Optimal Linear Estimation Fusion — Part VI: Sensor Data Compression[*]

Keshu Zhang   X. Rong Li   Peng Zhang

Department of Electrical Engineering, University of New Orleans, New Orleans, LA 70148

Phone: 504-280-7416, Email: xli@uno.edu

Haifeng Li

Department of Computer Science and Engineering, University of California, Riverside, CA 92521

Phone: 909-787-2882, Email: hli@cs.ucr.edu

Abstract − *In many engineering applications, estimation accuracy can be improved by data from distributed sensors. Due to limited communication bandwidth and limited processing capability at the fusion center, it is crucial to compress these data for the final estimation at the fusion center. One way of accomplishing this is to reduce the dimension of the data with minimum or no loss of information. Based on the best linear unbiased estimation (BLUE) fusion results obtained in the previous parts of this series, in this paper we present optimal rules for compressing data at each local sensor to an allowable size (i.e., dimension) such that the fused estimate is optimal. We show that without any performance deterioration, all sensor data can be compressed to a dimension not larger than that of the estimatee (i.e., the quantity to be estimated). For some simple cases, these optimal compression rules are given analytically; for the general case, they can be found numerically by an algorithm proposed here. Supporting simulation results are provided.*

Keywords: Estimation fusion, BLUE, MSE, Sensor compression rule

1   Introduction

In recent years, the applications of data fusion techniques have increased significantly, such as in target tracking, image processing, economic data analysis. The motivation behind using multiple sensors has many folds: to reduce error and uncertainty in the measurement, to obtain results that would not be accessible using a single sensor, etc. Data fusion techniques are used to combine the outputs of multiple sensors. 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 an unknown quantity – a parameter or process. These data sets are usually obtained from multiple sources (e.g., multiple sensors).

Estimation fusion has been investigated for more than two decades. Many results have been obtained (e.g., [1, 2, 9, 8, 7, 6]). These results focus on building the optimal fusion rules for distributed estimation fusion. In target tracking, the most commonly used distributed architecture — the standard distributed fusion [1] — only local estimates are available at the fusion center. However, for many applications, distributed fusion has a more general architecture. In distributed fusion, each local sensor sends linearly or nonlinearly processed data to the fusion center. How to define these local mappings is still an open problem. The reason is that a fusion rule can be easily defined by optimizing a certain criterion. For example, MMSE (minimum mean square error), BLUE (best linear unbiased estimation), and WLS (weighted least square) fusion rules minimize mean square error (matrix), mean square error within the class of linear rules, and the weighted data fitting error, respectively. However, within a single local sensor, there is no clear-cut criterion to define the best mapping for processing the sensor observations because our final goal is to achieve the optimal estimation at the fusion center rather than at the local sensor. For the standard estimation fusion architecture, local sensor uses the same optimality criterion as the fusion center and obtains the rule that maps local observations to the optimal local estimate. A benefit of this structure is that local sensors can also have their own optimal estimates. Limitations of this local sensor rule are obvious. This mapping does not optimize estimation at the fusion center, nor consider the communication capacity between the local sensors and the fusion center, as well as fusion center's processing ability.

An optimal sensor compression rule should yield the

---

optimal fusion at the center with the constraint on the communication bandwidth between the fusion center and the local sensors, and the processing capability of the fusion center. The problem of creating the data compression rule is a constrained optimization problem. Without constraints, the problem is trivial: the local sensors' observations can be directly sent to the fusion center, and then a globally optimal solution is guaranteed. Note that the size of the raw data is often large, which has a high demand on communication bandwidth and the fusion center must have good computation capability and large memory. For example, if we consider a uniform quantization for every dimension of the data, the larger the data size, the more bits we need to send. Thus it is crucial to consider data compression rules for the local sensors. In this paper, for the fusion rule, we consider the BLUE fusion. In the Gaussian case, BLUE is equivalent with MMSE. Many practical problems have the Gaussian assumption. Without loss of generality, we assume the dimension of each single observation is fixed. We consider compressing the local raw measurement to a lower dimension. Since it is very difficult to discuss general nonlinear transformations, our discussion is limited to linear rules. These linear rules are optimal in the sense that the fused estimates at the center are optimal under the constraints.

The remainder of this paper is organized as follows. Section 2 formulates the distributed estimation fusion problem. Section 3 presents the basic data compression rule for a single sensor in a distributed system. A general data compression rule for a multiple sensor system is discussed in Section 4, along with a Gauss-Seidel iteration algorithm to obtain the optimal solution. Several numerical examples are presented in Section 5 to support the theoretical results. Section 6 provides a summary.

2 Problem Formulation

Consider a distributed system with a fusion center and n sensors (local stations), each connected to the fusion center.

