Unified Optimal Linear Estimation Fusion—
Part I: Unified Models and Fusion Rules

X. Rong Li *
Dept. of Electrical Engineering
University of New Orleans
New Orleans, LA 70148
Phone: 504-280-7416, Fax: -3950
xli@uno.edu

Yunmin Zhu †
Dept. of Mathematics
Sichuan University
Chengdu, Sichuan
610064, P. R. China
ymzhu@scu.edu.cn

Chongzhao Han
Electronic & Information Engr
Xi’an Jiaotong University
Xi’an, Shaanxi
710049, P. R. China
czhan@xjtu.edu.cn

Abstract—This paper deals with data fusion for the purpose of estimation. Three fusion architectures are considered: centralized, distributed, and hybrid. A unified linear model and general framework for these three architectures are established. Optimal fusion rules in the sense of best linear unbiased estimation (BLUE), weighted least squares (WLS), and their generalized versions are presented for cases with either complete, incomplete, or no prior information. These rules are much more general and flexible than previous results. For example, they are in a unified form that are optimal for all the three fusion architectures with arbitrary correlation of local estimates or observation noises across sensors or across time. They are also in explicit forms convenient for implementation. The relationships among these rules are also presented.

Keywords: Estimation fusion, track fusion, fusion rules, BLUE, least squares, MMSE estimation

1 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 an unknown quantity—a parameter or process. These data sets are usually obtained from multiple sources (e.g., multiple sensors).

Evidently, estimation fusion has wide-spread applications since many practical problems involve data from multiple sources. One of the most important application areas of estimation fusion is target tracking using multiple (homogeneous or heterogeneous) sensors, where it is more commonly referred to as track fusion or track-to-track fusion. The multiple sets of data could also be collected from a single source over a time period. With this wider view of estimation fusion, the conventional problems of filtering, prediction, and particularly smoothing of an unknown process are special estimation fusion problems. In fact, some of the results presented in this paper have been applied to the development of optimal linear smoothers for linear systems with arbitrarily colored (not necessarily Markov) and cross-correlated process and measurement noise [12].

There are two basic fusion architectures: centralized and decentralized (or distributed) (referred to as measurement fusion and estimate fusion in target tracking, respectively), depending on whether raw data are sent to the fusion center or not. They have pros and cons in terms of optimality, channel requirements, reliability, survivability, information sharing, etc. Centralized fusion is nothing but a conventional estimation problem with distributed data. Distributed fusion is more challenging and has been a focal point of most fusion research for many years.

Estimation/track fusion has been investigated for more than two decades. Many results have been obtained (see, e.g., [9, 5, 14]). While these results do represent major progress, to the authors’ knowledge, they are still rather limited in several aspects. For example, it appears that a systematic approach to the development of distributed fusion rules is still lacking. Most of the existing optimal fusion formulas for distributed fusion were obtained by ingenious but ad hoc manipulations of formulas of local estimates so that they are equivalent to the corresponding centralized fusion formulas (see, e.g., [9, 4, 14]). Much of these results rely on the celebrated Kalman filters for the centralized fusion architecture. An obvious drawback of this ad hoc approach is that the fusion rules thus developed cannot be expected to have a wider applicability than that of the Kalman filter. The following are examples of the limitations of these fusion rules.

- Sensor observation errors have to be uncorrelated. This is a real limitation in many practical situations. For example, many sensor errors are dependent on the common random estimatee (the quantity to be estimated) and thus are correlated. Another example is that the estimatee is observed by multiple sensors in a common noisy environment, such as during noise jamming generated by a target.

- The local dynamic models have to be identical. In reality, local estimates may have been obtained based on different dynamic models of the estimatee to account for sensor-specific situations or requirements. For example, different sensors may use different

*Supported in part by ONR via grant N00014-00-1-0677, NSF via Grant ECS-9734285, LEQSF via Grant (1996–99)-RD-A-32.
†Supported in part by The National Key Project and Natural Science Foundation of China.
multiple models that are most suitable to take advantage of the merits of each local sensor.

- The network structure and information patterns of the distributed systems have to be simple. When a distributed system is too complicated, even a genius cannot devise a fusion rule by manipulating the local estimates such that it is equivalent to the corresponding centralized fusion rule.

It has been realized for many years following the original work of [2] that local estimates (tracks) have correlated errors. How to counter this cross correlation has been a central topic in distributed fusion. Several ingenious techniques, such as the tracklets technique [11, 10] and those based on the information graph [8], have been developed. Their main underlying idea is to reduce the correlation by smart processing and/or management at either the sensor or central level. It is, however, virtually impossible to annihilate this cross correlation completely in practical situations. Satisfactory approaches to distributed fusion in the presence of this cross correlation is still not available. In addition, these techniques cannot overcome the other limitations mentioned above.

