function defined on the convex set $\mathbb{S}_{++}^n$ and $\frac{\partial f}{\partial \Sigma}\Big|_{\Sigma=\Sigma^*} = 0$, Σ^* is the unique global minimizer of $f(\Sigma)$ over $\mathbb{S}_{++}^n$ [41].

In summary, the best fused estimate using our criterion is

$$\hat{x} = (\sum\nolimits_{i=1}^{N}(P^i)^{-1})^{-1}\sum\nolimits_{i=1}^{N}(P^i)^{-1}\hat{x}^i,$$
$$P = N \cdot (\sum\nolimits_{i=1}^{N}(P^i)^{-1})^{-1}.$$

Remark 3: Due to the constraint $\Sigma > 0$, there are only $n(n+1)/2$ independent variables within Σ. Without it, all $n \times n$ elements of Σ must be treated as independent variables and problem (2) would be much harder.

C. Discussion on our least distance based distributed fusion

Remark 4: It can be easily seen that our fusion rule differs from the SCC rule only in the fuser calculated MSE matrix by a scaling factor of N.

Remark 5: It can also be seen that our fusion rule turns out to be a special case of the CI rule with $\omega_1 = \omega_2 = \cdots = \omega_N = \frac{1}{N}$.

For the two-sensor case, our fusion rule is pessimistic semi-define in MSE matrix [42], i.e.,

$$P \geq P_{\text{true}},$$

where P_{true} is the true MSE matrix of the fused estimate $\hat{x}$. This can be shown as follows.

For the two-sensor case, it follows from our fusion rule that

$$\hat{x} = ((P^1)^{-1} + (P^2)^{-1})^{-1}((P^1)^{-1}\hat{x}^1 + (P^2)^{-1}\hat{x}^2).$$

Correspondingly,

$$\begin{aligned}\tilde{x} &= x - \hat{x} \\ &= ((P^1)^{-1} + (P^2)^{-1})^{-1}((P^1)^{-1}\tilde{x}^1 + (P^2)^{-1}\tilde{x}^2).\end{aligned}$$

Thus

$$\begin{aligned}P_{\text{true}} = \text{MSE}(\hat{x}) &= E[(x-\hat{x})(x-\hat{x})'] \\ &= ((P^1)^{-1} + (P^2)^{-1})^{-1}((P^1)^{-1} + (P^2)^{-1} \\ &+ (P^1)^{-1}P^{1,2}(P^2)^{-1} + (P^2)^{-1}P^{2,1}(P^1)^{-1}) \\ &\cdot ((P^1)^{-1} + (P^2)^{-1})^{-1},\end{aligned}$$

where $P^{i,j} = E[\tilde{x}^i(\tilde{x}^j)'],\ i \neq j$. Then we have

$$\begin{aligned}&P - P_{\text{true}} \\ &= ((P^1)^{-1} + (P^2)^{-1})^{-1}(2(P^1)^{-1} + 2(P^2)^{-1}) \\ &\quad \cdot ((P^1)^{-1} + (P^2)^{-1})^{-1} \\ &= ((P^1)^{-1} + (P^2)^{-1})^{-1}((P^1)^{-1} + (P^2)^{-1} \\ &\quad - (P^1)^{-1}P^{1,2}(P^2)^{-1} - (P^2)^{-1}P^{2,1}(P^1)^{-1}) \\ &\quad \cdot ((P^1)^{-1} + (P^2)^{-1})^{-1} \\ &= ((P^1)^{-1} + (P^2)^{-1})^{-1}E\{[(P^1)^{-1}\tilde{x}^1 - (P^2)^{-1}\tilde{x}^2][\cdot]'\} \\ &\quad \cdot ((P^1)^{-1} + (P^2)^{-1})^{-1} \\ &\geq 0.\end{aligned}$$

V. Illustrative Examples

In the following, we compare the local estimation and the fusion performance of the SCC, the OWLS, the CI and our method (abbreviated as KL). The performance evaluation measures used are true MSE, fuser calculated MSE, noncredibility index (NCI) [42] and inclination indicator (II) [42] over $5,000$ Monte Carlo runs.

Consider a two-sensor scalar case, where the MSE's of the local estimation along with the cross-correlation are described as

$$\mathbf{P}_2 = \begin{bmatrix} P^1 & P^{1,2} \\ P^{2,1} & P^2 \end{bmatrix} = \begin{bmatrix} P^1 & \rho\sqrt{P^1P^2} \\ \rho\sqrt{P^1P^2} & P^2 \end{bmatrix}.$$

Here ρ is the correlation coefficient between local estimation errors $\tilde{x}^1$ of sensor 1 and $\tilde{x}^2$ of sensor 2, and $P^1 = 9$.