Denote by z_i (an n_i-dimensional vector) the observations of the ith sensor of the estimatee (i.e., the quantity to be estimated) x. For a distributed fusion system, only data-processed observations are sent to the fusion center, that is, a non-trivial mapping g_i ($i = 1, \ldots, n$) is applied on sensor measurements:

$$g_i : z_i \rightarrow y_i$$

After sensor data processing, the available information at the fusion center is $y = [y_1', \ldots, y_n']'$. If $y_i = z_i$, it is known as centralized fusion, central-level fusion,
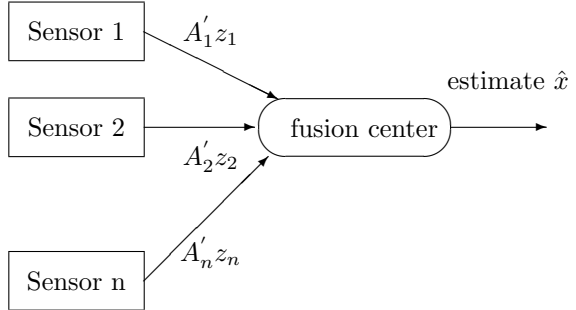

Figure 1: Framework of estimation fusion

or measurement fusion. If $y_i = \hat{x}_i$, that is, only local estimates (based on z_i) and its MSE matrix P_i are available at the fusion center, it is the standard distributed fusion. In general, the sensor rule $g_i(\cdot)$ could be any linear or nonlinear mapping. In this paper, only linear rules are considered, that is,

$$g_i(z_i) = A_i' z_i : R^{n_i} \rightarrow R^{d_i}$$

where $d_i \leq n_i$ and d_i is less than or equal to the dimensional requirement for the ith sensor due to communication or processing constraints of the system. After this data processing, sensor i sends $y_i = A_i' z_i$ to the fusion center. Since the original data size is reduced by the local processing, in the sequel, we call this linear transform the local sensor data compression.

At the fusion center, only linear unbiased estimation fusion is considered; that is, we consider the most commonly used linear estimation method: best linear unbiased estimation (BLUE). It is also known as linear minimum mean-square error (LMMSE), linear minimum variance (LMS), or linear unbiased minimum variance (LUMS) estimation. It is defined by, for the available information y at the fusion center,

$$\hat{x} = E^*(x|y) = \bar{x} + C_{xy} C_y^{-1}(y - \bar{y})$$
$$\text{MSE}(\hat{x}) = E[(x - \hat{x})(x - \hat{x})'] = C_x - C_{xy} C_y^{-1} C_{xy}'$$

where $\bar{x} = E(x)$, $C_x =\text{cov}(x)$, $C_{xy} =\text{cov}(x, y)$ and $C_y =\text{cov}(y)$ are prior information. If the inverse C_y^{-1} does not exist, it can be simply replaced with the unique *Moore-Penrose pseudoinverse* (MP inverse in short) C_y^+. With our sensor data compression, estimation fusion at the center can only base on the compressed data $y = [y_1', \ldots, y_n']'$, which is

$$y = A' z$$

where

$$A = \text{diag}[A_1, \ldots, A_n]$$
$$z = [z_1', \cdots, z_n']'$$

Now $C_{xy} = C_{xz}A$ and $C_y = A'C_z A$, and then

$$\hat{x} = \bar{x} + C_{xz}A(A'C_z A)^{-1}A'(z - \bar{z}) \qquad (1)$$
$$\text{MSE}(\hat{x}) = C_x - C_{xz}A(A'C_z A)^{-1}A'C'_{xz}$$

where

$$C_{xz} = \text{cov}(x, z) = [C_{xz_1}, \cdots, C_{xz_n}]$$

$$C_z = \text{cov}(z) = \begin{bmatrix} C_{z_1} & C_{z_1 z_2} & \cdots & C_{z_1 z_n} \\ C_{z_2 z_1} & C_{z_2} & & C_{z_2 z_n} \\ \vdots & & \ddots & \vdots \\ C_{z_n z_1} & & & C_{z_n} \end{bmatrix}$$

The problem of optimal sensor data compression, i.e., finding the optimal matrix $\tilde{A}$ which satisfies the dimensional requirement, is to solve the following constrained optimization problem:

$$A = \arg\min\{\text{MSE}(\hat{x})\}$$
$$= \arg\min\{C_x - C_{xz}A(A'C_z A)^{-1}A'C'_{xz}\}$$
$$= \arg\min\{\text{tr}[C_x - C_{xz}A(A'C_z A)^{-1}A'C'_{xz}]\}$$
$$\qquad (2)$$
$$= \arg\max_A J(A)$$
$$s.t.\ A = \text{diag}[A_1, \ldots, A_n] \text{ and } \dim(A_i) = n_i \times d_i$$

where

$$J(A) = \text{tr}[C_{xz}A(A'C_z A)^{-1}A'C'_{xz}] \qquad (3)$$

The third equality in (2) holds because of the uniqueness of optimal estimation $\hat{x}$ corresponding to the optimal $\tilde{A}$ [4]. (We will clearly see this in the following section.) If $A = I$, $y = [z'_1, z'_2, \ldots, z'_n]'$, raw data are sent to the fusion center. This is the centralized fusion and the estimation is globally optimal based on all measurements observed from local sensors. Generally

$$J(A) \le J(I) = \text{tr}[C_{xz}C_z^+ C'_{xz}] \qquad (4)$$

The equality in (4) holds if and only if the sensor data compression has no information loss for BLUE fusion.

3 Optimal Data Compression of a Single Sensor for Estimation

Suppose estimation is only based on the compressed data from a single sensor. In this system, the local sensor collects and processes the data. The compressed data are sent to the center for estimation. We first discuss this case because it is a good starting point to derive the optimal solution.