In this paper, a general and systematic approach to estimation/track fusion is presented. It is much more general and flexible than the above-mentioned approach based on the Kalman filter. A variety of fusion rules in unified forms that are optimal in the linear class for centralized, distributed, and hybrid fusion architectures are presented, along with their relationships. The key to this development is simple: The local estimates are nothing but “observations” of the estimate. These rules are optimal for an arbitrary number of sensors in the presence of the above-mentioned cross correlation in the sense of either the weighted least-squares (WLS) or best linear unbiased estimation (BLUE)—i.e., linear minimum variance (LMS) or mean-square error (LMMSE) [4]. The results presented do not suffer from the drawbacks mentioned above and are very easy to be implemented.

The remainder of this paper is organized as follows. Section 2 formulates the fusion problems. Section 3 introduces a unified data model for centralized, distributed, and hybrid architecture. Various optimal fusion rules are presented in Section 4. Section 5 clarifies the relationships of these rules. Section 6 provides a summary.

2 Formulation of Estimation Fusion

For simplicity, consider a distributed system with a fusion center and n sensors (local stations), each connected to the fusion center. Denote by $Z_i$ the set of observations (and the covariances $R_i$ of its associated noises) of the ith sensor of the estimatee (i.e., the quantity to be estimated) $x$. Denote by $\hat{x}_i$ the ith sensor’s local estimate of $x$ and $P_i = \text{cov}(\hat{x}_i)$, the covariance of its associated estimation error $\hat{x}_i = x - \hat{x}_i$. Let $\hat{x}$ and $P = \text{cov}(\hat{x})$ be the estimate of $x$ and the covariance of its associated error $\hat{x}_f = x - \hat{x}$ at the fusion center, respectively.

The problem considered in this paper is the following: Obtain $\hat{x}$ and $P$ using all available information $Y = \{y_1, \ldots, y_n\}$ at fusion center.

Estimation fusion can be classified into three categories, depending on what information is available at the fusion center:

- If all unprocessed\(^1\) data (or observations) are made available directly at the fusion center, that is, if $Y = Z \triangleq \{Z_1, \ldots, Z_n\}$, then it is known as a centralized fusion, central-level fusion, or measurement fusion problem.

- If only data-processed observations $Z_i$ are available at the fusion center, that is, if $Y = D \triangleq \{g_i(Z_1), \ldots, g_n(Z_n)\}$ for non-trivial mappings $g_i(\cdot), \forall i$, then the problem is known as decentralized fusion, distributed fusion, sensor-level fusion, or autonomous fusion. In the most commonly used distributed fusion systems, $Y = \{\{\hat{x}_1, P_1\}, \ldots, \{\hat{x}_n, P_n\}\}$; that is, only local estimates $\{\hat{x}_i, P_i\}$ (based on $Z_i$) are available at the fusion center. Such an architecture is known as estimate fusion in target tracking and will be referred to as the standard distributed fusion for convenience. Note, however, that distributed fusion in general does not need to have $g_i(Z_i) = \{\hat{x}_i, P_i\}$.

- If the information available to the fusion center includes unprocessed data from some sensors and processed data from other sensors (i.e., if $Y \neq Z$ and $Y \neq D$), then it is referred to as a hybrid fusion problem.

The results presented in this paper are valid for many other systems or architectures, such as those in which there is no fusion center, but there is communication among some or all of the local sensors.

3 Unified Linear Fusion Model

3.1 The Unified Model

Let $z_i$ be any (multi-dimensional) observation of the ith sensor; that is, $z_i$ is a generic observation in $Z_i$. It is said that $z_i$ is a linear observation if it is an affine function of $x$; that is, if it satisfies the following linear equation

$$ z_i = h_i x + \eta_i \tag{1} $$

where $h_i$ is not a function of $x$ and $\eta_i$ is the observation noise.

The key to the development of the unified linear observation model of this paper for centralized, distributed, and hybrid fusion is the following: A local estimate can be viewed as an observation of the estimate using the following identity

$$ \hat{x}_i = x + (\hat{x}_i - x) = x + (-\hat{x}_i) \tag{2} $$

\(^1\)By “unprocessed” it is meant that no data processing has been done, although observation data are usually obtained after signal processing. In general, data are obtained by processing certain signals.
where the estimation error \( \hat{\epsilon}_i \) acts as the observation noise. It should be emphasized that this “observation” model of the distributed estimation fusion is valid for every case without exception—it does not rely on any assumption. For example, it is valid for every linear or nonlinear estimator \( \hat{x}_i \) and every linear or nonlinear observation \( z_i \). This model shall be referred to as the data model for standard distributed fusion.

Clearly, the above two equations can be rewritten in a unified form. More generally, however, consider also distributed systems in which \( z_i \) is processed “linearly” first and then sent to the fusion center; that is,

\[
\tilde{y}_i = g_i(z_i) = a_i + B_i z_i = B_i h_i x + \tilde{\eta}_i
\]

is available at the fusion center, where \( a_i \) and \( B_i \) are known and are not functions of observations in \( Z_i \). This model shall be referred to as the linearly-processed data model for distributed fusion. Clearly, the centralized fusion model (1) is a special case of this model with \( B_i = I \) and \( \tilde{\eta}_i = \eta_i \). Also, for a linear estimator \( \hat{x}_i \) with linear observation \( z_i \), (2) is a special case of (3) with \( B_i h_i = I \) and \( \tilde{\eta}_i = -\hat{x}_i \). However, (2) holds universally, not limited to linear case. Note also that the data \( y_i \) sent to the fusion center in a distributed system could also be a nonlinear function of \( z_i \). However, this is not considered in this paper, which deals only with linear estimation fusion.