To obtain NCI and II, it is assumed that both local estimators are unbiased and the joint distribution of $\tilde{x}^1$ and $\tilde{x}^2$ is

$$[\ \tilde{x}^1 \quad \tilde{x}^2\ ]' \sim \mathcal{N}(\mathbf{0}_2, \mathbf{P}_2),$$

where $\mathbf{0}_2$ is the 2-dimensional zero column vector.

Suppose that $P^2 < P^1$, i.e., sensor 2 has better estimation accuracy than sensor 1. Then the fuser calculated MSE of the CI rule for this example is

$$P^{\text{CI}} = \frac{1}{\frac{\omega_1}{P^1} + \frac{\omega_2}{P^2}} = \frac{P^1P^2}{P^2\omega_1 + P^1\omega_2} = \frac{P^1P^2}{(P^1 - P^2)\omega_2 + P^2},$$

where the property $\omega_1 + \omega_2 = 1,\ \omega_1, \omega_2 \geq 0$ has been used.

Since P^{CI} is a scalar, it does not matter whether the determinant or trace is minimized to get ω_1 and ω_2. Furthermore, from the above, it can be easily seen that

$$\min_{\omega_1,\omega_2} P^{\text{CI}},\ \text{s.t. } \omega_1 + \omega_2 = 1,\ \omega_1, \omega_2 \geq 0$$

is equivalent to

$$\max_{\omega_2}(P^1 - P^2)\omega_2 + P^2,\ \text{s.t. } 1 \geq \omega_2 \geq 0.$$

It achieves the maximum at $\omega_2^* = 1$ because $P^1 > P^2$. Therefore, the optimal ω_1 and ω_2 are

$$\omega_1^* = 0,\ \omega_2^* = 1.$$

That is, the CI fusion rule simply chooses the more accurate local estimate (sensor 2) as the fused one.

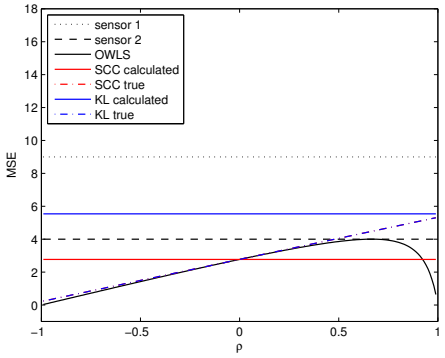

Figure 1. MSE of Case 1. Note that SCC true overlaps with KL true.

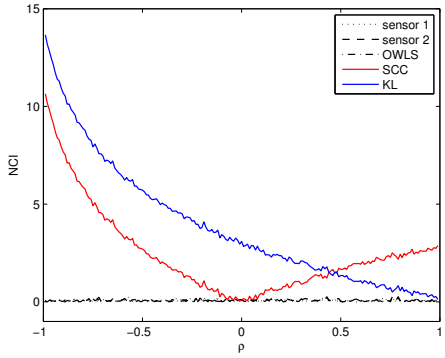

Figure 2. NCI of Case 1.

A. Case 1

In this case, the local MSE of sensor 2 is $P^2 = 4$, that is, the two sensors have quite different estimation accuracy.

Fig. 1 shows the MSEs of the two local sensors and of all four distributed fusers.

It can be easily seen that our new fuser is pessimistic semi-definite in MSE [42] since the fuser calculated MSE is always greater than or equal to its true MSE. Also, the true MSE is a linearly increasing function of the correlation coefficient ρ. The smaller ρ is, the greater the difference between the fuser calculated MSE and the true MSE is. As ρ approaches 1, the fuser calculated MSE gradually approaches the true MSE. In terms of true MSE, our fuser is better than both local estimators in most cases and it also performs very close to the OWLS fuser, which is the best among all, in most cases.

It can be seen that the SCC fuser is neither optimistic nor pessimistic semi-definite and it is not perfectly credible either, though its calculated MSE is better than the MSE of any local estimator. This is because there is no fixed relationship between the fuser calculated MSE and the true MSE when the SCC fuser is used. As can be seen from Fig. 1, in half of the cases, the fuser calculated MSE is less than the true MSE and in the other half cases, it is the opposite. It can thus be concluded that compared with the SCC fuser, the amplification factor N in our fuser contributes to make our fuser pessimistic semi-definite.