In this case, $n = 1$, $z = z_1$ and $A' = A'_1$ and thus $y = y_1 = A'_1 z_1$ compresses the data from dimension n_1 to d_1 ($d_1 < n_1$). According to (2), the optimal

compression is the solution of the following constrained optimization problem.

$$A = \arg\max_A J(A) \qquad (5)$$
$$= \arg\max_A \text{tr}[C_{xz}A(A'C_z A)^{-1}A'C'_{xz}]$$
$$s.t.\ \dim(A_1) = n_1 \times d_1$$

As before, $(A'C_z A)^{-1}$ can be replaced with the MP inverse $(A'C_z A)^+$ if the inverse does not exist. In general, however, we can always find some A such that $A'C_z A$ is nonsingular: when $A'C_z A$ is not invertible, we can always find a new $\bar{A}$ with a lower dimension, but $\bar{A}'C_z\bar{A}$ becomes the nonsingular matrix.

Lemma 3.1: If $A'C_z A$ is singular, we can find a lower dimensional matrix $\bar{A}$ such that $\bar{A}'C_{z_1}\bar{A}$ is invertible and satisfies

$$J(\bar{A}) = J(A)$$

Proof: Omitted.

Lemma 3.2: For any nonsingular matrix D, we have

$$J(A) = J(AD)$$

Proof: Omitted.

Based on Lemma 3.1, in the following, we only consider the case that $A'C_z A$ is invertible.

Theorem 3.1: The optimal solution $\tilde{A}$ of (5) is

$$\tilde{A} = C_z^+ \bar{K} \qquad (6)$$

where the column vectors of $\bar{K}$ are the eigenvectors corresponding to the d_1 largest eigenvalues of $(C'_{xz}C_{xz}C_z^+)$, and J is the sum of these d_1 eigenvalues.

Proof: Omitted.

Based on Lemma 3.2., the optimal solution of (5) is not unique[3], because if $\tilde{A}$ is an optimal solution, $\tilde{A}D$ is another optimal solution for any nonsingular matrix D. However different optimal solutions $\tilde{A}$ and $\tilde{A}D$ correspond to the same estimator:

$$E^*(x|\tilde{A}z) = E^*(x|\tilde{A}Dz) \qquad (7)$$

This can be shown easily from the definition of BLUE estimator (1). It can be shown that any two elements A and $\tilde{A}$ in the optimal solution set Ω are related by $A = \tilde{A}B$ for some nonsingular matrix B. So (7) verifies the uniqueness of the optimal estimator $\hat{x}$ corresponding to Ω.

Suppose $\text{rank}(C'_{xz}C_{xz}C_z^+) = m$, $m \le \min\{n_1, n_x\}$, where n_1 is the dimension of observation z_1; n_x is the dimension of estimatee x. So $C'_{xz}C_{xz}C_z^+$ has only m nonzero eigenvalues, which means that if the dimensional requirement for the sensor is larger than m, we can always project the observation into a subspace of

dimension m in that we can choose the optimal A such that $d_1 = m$. It should be realized that there is no information loss with this data compression. Also, $m = \min\{n_1, n_x\} = n_1$ means the observation dimension is not larger than the dimension of the estimatee. Since $J(A)$ is the sum of the d_1 largest eigenvalues and $J(I)$ is the sum of all nonzero n_1 eigenvalues of $(C'_{xz} C_{xz} C_z^+)$, if we compress the observations with $d_1 < m \leq n_1$, then $J(A) < J(I)$. In this situation, we can not compress the data without information loss (performance deterioration). If $d_1 < m$, there is information loss even for the optimal compression $y = A'z$, that is, the globally optimal estimation is not achievable using the optimally compressed data:

$$J(A) < J(I)$$

Note that sensor data compression is optimal in that the most important information in the data is extracted for estimating x at the fusion center.

4 Optimal Data Compression for BLUE Fusion

For multiple-sensor estimation fusion $(n > 1)$, the linear mapping $y = A'z$ compresses each local sensor measurement from dimension n_i to d_i $(d_i < n_i)$, $i = 1, \ldots, n$, where $z = [z'_1, z'_2, \ldots, z'_n]'$ and $A = \mathrm{diag}[A_1, \ldots, A_n]$. In this section, we first discuss three special cases which have simple optimal solutions. The most general and difficult case is discussed last.

4.1 Uncorrelated sensors

In this case, $C_{z_i z_j} = 0$ for any $i \neq j$ where $i, j = 1, \ldots, n$. Since C_z and A are all block diagonal matrices,

$$A'C_z A = \mathrm{diag}[A'_1 C_{z_1} A_1, \ldots, A'_n C_{z_n} A_n]$$

we have

$$(A'C_z A)^{-1} = \mathrm{diag}[(A'_1 C_{z_1} A_1)^{-1}, \ldots, (A'_n C_{z_n} A_n)^{-1}]$$

Also

$$
A'C'_{xz} C_{xz} A
$$
$$
= \begin{bmatrix} A'_1 C'_{xz_1} C_{xz_1} A_1 & \cdots & A'_1 C'_{xz_1} C_{xz_n} A_n \\ \vdots & \ddots & \vdots \\ A'_n C'_{xz_n} C_{xz_1} A_1 & \cdots & A'_n C'_{xz_n} C_{xz_n} A_n \end{bmatrix}
$$