It is now ready to describe the unified model. Let

\[
\begin{align*}
\text{CL} & = \text{centralized with linear observation} \\
\text{SD} & = \text{standard distributed} \\
\text{DL} & = \text{distributed with linearly-processed data}
\end{align*}
\]

\[
y_i = \begin{cases} 
z_i & \text{CL} \\
\hat{x}_i & \text{SD} \\
\hat{y}_i & \text{DL}
\end{cases}, \\
v_i = \begin{cases} 
\eta_i & \text{CL} \\
-\hat{x}_i & \text{SD} \\
\tilde{\eta}_i & \text{DL}
\end{cases}
\]

\[
H_i = \begin{cases} 
\tilde{h}_i & \text{CL} \\
I & \text{SD} \\
B_i h_i & \text{DL}
\end{cases}
\]

\[
y^n = \begin{bmatrix} y_1 \\
\vdots \\
y_n \end{bmatrix}, \quad H = \begin{bmatrix} H_1 \\
\vdots \\
H_n \end{bmatrix}, \quad v^n = \begin{bmatrix} v_1 \\
\vdots \\
v_n \end{bmatrix}
\]

Then, a unified linear model of the data available to the fusion center is

\[
y_i = H_i x + v_i
\]

The corresponding batch form for all data is

\[
y^n = H x + v^n
\]

Clearly, this model is valid not only for centralized and distributed fusion, but also for hybrid fusion. It is the foundation for the unified treatment of linear estimation fusion presented in this paper. The key to this unified model is to view each piece of data (in particular, \( \hat{x}_i \)) available at the fusion center as an observation of \( x \).

3.2 Correlation in the Unified Model

In general, noise components \( v_1, \ldots, v_n \) in the unified model are correlated; that is, the following matrix is not necessarily block diagonal:

\[
C = \text{cov}(v^n) = \begin{bmatrix} R & \Sigma \\
\Sigma & SD \\
R & DL \end{bmatrix}
\]

where \( R = \text{cov}(\eta_1, \ldots, \eta_n) \) is the observation noise covariance; \( \Sigma = \text{cov}(\hat{x}_1, \ldots, \hat{x}_n) \) is the joint error covariance of local estimates; and \( R = \text{cov}(\hat{\eta}_1, \ldots, \hat{\eta}_n) \) is the covariance of equivalent observation noise. It should be emphasized that the framework of the unified model requires that \( \Sigma \) be the a priori covariance before any data is made available. \( R \) is almost always assumed block-diagonal in the literature. Unfortunately, this is not true for many important practical situations. First, as shown in a forthcoming paper, the observation noises of a discrete-time multisensor system obtained by sampling a continuous-time multisensor system are correlated. Second, \( \eta_1, \ldots, \eta_n \) are in general correlated if \( x \) is observed in a common noisy environment, such as when a target state is observed in the presence of countermeasures (e.g., noise jamming) or atmospheric noise. This is particularly true when the sensors are on the same platform. Third, observation errors of many practical sensors are coupled through their dependence on \( x \). For example, the observation noise statistics is dependent on how far the target is because of difference in such quantities as cross-sections or SNR.

In general, \( \Sigma \) is almost never block-diagonal even if \( R \) is. This is easily understandable and is analogous to the fact that the estimation errors of a Kalman filter at different time instants are generally correlated even if the observation noise is white. A more detailed discussion is given in Part II [13].

Moreover, the noise \( v \) and the estimate \( x \) (if \( x \) is random) could be correlated in general:

\[
\text{cov}(x, v^n) = A_i, \quad \text{cov}(x, v_i) = A_i, \quad i \geq 1
\]

For example, for the standard distributed system with optimal linear unbiased local estimates with prior mean \( \tilde{x} \), one has

\[
A_i = \text{cov}(x, v_i) = \text{cov}(\hat{x}_i + \hat{x}_i - x), \tilde{x}_i \\
= -\text{cov}(\hat{x}_i) - E[\hat{x}_i \hat{x}_i'] + E[\hat{x}_i] = -P_i
\]

where the last equation above follows from the orthogonality principle.

In summary, (4)-(5) is indeed a unified data model for all linear centralized, distributed, and hybrid fusion problems. However, the price paid is that the observation noises may be correlated with each other and with the estimate. As a consequence, any optimal fusion rule based on this unified model must be valid for each correlation.

3.3 An Alternative Distributed Fusion Architecture

It is clear that the standard distributed fusion model (2) is universally valid. However, it leads to correlation among the equivalent noises of different sensors and between the noise and the estimatee. These two types of correlation may be avoided by using instead the linearly-processed data
model (3) for distributed fusion. This model is, however, not universally valid. It works only for linear observations with the uncorrelated actual observation noises under some assumed structure of the local estimators.