It should be noted that the CI fuser for this case overlaps with sensor 2 as explained above. Thus it is perfectly credible. In terms of true MSE, it can be seen that only when ρ is large, it performs between our fuser (also the SCC fuser) and the OWLS fuser. Otherwise, it is the worst one of all fusers.

Figs. 2 and 3 show the credibility of the two local sensors and of all four fusers in terms of NCI and II.

It can be seen that our fuser is pessimistic semi-definite in MSE since its II is always negative. The SCC fuser is neither optimistic nor pessimistic semi-definite since its II can be both positive and negative, depending on ρ. And it is not perfectly credible either. Using NCI and II, it can be seen that sensor 1, sensor 2 (also the CI fuser) and the OWLS fuser are all perfectly credible.

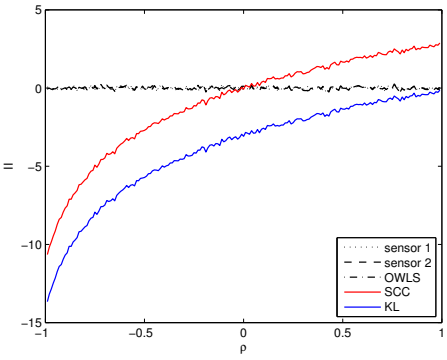

Figure 3. II of Case 1.

B. Case 2

In this case, the local MSE of sensor 2 is $P^2 = 8.5$, that is, the two sensors have quite close estimation accuracy.

Figs. 4 to 6 show the estimation performance of the two local sensors and of all four fusers.

It can be easily seen that the trend for all measures is pretty much the same as in Case 1. The main difference is that the MSE of the OWLS fuser is very close to the true MSE of our fuser (also the SCC fuser) for almost all ρ values. This further

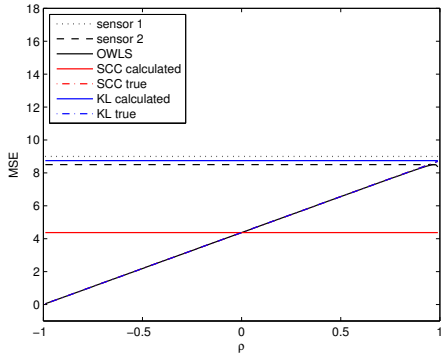

Figure 4. MSE of Case 2. Note that SCC true overlaps with KL true. Also note that KL true almost overlaps with OWLS except when ρ is very close to 1.

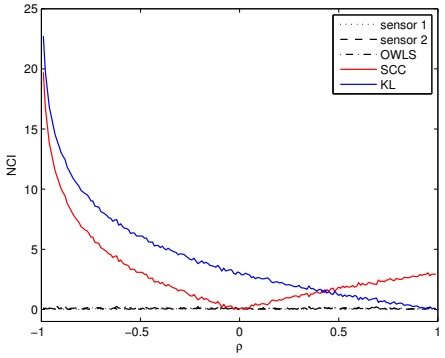

Figure 5. NCI of Case 2.

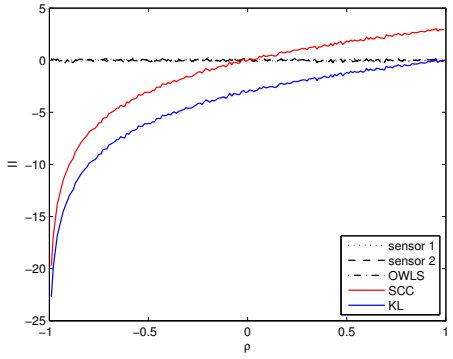

Figure 6. II of Case 2.

supports the use of our fuser since the OWLS fuser is the best among all.

VI. Conclusions

This work considers the multi-sensor distributed estimation fusion problem with unavailable/unknown cross-correlation of local estimation errors across sensors. Our framework is the same as the one used in most existing work for this problem: set up an optimality criterion first and then optimize it over all possible cross-correlation. However, a new objective function using the minimum sum of statistical distances between the fused density and the posterior densities from local sensors is proposed. This is inspired by the least squares fitting for parameter estimation and the fact that distance between densities is a measure of their similarity. The fused density is then the most similar one to the local posterior densities. By choosing the KL divergence as the distance in the new criterion and using Gaussian assumptions for the first two moments, which are the only quantity used in point estimation, it is shown that the new optimality criterion leads to a convex optimization problem. Its unique globally optimal analytical solution provides the same fused estimate as the simple convex combination method does. However, its fused MSE is that calculated by the simple convex combination method times the number of sensors used. This helps guarantee the new fuser to be pessimistic semi-definite. Numerical examples presented have shown that the new fuser is advantageous over several widely used fusers in terms of both estimation accuracy and estimator credibility.