According to the property of trace, the objective function (3) can be rewrite as

$$
J(A) = \mathrm{tr}[A'C'_{xz} C_{xz} A (A'C_z A)^{-1}]
$$
$$
= \sum_{i=1}^{n} \mathrm{tr}[A'_i C'_{xz_i} C_{xz_i} A_i (A'_i C_{z_i} A_i)^{-1}]
$$
$$
= \sum_{i=1}^{n} J_i(A_i)
$$

where

$$J_i(A_i) = \mathrm{tr}[A'_i C'_{xz_i} C_{xz_i} A_i (A'_i C_{z_i} A_i)^{-1}]$$

Now the constrained optimization problem (2) concerning $J(A)$ can be divided into the following n individual constrained optimization problems

$$A_i = \arg\max_{A_i} J_i(A_i), \; i = 1, 2, \ldots, n \quad (8)$$
$$s.t. \; \dim(A_i) = n_i \times d_i$$

For each A_i, the optimization problem is the same as the single sensor case, and so we have the optimal solution $\tilde{A}_i = C_{z_i}^+ \bar{K}_i$, where the column vectors of $\bar{K}_i$ are the eigenvectors corresponding to the d_i largest eigenvalues of $C'_{xz_i} C_{xz_i} C_{z_i}^+$ or $\tilde{A}D$ with any nonsingular matrix D, and J_i is the sum of the d_i eigenvalues.

4.2 Perfectly correlated sensors

In this case, $C_{z_i z_j} = C_{z_1}$ and $C_{xz_j} = C_{xz_1}$ for $i, j = 1, \ldots, n$. It happens if local sensors receive statistically the same observations. Now

$$
A'C_z A = \begin{bmatrix} A'_1 \\ \vdots \\ A'_n \end{bmatrix} C_{z_1} \begin{bmatrix} A'_1 \\ \vdots \\ A'_n \end{bmatrix}'
$$

$$
A'C'_{xz} C_{xz} A = \begin{bmatrix} A'_1 \\ \vdots \\ A'_n \end{bmatrix} C_{xz_1}^T C_{xz_1} \begin{bmatrix} A'_1 \\ \vdots \\ A'_n \end{bmatrix}'
$$

Let $A_{\mathrm{new}} = [A_1, \cdots, A_n]$. Then the objective function (3) is

$$J(A_{\mathrm{new}}) = \mathrm{tr}[A'_{\mathrm{new}} C'_{xz_1} C_{xz_1} A_{\mathrm{new}} (A'_{\mathrm{new}} C_{z_1} A_{\mathrm{new}})^{-1}]$$

and the constrained optimization problem is

$$A_{\mathrm{new}} = \arg\max_{A_{\mathrm{new}}} J(A_{\mathrm{new}}) \quad (9)$$
$$s.t. \; \dim(A_{\mathrm{new}}) = n_1 \times d$$

where $d = \sum_{i=1}^{n} d_i$. Now the optimization problem (9) is the same as the single sensor case (5). So the optimal solution is $\tilde{A}_{\mathrm{new}} = C_{z_1}^+ \bar{K}_1$, where the column

vectors of $\bar{K}_1$ are the eigenvectors corresponding to the d largest eigenvalues of $C'_{xz_1}C_{xz_1}C^+_{z_1}$ or $\tilde{A}D$ with any nonsingular matrix D, and J is the sum of the d eigenvalues.

It is obvious that when the total dimensional requirement of all sensors (i.e., d) is not smaller than $m = \text{rank}(C'_{xz_1}C_{xz_1}C^+_{z_1})$, there is no information loss for the estimation at the fusion center.

4.3 Sensor dimensional requirement larger than rank of $C'_{xz}C_{xz}C^+_z$

When the dimensional requirement for each sensor is equal to or larger than $m = \text{rank}(C'_{xz}C_{xz}C^+_z)$, we can simply create the sensor data compression $y_i = A'_i z_i$ such that $d_i = m$ $(i = 1, \ldots, n)$. We will see there is no information loss for estimation fusion at the center with this sensor data compression.

Since all A_i have the same number of columns, we can define
$$A_{\text{new}} = [A'_1, A'_2, \ldots, A'_n]'$$
Then the objective function (3) is equivalent to
$$J(A_{\text{new}}) = \text{tr}[A'_{\text{new}}C'_{xz}C_{xz}A_{\text{new}}(A'_{\text{new}}C_z A_{\text{new}})^{-1}]$$
Now the constrained optimization problem for this case is the same as for the single sensor case, that is,
$$A_{\text{new}} = \arg\max_A J(A_{\text{new}})$$
$$s.t. \ \dim(A_{\text{new}}) = (\sum_{i=1}^n n_i) \times m$$

Then the optimal solution is $\tilde{A}_{\text{new}} = C^+_z \bar{K}$, where the m column vectors of $\bar{K}$ are the eigenvectors corresponding to the m nonzero eigenvalues of $C'_{xz}C_{xz}C^+_z$ or $\tilde{A}D$ with any nonsingular matrix D, and J is the sum of the m eigenvalues. Obviously $J(A_{\text{new}}) = \text{tr}(C'_{xz}C_{xz}C^+_z) = J(I)$, and so there is no information loss with this sensor data compression.