Assume that \( \hat{x}_i \) is given in the following form
\[
\hat{x}_i = \bar{x}_i + K_i[Z_i - h_i \bar{x}_i] = [I - K_i h_i] \bar{x}_i + K_i z_i
\]
with linear observations \( z_i = h_i x + \eta_i \). Note that this holds true for all linear unbiased estimators (not necessarily optimal in the linear class) with linear observations. Then
\[
y_i = \hat{x}_i - [I - K_i h_i] \bar{x}_i = K_i h_i x + K_i \eta_i = B_i h_i x + v_i
\]
is a linearly processed observation. Note that the covariance of the corresponding noise \( v \) is
\[
C = \text{cov}(v) = [C_{ij}] = [K_i R_{ji} K_j'] = K R K'
\]
where \( R = \text{cov}(\eta) \) and \( K = \text{diag}(K_1, \ldots, K_n) \). Thus, \( C \) is block-diagonal provided \( R \) is block-diagonal. This avoids the difficulty associated with the unified model based on the standard distributed model, which has a non-block-diagonal \( C \). In this model, in addition to the data \( y_i = \hat{x}_i - [I - K_i h_i] \bar{x}_i \), gain \( K_i \) needs to be sent to or computed at the fusion center.

This distributed architecture will be referred to as the simple non-standard distributed fusion.

4 Unified Optimal Fusion Rules

Only linear unbiased estimation fusion will be considered; that is, it is assumed in this paper that \( \hat{x}_i \) is linear unbiased estimators of \( x \) and only linear and unbiased fusion rules \( \hat{x} \) will be considered.

Consider two most commonly used types of linear estimation methods: optimal (linear) weighted least squares (WLS) and best linear unbiased estimation (BLUE). The latter is also known as linear minimum mean-square error (LMMSE), linear minimum variance (LMS), or linear unbiased minimum variance (LUMS) estimation. They are defined by, for the unified model (5),
\[
\hat{x}_{\text{BLUE}}^{\text{BLUE}} = \arg \min_{\hat{x}} E[(x - \hat{x})(x - \hat{x})'|Y]
\]
\[
\hat{x}_{\text{WLS}} = \arg \min_{\hat{x}} (y^n - H \hat{x})' C^{-1} (y^n - H \hat{x})
\]
where \( a \) and \( B \) do not depend on data \( Y \). They minimize mean-square error and fitting error, respectively.

When multiple solutions \( \hat{x} \) of the minimization problem \( \min_{\hat{x}} (y^n - H \hat{x})' C^{-1} (y^n - H \hat{x}) \) exist, it is assumed that the one with the minimum norm (i.e., smallest length \( \| \hat{x} \| \)) is always chosen since this solution is the one that has smallest error covariance \( E[(x - \hat{x})(x - \hat{x})'] \) and is unique. As a consequence, there is no need to assume that \( H' C^{-1} H \) is nonsingular. However, it is normally assumed that \( C^{-1} \) is positive definite (and thus nonsingular) for least squares fusion, otherwise the least squares approach is very questionable—zero fitting error does not imply zero estimation error. This assumption will not be stated every time it is needed.

In the case where \( C \) is block-diagonal (and \( A = 0 \)), both batch and recursive forms of these two estimators are simple and well-known. To be valid for the general case, in particular for the standard distributed fusion, it is most desirable to develop BLUE and WLS fusers that are valid for observation noises that are correlated with each other and with the estimatee.

4.1 BLUE Fusion with Complete Prior Knowledge

For convenience, it is said that a BLUE fusion problem has complete prior knowledge if both the prior mean \( \bar{x} \) and the prior covariance \( P_0 \) of the estimatee (as well as its correlation \( A \) with the observation error \( v \)) are known. Because the only prior information of the estimatee used by a BLUE estimator is its first two moments. The following theorem presents the BLUE fusion rule in the case of complete prior knowledge.

**Theorem 1.** Using data from (5), the unique, unified, optimal, BLUE fuser (10) of \( x \) with prior knowledge \( \bar{x} = E[x], P_0 = \text{cov}(x), \) and \( A = \text{cov}(x, v^n) \) is
\[
\hat{x} = \bar{x} + K[y^n - H \bar{x}] \tag{12}
\]
\[
P = \text{cov}(x - \hat{x}|Y) = P_0 - K S K'
\]
where
\[
S = H P_0 H' + C + H A + (H A)' \tag{14}
\]
\[
K = (P_0 H' + A) S^{-1} \tag{15}
\]
The error covariance and the gain matrices have the following alternative forms
\[
P = U P_0 U' + K C K' - U A K' - [U A K']' \tag{16}
\]
\[
K = (P H' + A)(C + H A)^{-1} \tag{17}
\]
where the inverse is assumed to exist and \( U = I - K H \).

In the case where \( S \) is singular, Theorem 1 is still valid with \( S^{-1} \) replaced by \( S^+ \), the Moore-Penrose pseudoinverse of \( S \). Note, however, that (17) is no longer valid in this case. This BLUE fuser reduces to the familiar BLUE estimator with uncorrelated errors if \( A = 0 \).