Appendix

For convenience, two properties of matrix algebra from [43] are summarized below for easy reference:

1) If $X \in \mathbb{R}^{m\times m}$, $A \in \mathbb{R}^{m\times m}$ and $X = X'$, then $\frac{\partial \mathrm{tr}(AX)}{\partial X} = A + A' - \mathrm{diag}(A)$, where $\mathrm{diag}(A)$ is the diagonal matrix having the same diagonal elements as A.
2) If $X \in \mathbb{R}^{m\times m}$, $|X| > 0$ and $X = X'$, then $\frac{\partial \ln |X|}{\partial X} = 2X^{-1} - \mathrm{diag}(X^{-1})$.

References

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

[2] Z. S. Duan and X. R. Li, "Lossless linear transformation of sensor data for distributed estimation fusion," *IEEE Transactions on Signal Processing*, vol. 59, no. 1, pp. 362–372, January 2011.

[3] ——, "Multi-sensor estimation fusion for linear equality constrained dynamic systems," in *Proceedings of the 16th International Conference on Information Fusion*, Istanbul, Turkey, 9-12 July 2013, pp. 93–100.

[4] K. S. Zhang, X. R. Li, P. Zhang, and H. F. Li, "Optimal linear estimation fusion - part VI: sensor data compression," in *Proceedings of the 6th International Conference on Information Fusion*, Cairns, Quensland, Australia, 8-11 July 2003, pp. 221–228.

[5] Y. M. Zhu, E. B. Song, J. Zhou, and Z. S. You, "Optimal dimensionality reduction of sensor data in multisensor estimation fusion," *IEEE Transactions on Signal Processing*, vol. 53, no. 5, pp. 1631–1639, May 2005.

[6] E. B. Song, Y. M. Zhu, and J. Zhou, "Sensors' optimal dimensionality compression matrix in estimation fusion," *Automatica*, vol. 41, no. 12, pp. 2131–2139, December 2005.

[7] I. D. Schizas, G. B. Giannakis, and Z. Q. Luo, "Distributed estimation using reduced-dimensionality sensor observations," *IEEE Transactions on Signal Processing*, vol. 55, no. 8, pp. 4284–4299, August 2007.

[8] C. Y. Chong, S. Mori, W. H. Barker, and K. C. Chang, "Architectures and algorithms for track association and fusion," *IEEE Aerospace and Electronic Systems Magazine*, vol. 15, no. 1, pp. 5–13, January 2000.

[9] S. Mori, W. H. Barker, C. Y. Chong, and K. C. Chang, "Track association and track fusion with nondeterministic target dynamics," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 38, no. 2, pp. 659–688, April 2002.

[10] Y. Bar-Shalom, "On the track-to-track correlation problem," *IEEE Transactions on Automatic Control*, vol. 26, no. 2, pp. 571–572, Apr 1981.

[11] 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, November 1986.

[12] K. C. Chang, R. K. Saha, and Y. Bar-Shalom, "On optimal track-to-track fusion," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 33, no. 4, pp. 1271–1276, October 1997.

[13] K. H. Kim, "Development of track to track fusion algorithms," in *Proceedings of the 1994 American Control Conference*, Baltimore, MD, 29 June-1 July 1994, pp. 1037–1041.

[14] H. M. Chen, T. Kirubarajan, and Y. Bar-Shalom, "Performance limits of track-to-track fusion versus centralized estimation: theory and application," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 2, pp. 386–400, April 2003.

[15] K. C. Chang, Z. Tian, and S. Mori, "Performance evaluation for MAP state estimate fusion," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 40, no. 2, pp. 706–714, April 2004.

[16] X. Tian and Y. Bar-Shalom, "Exact algorithms for four track-to-track fusion configurations: all you wanted to know but were afraid to ask," in *Proceedings of the 12th International Conference on Information Fusion*, Seattle, Washington, 6-9 July 2009, pp. 537–544.

[17] X. R. Li and K. S. Zhang, "Optimal linear estimation fusion - part IV: optimality and efficiency of distributed fusion," in *Proceedings of the 4th International Conference on Information Fusion*, Montreal, QC, Canada, August 2001, pp. WeB1–19–WeB1–26.