Observations may have a higher dimension than that of the estimatee. So there is no estimation accuracy degradation if all sensors compress their observations to the dimension of the estimatee. However this data compression needs to consider correlation between sensors. This means that if a sensor compresses its observations by only considering its local information, generally there is information loss. This is related to the fact that the standard distributed fusion can not achieve the same performance as the centralized fusion in many cases [5].

4.4 Arbitrary sensor dimensional requirement

The general case has an arbitrary sensor dimensional requirement. The sensors have different dimensional requirements according to the system restriction. In particular, some have a demanding dimensional requirement of $d_i < m = \text{rank}(C'_{xz}C_{xz}C^+_z)$. So we can not construct the same sensor data compression as above, and data compression generally has information loss.

By (2), the optimal sensor compression is the solution of the following constrained optimization problem:
$$A = \arg\max_A J(A)$$
$$A = \text{diag}[A_1, \ldots, A_n] \text{ and } \dim(A_i) = n_i \times d_i$$
where
$$J(A) = \text{tr}[C_{xz}A(A'C_z A)^{-1}A'C'_{xz}]$$

Although the objective function $J(A)$ has the same form as the single-sensor case, we can not directly borrow the solution there, because there is one more constraint for the matrix A which requires A to be block diagonal. In the single-sensor case, if we look at the objective function $J(A_1)$ as a single variable function of A_1, the objective function $J(A_1, A_2, \ldots, A_n)$ for the n-sensor case should be a multivariate function. Now the problem become a multivariable optimization problem.

For each $i = 1, \ldots, n$, we denote
$$y_{(i)} = [y'_1, \ldots, y'_{i-1}, y'_{i+1}, \ldots, y'_n]'$$
$$z_{(i)} = [z'_1, \ldots, z'_{i-1}, z'_{i+1}, \ldots, z'_n]'$$
$$A_{(i)} = \text{diag}[A_1, \ldots, A_{i-1}, A_i, \ldots, A_n]$$

At the fusion center, based on the BLUE fusion and the recursive BLUE formulas, we have
$$\hat{x}^{\text{BLUE}} = E^*(x|y) = E^*(x|y_{(i)}, y_i)$$
$$= E^*(x|y_{(i)}) + C_{x\tilde{y}_{i|y_{(i)}}}C^{-1}_{\tilde{y}_{i|y_{(i)}}}\tilde{y}_{i|y_{(i)}}$$
$$\text{MSE}(\hat{x}^{\text{BLUE}}) = C_{x|y_{(i)}} - C_{x\tilde{y}_{i|y_{(i)}}}C^{-1}_{\tilde{y}_{i|y_{(i)}}}C'_{x\tilde{y}_{i|y_{(i)}}}$$
where $\tilde{y}_{(i)} = y_{(i)} - \bar{y}_{(i)}$, $\tilde{z}_{(i)} = z_{(i)} - \bar{z}_{(i)}$, $\bar{y}_{(i)} = E(y_{(i)})$, $\bar{z}_{(i)} = E(z_{(i)})$, and
$$\tilde{y}_{i|y_{(i)}} = A'_i\tilde{z}_{i|y_{(i)}}$$
$$\tilde{z}_{i|y_{(i)}} = z_i - E^*(z_i|y_{(i)})$$
$$= z_i - \bar{z}_i - C_{z_i y_{(i)}}C^{-1}_{y_{(i)}}\tilde{y}_{(i)}$$
$$= (z_i - \bar{z}_i) - C_{z_i z_{(i)}}A_{(i)}(A'_{(i)}C_{z_{(i)}}A_{(i)})^{-1}A'_{(i)}\tilde{z}_{(i)}$$

So
$$\text{MSE}(\hat{x}^{\text{BLUE}})$$
$$= C_{x|y_{(i)}} - C_{x\tilde{z}_{i|y_{(i)}}}A_i[A'_iC_{\tilde{z}_{i|y_{(i)}}}A_i]^{-1}A'_iC'_{x\tilde{z}_{i|y_{(i)}}}$$

where

$$C_{x|y_{(i)}} = C_x - C_{xz_{(i)}}A_{(i)}(A'_{(i)}C_{z_{(i)}}A_{(i)})^{-1}A'_{(i)}C'_{xz_{(i)}}$$

$$C_{x\tilde{z}_{i|y_{(i)}}} = C_{xz_i} - C_{xz_{(i)}}A_{(i)}(A'_{(i)}C_{z_{(i)}}A_{(i)})^{-1}A'_{(i)}C'_{z_iz_{(i)}}$$

$$C_{\tilde{z}_{i|y_{(i)}}} = C_{z_i} - C_{z_iz_{(i)}}A_{(i)}(A'_{(i)}C_{z_{(i)}}A_{(i)})^{-1}A'_{(i)}C'_{z_iz_{(i)}}$$

Let

$$J_i(A_i) = \text{tr}[C_{x\tilde{z}_{i|y_{(i)}}}A_i(A'_iC_{\tilde{z}_{i|y_{(i)}}}A_i)^{-1}A'_iC'_{x\tilde{z}_{i|y_{(i)}}}]$$