4.2 BLUE Fusion Without Prior Knowledge

The BLUE fusion rule of Theorem 1 is not valid if either there is no prior knowledge about the estimatee or the knowledge is incomplete (e.g., the prior covariance is not known or does not exist). In either case, the popular BLUE estimation formulas are not applicable. However, the BLUE fusion rules as defined by (10) for these cases still exist. They are presented as Theorems 2 and 3 in this and the next subsections.

**Theorem 2.** The unique, unified, optimal, BLUE fuser (10) of \( x \) using data from (5) without prior knowledge is
\[
\hat{x} = K y^n = \bar{K} y^n \tag{18}
\]
\[
P = K C K' = \bar{K} C \bar{K}' \tag{19}
\]
\[
\bar{K} = H^+ [I - C (T C T')^+] \tag{20}
\]
where the optimal gain matrix \( K \) is given uniquely by
\[
K = \hat{K} = H^+[I - C(TCT)^+] = H^+[I - CT\hat{T}(TCT)^+]^{-1}T^+
\]
if and only if \([H, C\hat{T}]\) has full row rank, where \( T \) is a full-rank square-root matrix of \( T = I - HH^+ \) and \( C\hat{T} \) is any square-root matrix of \( C \). Otherwise,
\[
K = \hat{K} + \theta T
\]
where \( \theta \) is any matrix of compatible dimension satisfying \( C\hat{T}T\theta = 0 \) and the superscript \( + \) stands for the unique Moore-Penrose pseudoinverse.

Note that an alternative form of \( \hat{K} \) is
\[
\hat{K} = PH'C^{-1}
\]
if and only if \( C \) is nonsingular. This is the same as (17) with \( A = 0 \).

For the standard distributed BLUE fusion without prior information, one has a clear expression for each element \( \hat{K}_{ij} \) of \( \hat{K} = [\hat{K}_1, \ldots, \hat{K}_n] \), where \( H = [I, \ldots, I] \) and thus \( H^+ = \frac{1}{n}H \). Denote by \( M_{ij} \) the \((i, j)\)th \( n \times n \) submatrix of \([I - C(TCT)^+] = [M_{ij}]\). Then \( \hat{K}_{ij} \) is the average of \( M_{ij} \) over the \( j \)th column and thus the estimate follows
\[
\hat{K}_{ij} = \frac{1}{n} \sum_{i=1}^{n} M_{ij}
\]

\[
\hat{x} = \hat{K}y^n = \sum_{j=1}^{n} \hat{K}_j y_j, \quad P = \hat{K}C\hat{K}^T
\]

### 4.3 BLUE Fusion With Incomplete Prior Knowledge

In practice, it is sometimes more desirable to define a singular \( P_0^{-1} \) and thus the corresponding \( P_0 \) does not exist\(^2\). This would be the case if prior information about some but not all components of \( x \) were not available. For example, when tracking an aircraft just taking off from an airport, it is easy to determine the prior velocity vector with certain covariance, but not the prior position vector. A practical means of specifying such knowledge is to set the corresponding elements (or eigenvalues) of \( P_0 \) to infinity, or more appropriately, set the corresponding elements (or eigenvalues) of \( P_0^{-1} \) to zero.

**Theorem 3.** Given the following as the incomplete prior knowledge of \( x: \pi = E[x] \), a positive semi-definite symmetric but singular matrix \( P_0^{-1} \), and \( \text{cov}(x, v^n) = A \), the unique, unified, optimal, BLUE fuser (10) for data model (5) is given by
\[
\hat{x} = VK[V'\pi], \quad y^n = [V'\pi']
\]
\[
P = VK\hat{C}K'V'
\]

where the optimal gain matrix \( K \) is as given in Theorem 2 with \( H \) and \( C \) replaced by \( \hat{H} \) and \( \hat{C} \):
\[
H_{V} = \begin{bmatrix} [I_{r \times r}, 0] \end{bmatrix}, \quad \hat{C} = \begin{bmatrix} \Lambda_1 & -V'A \\ -V'A' & C \end{bmatrix}
\]

\[V = [V_1, V_2] \] is the orthogonal matrix that diagonalizes \( P_0^{-1}: P_0^{-1} = V\text{diag}(\Lambda_1^{-1}, 0, \ldots, 0)V' \) such that \( \Lambda_1 = \text{diag}(\lambda_1, \ldots, \lambda_r) > 0 \), \( r = \text{rank}(P_0^{-1}) \).

### 4.4 Optimal Weighted LS Fusion

**Theorem 4.** The unique, unified, optimal, WLS fuser (11) having minimum norm using data from (5) is
\[
\hat{x} = Ky^n
\]
\[
P = KCK' = [H'C^{-1}H]^+
\]
where the gain matrix is given by
\[
K = PH'C^{-1}
\]
This minimum-norm fuser becomes the unique optimal WLS fuser if and only if \( H'C^{-1}H \) is nonsingular.