[18] X. R. Li, K. S. Zhang, J. Zhao, and Y. M. Zhu, "Optimal linear estimation fusion - part V: relationships," in *Proceedings of the 5th International Conference on Information Fusion*, Annapolis, MD, July 2002, pp. 497–504.

[19] X. R. Li, "Optimal linear estimation fusion - part VII: dynamic systems," in *Proceedings of the 6th International Conference on Information Fusion*, Cairns, Australia, July 2003, pp. 455–462.

[20] K. C. Chang, Z. Tian, and R. K. Saha, "Performance evaluation of track fusion with information matrix filter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 38, no. 2, pp. 455–466, April 2002.

[21] C. Y. Chong, "Hierarchical estimation," in *Proceedings of the MIT/ONR Workshop on C3*, Monterey, CA, 1979.

[22] 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, January 1988.

[23] E. B. Song, Y. M. Zhu, J. Zhou, and Z. S. You, "Optimal Kalman filtering fusion with cross-correlated sensor noises," *Automatica*, vol. 43, no. 8, pp. 1450–1456, August 2007.

[24] N. A. Carlson, "Federated square root filter for decentralized parallel processors," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 26, no. 3, pp. 517–525, May 1990.

[25] ——, "Federated filter for computer-efficient, near-optimal GPS integration," in *Proceedings of 1996 IEEE Position Location and Navigation Symposium*, Atlanta, GA, 22-26 April 1996, pp. 306–314.

[26] S. J. Julier and J. K. Uhlmann, "A non-divergent estimation algorithm in the presence of unknown correlations," in *Proceedings of the 1997 American Control Conference*, vol. 4, Albuquerque, NM, 4-6 June 1997, pp. 2369–2373.

[27] ——, "General decentralized data fusion with covariance intersection," in *Handbook of multisensor data fusion: theory and practice*, 2nd ed., M. E. Liggins, D. L. Hall, and J. Llinas, Eds. Boca Raton, FL: CRC Press, 2009, ch. 14, pp. 319–343.

[28] 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, Montreal, QC, Canada, 7-10 August 2001, pp. WeA2.11–WeA2.18.

[29] M. B. Hurley, "An information-theoretic justification for covariance intersection and its generalization," in *Proceedings of the 5th International Conference on Information Fusion*, Annapolis, MD, 8-11 July 2002, pp. 505–511.

[30] Y. M. Wang and X. R. Li, "Distributed estimation fusion under unknown cross-correlation: an analytic center approach," in *Proceedings of the 13th International Conference on Information Fusion*, EICC Edinburgh, UK, 26-29 July 2010, pp. 1–8.

[31] ——, "Distributed estimation fusion with unavailable cross-correlation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 1, pp. 259–278, January 2012.

[32] X. R. Li and P. Zhang, "Optimal linear estimation fusion - part III: cross-correlation of local estimation errors," in *Proceedings of the 4th International Conference on Information Fusion*, Montreal, QC, Canada, August 2001, pp. WeB1–11–WeB1–18.

[33] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.

[34] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, March 2008.

[35] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, 2006.

[36] K. Ito and K. Q. Xiong, "Gaussian filters for nonlinear filtering problems," *IEEE Transactions on Automatic Control*, vol. 45, no. 5, pp. 910–927, May 2000.

[37] I. Arasaratnam, S. Haykin, and R. J. Elliott, "Discrete-time nonlinear filtering algorithms using Gauss-Hermite quadrature," *Proceedings of the IEEE*, vol. 95, no. 5, pp. 953–977, May 2007.

[38] S. Julier, J. Uhlmann, and H. F. Durrant-Whyte, "A new method for nonlinear transformation of means and covariances in filters and estimators," *IEEE Transactions on Automatic Control*, vol. 45, no. 3, pp. 477–482, March 2000.

[39] M. Norgaard, N. K. Poulsen, and O. Ravn, "New developments in state estimation for nonlinear systems," *Automatica*, vol. 36, no. 11, pp. 1627–1638, November 2000.

[40] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2009.

[41] E. K. P. Chong and S. H. Zak, *An Introduction to Optimization*, 4th ed. John Wiley & Sons, 2013.

[42] X. R. Li and Z. L. Zhao, "Measuring estimator's credibility: noncredibility index," in *Proceedings of the 9th International Conference on Information Fusion*, Florence, Italy, 10-13 July 2006, pp. 1–8.

[43] D. A. Harville, *Matrix Algebra from A Statistician's Perspective*. Springer, 1997.