Then

$$\begin{aligned}J(A) &= J(A_{(i)}, A_i) \\ &= \text{tr}[C_{xz_{(i)}}A_{(i)}(A'_{(i)}C_{z_{(i)}}A_{(i)})^{-1}A'_{(i)}C'_{xz_{(i)}}] \\ &\quad + J_i(A_i)\end{aligned}$$

The objective function J is a multivariate function of $A_1, \ldots, A_n$. According to the necessary conditions for its extreme points, A_i must be the optimal solution of the following optimization problem

$$A_i = \arg\max_{A_i} J_i(A_i)$$
$$s.t. \ \dim(A_i) = n_i \times d_i$$

This problem is same as that for the single-sensor case. Then the optimal solution is

$$\tilde{A}_i = C^+_{\tilde{z}_{i|y_{(i)}}}\bar{K}_i$$

where the column vector of $\bar{K}_i$ is the d_i eigenvectors corresponding to the d_i largest eigenvalues of $C'_{x\tilde{z}_{i|y_{(i)}}}C_{x\tilde{z}_{i|y_{(i)}}}C^+_{\tilde{z}_{i|y_{(i)}}}$. Note, however, that the optimal solution $\tilde{A}_i$ depends on the value of $A_{(i)}$ through $\tilde{z}_{i|y_{(i)}}$. It is not easy to give an explicit solution for each A_i. In the following, we give an iterative algorithm to search for the optimal $\tilde{A}$.

Define the operator $\Gamma = (\Gamma_1, \ldots, \Gamma_n)$, for $i = 1, \ldots, n$

$$\Gamma_i(A_1, \ldots, A_{i-1}, A_{i+1}, \ldots, A_n) = C^+_{\tilde{z}_{i|y_{(i)}}}\bar{K}_i$$

Then we can construct a Gauss-Seidel iteration to search for the optimal solution $(A_1, \ldots, A_n)$. Suppose the nonzero initial value is $(A_1^{(0)}, \ldots, A_n^{(0)})$. At each iteration $k = 1, 2, \ldots$,

$$\begin{aligned}A_1^{(k+1)} &= \Gamma_1(A_2^{(k)}, \ldots, A_n^{(k)}) \\ A_2^{(k+1)} &= \Gamma_2(A_1^{(k+1)}, A_3^{(k)}, \ldots, A_n^{(k)}) \qquad (10) \\ &\vdots \\ A_n^{(k+1)} &= \Gamma_n(A_1^{(k+1)}, \ldots, A_{n-1}^{(k+1)})\end{aligned}$$

After each iteration (10), we have

$$A^{(k)} = \text{diag}[A_1^{(k)}, \ldots, A_n^{(k)}]$$

The iteration stops once the objective function $J(A)$ satisfies

$$J(A^{(k+1)}) - J(A^{(k)}) < \epsilon$$

where ϵ is some predetermined small number.

Since each step maximizes one term of J while fixing the other terms, we have, for each $i = 1, \ldots n$,

$$\begin{aligned}&J(A_1^{(k+1)}, \ldots, A_i^{(k+1)}, A_{i+1}^{(k)}, \ldots, A_n^{(k)}) \\ &\geq J(A_1^{(k+1)}, \ldots, A_{i-1}^{(k+1)}, A_i^{(k)}, \ldots, A_n^{(k)})\end{aligned}$$

This implies,

$$J(A^{(k)}) \geq J(A^{(k-1)})$$

The equality holds if and only if $A_i^{(k+1)} = A_i^{(k)}D_i$, where D_i is any nonsingular matrix. Thus the sequence $\{J(A^{(k)})\}$ is monotonically increasing.

Lemma 4.1 Function $J(A)$ has an upper bound.

Proof: Omitted.

Since $\{J(A^{(k)})\}$ is monotonically increasing and has an upper bound, it has a limit ζ:

$$\lim_{k \to \infty} J(A^{(k)}) = \zeta$$

Combined with the continuity of function $J(A^{(k)})$ without considering the point at which $A'C_zA$ becomes singular, there exists $\tilde{A}$ such that

$$\lim_{k \to \infty} J(A^{(k)}) = J(\tilde{A})$$

So $\tilde{A}$ is a limit of $\{A^{(k)}\}$.

Theorem 4.1 The limit point $\tilde{A}$ in Theorem 4.1 is a stationary point of the objective function, i.e., $\frac{\partial}{\partial A_i}J(A)|_{A=\tilde{A}} = 0$ for $i = 1, \ldots, n$.

Proof: Omitted.

Theorem 4.1 is important. It implies that when the iteration ends, the solution is a stationary point of the objective function J. It may be a maximizer because extreme points are critical points. Unfortunately we are not able to provide further theoretical results concerning the convergence of $\{A^{(k)}\}$ to the globally optimal solution at this stage, because the objective function $J(A)$ is too complex to analyze for block diagonal matrix A. However, we found from simulation that almost every time we achieve the globally optimal solution, which means starting from different initial points, the iteration will end with the same value of J and the same estimator $\hat{x}$.