Although new for fusion, this is a well-known result for estimation.

For the standard distributed WLS fusion, the following theorem holds for each submatrix of the gain matrix \( K \).

**Theorem 5.** Each element \( K_{ij} \) of \( K = [K_1, \ldots, K_n] \) of the optimal WLS rule for the standard distributed fusion is given explicitly in terms of \( C^{-1} \) alone by
\[
K_{ij} = \left( \sum_{i=1}^{n} C_{ij}^{-1} \right)^{-1} \sum_{i=1}^{n} C_{ij}^{-1}
\]
where \( C_{ij}^{-1} \) is the \((i, j)\)th \( n \times n \) submatrix of \( C^{-1} \).

This theorem is attractive mainly in a theoretical sense. It makes the dependence of the weights of the optimal fuser on \( C^{-1} \) explicit. Note that these weights are uniquely determined by \( C^{-1} \) alone for the standard distributed fusion. It can be shown that many existing fusion formulas, such as the standard distributed fusion with feedback of [7] and the two-sensor track fusion formula of [3], are special cases of this theorem. The fuser of this theorem is, however, computationally expensive.

### 4.5 Optimal Generalized Weighted LS Fusion

Consider now a “generalized WLS” problem of using the (non-Bayesian) least-squares method to provide an optimal fused estimate of a random variable \( x \) that has prior mean \( \pi \) and covariance \( P_0 \), and is correlated with the data errors, with the cross-covariance given by \( \text{cov}(x, v) = A \).

Consider the following model
\[
y_0 = x + v_0 = \pi
\]
with \( \text{cov}(v_0) = \text{cov}(x - x) = P_0 \). Let
\[
\hat{y}^n = \begin{bmatrix} \pi \\ y^n \end{bmatrix}, \quad \hat{H} = \begin{bmatrix} I \\ H \end{bmatrix},
\]
\[
\hat{v}^n = \begin{bmatrix} v_0 \\ v^n \end{bmatrix}, \quad \hat{C} = \begin{bmatrix} P_0^{-1} & -A \\ -A' & C \end{bmatrix}
\]

Then,
\[
\hat{y}^n = \hat{H}\pi + \hat{v}^n
\]
and the corresponding optimal WLS estimator of $x$ using data set $\tilde{y} = \{\tilde{x}, y_1, \ldots, y_n\}$ is

$$\hat{x} = \arg \min_{\tilde{x}} (\tilde{y}^n - \tilde{H}\tilde{x})^T \tilde{C}^{-1} (\tilde{y}^n - \tilde{H}\tilde{x}) = K\tilde{y}^n \tag{36}$$

$$P = E[(x - \hat{x})(x - \hat{x})^T][\tilde{y}] = K\tilde{C}K' = [\tilde{H}^T\tilde{C}^{-1}\tilde{H}]^{-1} \tag{37}$$

$$K = \tilde{P}\tilde{H}^T \tilde{C}^{-1} \tag{38}$$

which follows from the optimal WLS fuser above by replacing $y^n$, $H$, $C$ with $\tilde{y}^n$, $\tilde{H}$, $\tilde{C}$.

5 Relationships Among Fusion Rules

5.1 Equivalence of Optimal WLS and BLUE Fusion

**Theorem 6.** Assume that there is no prior knowledge about the estimatee $x$ and $C$ is positive definite. Then the unique BLUE fuser and the unique minimum-norm optimal WLS fuser are identical (i.e., they coincide).

**Remark 1:** In the case where $H^TC^{-1}H$ is singular, the optimal WLS problem amounts to the undetermined case of a system of linear equations, and thus has infinitely many solutions; any of these solutions $\hat{x}$ yields zero fitting error: $(y^n - \hat{H}\hat{x})^T C^{-1} (y^n - \hat{H}\hat{x}) = 0$. Consequently, an optimal WLS fuser does not exist in a strict sense since there is no way of selecting one out of these solutions without using an additional criterion. The solution given by Theorem 4 is the one having smallest norm and is unique. This theorem states that this minimum-norm solution turns out to have the smallest error covariance.

**Remark 2:** The above equivalence is roughly the best one can expect since $C$ is always assumed positive definite whenever the optimal WLS fusion is considered. If $C$ is not positive definite, the WLS fuser of Theorem 4 with the following replacement $P = KCK'$ and $K = [H^TC^{-1}H]^+ H^TC$ does not coincide with the BLUE fuser, and in fact, is often a meaningless fuser.

**Remark 3:** The optimal WLS fusers using data of uncorrelated errors are statistically consistent in the sense the estimation error approaches zero as more and more data are used. When the data have correlated errors, these fusers are in general no longer consistent. By this theorem, the BLUE fuser for data of correlated errors could also be inconsistent.