5 Simulation

Several simple numerical examples are given in this section to verify formulas presented and the optimality of our Gauss-Seidel iteration based search solutions. All examples are for the following multi-sensor target tracking system:

We consider a constant-velocity moving target in 2-dimensional x-y space. The estimatee $x(t)$ is the state process, consisting of position and velocity components: $x = [\mathsf{x}, \dot{\mathsf{x}}, \mathsf{y}, \dot{\mathsf{y}}]'$. We set up 6 observation stations S_i, $i = 1, ..., 6$, of 3 types. The first two stations have a linear observation model: $z_i(t) = H_i x(t) + v_i(t)$, where $H_i = I_4$. The third and fourth stations have the same linear observation model but with $H_i = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$, that is, they only observe position of the target. The last two stations have a nonlinear observation model:

$$z_i(t) = \begin{bmatrix} \sqrt{\mathsf{x}(t)^2 + \mathsf{y}(t)^2} \\ \tan^{-1}(\frac{\mathsf{y}(t)}{\mathsf{x}(t)}) \end{bmatrix} + v_i(t)$$

In the above, $v_i(t)$ is zero-mean white noise with covariance $R_i = \mathrm{cov}(v_i(t))$, $i = 1, .., 6$.

$$R_1 = 10 \begin{bmatrix} 1000 & 20 & 10 & 5 \\ 20 & 100 & 5 & 10 \\ 10 & 5 & 4000 & 50 \\ 5 & 10 & 50 & 300 \end{bmatrix},$$

$$R_2 = 1000 \begin{bmatrix} 40 & 2 & 0 & 0 \\ 2 & 6 & 0 & 0 \\ 0 & 0 & 10 & 1 \\ 0 & 0 & 1 & 5 \end{bmatrix}$$

$$R_3 = 1000 \begin{bmatrix} 20 & 1 \\ 1 & 30 \end{bmatrix}, R_4 = 1000 \begin{bmatrix} 15 & 4 \\ 4 & 10 \end{bmatrix}$$

$$R_5 = \begin{bmatrix} 0.004 & 0 \\ 0 & 90000 \end{bmatrix}, R_6 = \begin{bmatrix} 1 & 0 \\ 0 & 1.6 \times 10^8 \end{bmatrix}$$

In this setting, the estimatee is 4-dimensional and the observations of the six sensors at each sampling interval T_i ($i = 1, \ldots, 6$) have the dimensions 4, 4, 2, 2, 2, 2, respectively. It is not realistic that each sensor's data transmitting rate is the same as sampling rate according to the channel capability. The most often case is that each sensor transmits information to the fusion center every N_i sampling intervals. Then the stacked observations sent by each sensor have dimensions $4N_1, 4N_2, 2N_3, 2N_4, 2N_5, 2N_6$, respectively. We use Monte Calro method to calculate the covariance matrices C_{xz} and C_z with conversion of data to the same time [5], since we use nonlinear observation models for sensors 5 and 6. In the following,

we try to get the optimal compression $y = A'z$ for several fusion systems and compare the mean-square error $\mathbf{mse}(\hat{x}(y)) = \mathrm{tr}(C_x - C_{xz}A(A'C_zA)^{-1}A'C'_{xz})$ with the minimum possible $\mathbf{mse}(\hat{x}(z)) = \mathrm{tr}(C_x - C_{xz}C_z^{-1}C'_{xz})$, in order to check if the compression lose information or not.

5.1 Single Sensor

In this case, we only use sensor 1, 3 or 5. Here $N_1 = N_3 = N_5 = 10$, which means every 10 observations are stacked in each sensor to be compressed. At the fusion center, we use the compressed data to estimate the state. By (6) we can get the optimal compression for each system with different data dimensional requirement d. In Table 1, we compare the $\mathbf{mse}$ for all cases: The estimation accuracy improves

Table 1: $\mathbf{mse}$ of estimates using compressed data from a single sensor

$\mathrm{mse}(\hat{x}(y))$	with S_1	with S_3	with S_5
$d = 1$	506.1702	511.8113	603.4013
$d = 2$	483.3577	498.4204	596.2738
$d = 3$	483.3381	498.4161	596.2643
$d = 4$	483.3308	498.4141	596.2609
$\mathrm{mse}(\hat{x}(z))$	483.3308	498.4141	596.2609

if data of a higher dimension are allowed to be sent. Since $\mathrm{rank}(C'_{xz}C_{xz}C_z^{-1}) = 4$, our analysis states that $\mathbf{mse}(\hat{x}(y)) = \mathbf{mse}(\hat{x}(z))$ if and only if $d \geq 4$ for the optimal compression $y = A'z$. This is verified by Table 1.

It is also interesting to note that the compression to $d = 2$ loses little estimation accuracies in terms of $\mathbf{mse}$.

5.2 Multi-Sensor Fusion System

In this case, we construct the fusion system by using some combinations of the 6 sensors. Vector $(d_1, d_2, d_3, d_4, d_5, d_6)$ denotes the dimensional requirement for the sensors. So $d_i = 0$ means sensor i does not send out any information. Here $N_i = 10$, that is, each sensor stacks 10 observations. The Gauss-Seidel iteration was used to get the optimal solution. In Table 2, we compare the $\mathbf{mse}$ for all cases, where $\hat{x}(z)$ denotes the centralized fusion by using all observations from all active ($d_i \neq 0$) sensors. We also label the required iteration steps for searching for the optimal solution with the proposed Gauss-Seidel algorithm to reach $\epsilon \leq 0.0001$.