Conversely, an optimal WLS fusion problem (i.e., without prior knowledge about $x$) can be treated as a BLUE fusion problem as follows. Use the first piece (or block) of data $y_1$ to obtain the optimal WLS (or equivalently, the BLUE) fused estimate of $x$, which is given by $P^1 = [H_1^TC_1^{-1}H_1]^+$, $\hat{x}^1 = P^1H_1^TC_1^{-1}y_1$. Then use the remaining data to compute the BLUE fused estimates by treating $\hat{x}^1$ and $P^1$ as $\bar{x}$ and $\bar{P}$ in the usual formulation of BLUE fusion. Note that if the data errors are correlated, the errors of the remaining data are in general correlated with $\tilde{x} = x - \bar{x} = x - \hat{x}^1$.

Theorem 6 above also guarantees that the optimal generalized WLS fuser coincides with the BLUE fuser if $C > 0$.

5.2 Relationships of Various Fusion Rules

Of all fusion rules presented above, the BLUE fusion rule with incomplete prior knowledge is the most general one. It clearly reduces to the BLUE fusion rule without prior knowledge and with complete prior knowledge in the case of $P_0^{-1} = 0$ and $P_0^{-1} \neq 0$, respectively. Somewhat against many people’s intuition, the BLUE fusion rule with complete prior knowledge is a special case of the BLUE fusion rule without prior knowledge since the prior mean and the covariance can be treated as the initial data and the associated error covariance. From Theorem 6, it is clear that the optimal WLS fusion rule is a special case of the BLUE fusion rule without prior knowledge. It is also easy to see that the optimal generalized WLS fusion rule is a special case of the BLUE fusion rule with complete prior knowledge. Note that almost all optimal linear fusion formulas presented before in the literature are special cases of the optimal WLS fusion rules of this paper.

6 Summary

Centralized, distributed, and hybrid estimation fusion architectures have been considered in a general framework. A unified linear data model for these three architectures have been presented. The best linear unbiased estimation (BLUE), the optimal weighted least squares (WLS), and the optimal generalized WLS fusion rules for an arbitrary number of sensors, with complete, incomplete, or no prior information about the estimatee, have been presented in explicit forms. These rules have unified forms for the above three fusion architectures and are extremely convenient for implementation. They are much more general and flexible than previously available results. For example, they remain to be optimal for the cases where observation errors are arbitrarily correlated across sensors, over time, and/or arbitrarily correlated with the estimatee. They are also optimal for cases with complete, incomplete, or no prior information. The fusion rules for the standard distributed architecture are extremely general in the sense that they virtually do not rely on any assumption on the local estimates. For example, they are optimal given the cross covariance of local estimates no matter what these estimates are and how they were obtained. Relationships among the various fusion rules presented have also been provided.

A Appendix: Proofs

A.1 Proof of Theorem 1

The unique BLUE estimator of $x$ using $Y = \{y_1, \ldots, y_n\}$ is given by (see, e.g., [6])

$$\hat{x}^n = E[x] + \text{cov}(\tilde{x}, \tilde{y}^n)\text{cov}^T(\tilde{y}^n)(y^n - E[y^n])$$

$$= \bar{x} + K[y^n - H\bar{x}]$$

$$P^n = P_0 - \text{cov}(\tilde{x}, \tilde{y}^n)\text{cov}^T(\tilde{y}^n, \tilde{x})$$

$$= P_0 - KSK' \tag{39}$$

$$P^n = P_0 - \text{cov}(\tilde{x}, \tilde{y}^n)\text{cov}^T(\tilde{y}^n, \tilde{x})$$

$$= P_0 - KSK' \tag{40}$$
where \( \hat{\theta} = x - \bar{x}, \bar{y} = y - H \bar{x} = H \hat{\theta} + v^n \)

\[
K = \text{cov}(\hat{\theta}, \bar{y}) = \text{cov}(\hat{\theta}, y^n) && (41)
\]

\[
S = \text{cov}(\bar{y}) = \text{cov}(H \hat{\theta} + v^n) = HP_0H' + C + HA + (HA)' && (42)
\]

(15) follows from

\[
\text{cov}(\hat{\theta}, \bar{y}) = E[\hat{\theta}(H \hat{\theta} + v^n)] = P_0H' + A
\]

The estimation error is

\[
\hat{\theta} = x - \hat{\theta} = x - [\bar{x} + K(y^n - H \bar{x})] = U^n \bar{x} - K v^n
\]

(16) thus follows from the property of the BLUE estimator that \( P^n = \text{cov}(\hat{\theta})Y = \text{cov}(\hat{\theta}) \). Its equivalence to \( P^n = P_0 - KSK' \) can also be shown by algebraic manipulations. (17) follows from below:

\[
(P^nH' + A)(C + HA)^{-1} = \left\{ [P_0 - (P_0H' + A)(HP_0H' + C + HA + (HA)')^{-1} \cdot (P_0H' + A)]^H H' + A \right\}(C + HA)^{-1} = (P_0H' + A)(H_0H' + C + HA + (HA)')^{-1} \cdot (HP_0H' + (HA)')(C + HA)^{-1} = (P_0H' + A)(H_0H' + C + HA + (HA)')^{-1} \cdot (C + HA)(C + HA)^{-1} = (P_0H' + A)S^{-1}
\]