In Fig.2, we plot for the $(1, 2, 3, 4, 1, 3)$ case the $\mathbf{mse}$ **ratio**$= \frac{\mathrm{mse}(\hat{x}(z))}{\mathrm{mse}(\hat{x}(y))}$ of the centralized fusion to distributed fusion versus iteration number k to demon-

Table 2: **mse** for multi-sensor system

$(d_1,d_2,d_3,$ $d_4,d_5,d_6)$	Gauss-Seidel Solution:mse$(\hat{x}(y))$	mse$(\hat{x}(z))$
$(1,1,1,1,0,0)$	369.6550 (step=11)	291.1871
$(2,2,2,2,0,0)$	291.2524 (step=9)	291.1871
$(3,3,3,3,0,0)$	291.2082 (step=9)	291.1871
$(1,2,3,4,0,0)$	308.1658 (step=9)	291.1871
$(4,4,4,4,0,0)$	291.1871 (step=8)	291.1871
$(1,1,1,1,1,1)$	348.9785 (step=12)	265.0382
$(2,2,2,2,2,2)$	265.1296 (step=13)	265.0382
$(3,3,3,3,3,3)$	265.0646 (step=13)	265.0382
$(1,2,3,4,1,3)$	289.2232 (step=12)	265.0382
$(4,4,4,4,4,4)$	265.0382 (step=12)	265.0382

compression rule for general multiple-sensor system with a sensor dimensional requirement.

Theoretically, our analysis concludes that there is no estimation accuracy degradation if the dimensional requirement for sensor data compression is not smaller than the dimension of the estimatee. But the BLUE fusion given by a standard distributed fusion system generally suffers from accuracy loss when local sensors are correlated.

Simulation results demonstrate that the algorithm for searching the optimal sensor compression rule for general multiple-sensor systems with Gaussion-Seidel iteration is efficient. In the future, we would like to provide further theoretical support for the optimality of the searching algorithm.

strate the convergence rate and optimality of the search method. The three lines are for three different initializations $A^{(0)}$.
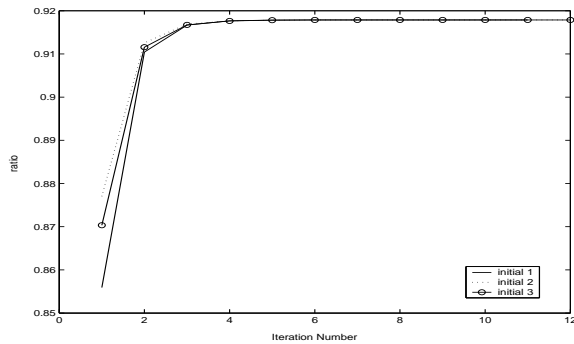

Figure 2: **mse ratio** for different initializations $A^{(0)}$

From Table 2 and Fig.2, we see that the Gaussian-Seidel iterative search for the optimal sensor data compression is efficient and yields the optimal solution, because it always converges to the same estimator for any initialization $A^{(0)}$, and the convergence rate is high.

6 Conclusions

In this paper, we propose that in a multi-sensor distributed estimation fusion system, the local sensor data processing should be based on the fusion rule at the center, the channel capacity between sensors and the center and the processing capability of the center. We formulate the system restriction as the sensor data dimensional requirement. Based on BLUE fusion, we present linear sensor compression rules. Explicit solution for local sensor data compression is given for single-sensor systems and some particular multiple-sensor systems. An algorithm based on Gaussion Seidel iteration is presented for searching the optimal

References

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

[2] K. C. Chang, S. Mori, and C. Y. Chong. Evaluating a Multiple-Hypothesis Multitarget Tracking Algorithm. *AES*, AES-30(2):578–590, Apr. 1994.

[3] F. Keinosuke. *Introduction to Statistical Pattern Recognition*. Academic Press, INC, San Diego, CA, USA, 1990.

[4] X. R. Li. *Applied Estimation and Filtering*. Course Note, 2002.

[5] X. R. Li and C. He. Optimal Initialization of Linear Recursive Filters. In *37*, pages 2335–2340, Tampa, FL, Dec. 1998.

[6] X. R. Li and K. S. Zhang. Optimal Linear Estimation Fusion—Part IV: Optimality and Efficiency of Distributed Fusion. In *Fusion 2001*, pages WeB1.19–WeB1.26, Montreal, QC, Canada, Aug. 2001.

[7] X. R. Li and P. Zhang. Optimal Linear Estimation Fusion—Part III: Cross-Correlation of Local Estimation Errors. In *Fusion 2001*, pages WeB1.11–WeB1.18, Montreal, QC, Canada, Aug. 2001.

[8] X. R. Li, Y. M. Zhu, and C. Z. Han. Unified Optimal Linear Estimation Fusion—Part I: Unified Models and Fusion Rules. In *Fusion 2000*, pages MoC2.10–MoC2.17, Paris, France, July 2000.

[9] M. Liggins, C. Y. Chong, I. Kadar, M. G. Alford, V. Vannicola, and S. Thomopoulos. Distributed Fusion Architectures and Algorithms for Target Tracking. *Proceedings of the IEEE*, 85(1):95–107, Jan. 1997.