A.2 Proof of Theorem 2

Let

\[
\hat{\theta} = a + Ky^n = a + K(Hx + v^n)
\]

where \( a \) and \( K \) do not depend on \( Y = \{y_1, \ldots, y_n\} \). When there is no prior knowledge about \( x \), the unbiasedness of \( \hat{\theta} \) requires that \( \hat{\theta} \) be unbiased for all possible (and unknown) \( \bar{x} \), that is, \( E[\hat{\theta}] = \bar{x}, \forall \bar{x} \). This is true iff \( a = 0 \) and \( KH = I \). This also agrees with the fact that in order for the BLUE estimator

\[
\hat{\theta} = \bar{x} + K(y^n - H\bar{x}) = (I - KH)\bar{x} + Ky^n
\]

to be unbiased for all possible \( \bar{x} \), it is required that \( \hat{\theta} = Ky^n \) with \( K \) satisfying \( KH = I \). Consequently, the BLUE estimator is given by \( \hat{\theta} = Ky^n \), where \( K \) satisfies \( KH = I \), or equivalently \( H'K' = I \), and minimizes the corresponding error covariance \( E[(x - \hat{\theta})(x - \hat{\theta})'] = KCK' \) since

\[
x - \hat{\theta} = Ix - Ky^n = KHx - K(Hx + v^n) = -Kv^n
\]

In other words, \( \hat{\theta} = Ky^n \), where

\[
K' = \arg \min_{K'} \langle (K')'CK' \rangle
\]

Note that this is a weighted least squares problem with linear constraint \( H'K' = I \).

It is well known that the general solution of \( H'K' = I \) is

\[
K' = (H')^{-1} - T \xi
\]

where \( \xi \) is any matrix of suitable dimension and \( T = I - (H')'H'^+ = I - HH'^+ \) (since \( T' = T \)). It is straightforward to show from the basic properties of the matrix pseudoinverse that

\[
C^{-1}_T(TC^{-1}_T) + TC^{-1}_T
\]

Substituting the last two equations above into the quadratic form in (43) yields

\[
\{[H' - \xi(T)C^{-1}_T] [H' - \xi(T)C^{-1}_T]'\} = \{[\xi' - H'^{-1}C^{-1}_T (TC^{-1}_TC^{-1}_T) + H''^{-1}C^{-1}_T (TC^{-1}_TC^{-1}_T)]\} ^+ \cdot \xi' - H'^{-1}C^{-1}_T (TC^{-1}_TC^{-1}_T) + H''^{-1}C^{-1}_T (TC^{-1}_TC^{-1}_T) \cdot (H'^{-1}C^{-1}_T (TC^{-1}_TC^{-1}_T) + H''^{-1}C^{-1}_T (TC^{-1}_TC^{-1}_T)) \}
\]

Clearly, minimizing the quadratic objective function in (43) thus amounts to solving

\[
[H' - \xi(T)C^{-1}_T] [H' - \xi(T)C^{-1}_T]' = 0
\]

for \( \xi \). It can be shown that

\[
T(TCT) = (TCT) = (TCT)^+
\]

Hence, the solution is

\[
\xi = (TCT)^+C(H)' \cdot \theta, \quad \forall \theta : C^{-1}_T T \theta = 0
\]

It thus follows from (44) that the general solution of (43) is

\[
K' = [I - (TCT)^+C(H)' + T \theta] = K' + T \theta
\]

where \( \theta \) is any matrix of suitable dimension satisfying \( C^{-1}_T T \theta = 0 \). (23) thus follows.

To show necessary and sufficient condition for the uniqueness of \( K' \), note first that only \( K' \) of (45) with \( T \theta = 0 \) can be a unique solution, otherwise \( K' \) of (45) with \( aT \theta \) for any real number \( a \neq 0 \) would be a distinct solution since \( aC^{-1}_T T \theta = 0 \). When \( T \theta = 0 \), (45) gives a unique solution due to the uniqueness of the M-P pseudoinverse. Thus \( K' \) is unique iff \( T \theta = 0 \). A necessary and sufficient condition for \( T \theta = 0 \) when \( C^{-1}_T T \theta = 0 \) is that the vector \( T \theta \) is in the row space of \( C^{-1}_T \). Since \( T = I - (H')'H' \) is a projector onto the orthogonal complement of the row space (i.e., subspace spanned by the row vectors) of \( H' \), the above necessary and sufficient condition holds iff the row space of \( C^{-1}_T \) is the orthogonal complement of the row space of \( H' \), which is equivalent to \( [H, C^{-1}_T] \) having full row rank. The second equation in (21) follows from the first one because it can be shown that

\[
(TCT)^+ = T^{-1} (T^{-1}C^{-1}_T)^{-1} T^{-1}
\]

Note that

\[
y^n' T \eta = (x'H' + v'v)^T \theta
\]

\[
x^n' [H' - H' (H')' + H'] \theta + v'v \theta
\]

\[
= y^n' C^{-1}_T T \theta = 0
\]

where \( v^n = C^{-1}_T v^n \). Thus, \( \hat{\theta} = Ky^n = (K' + T \theta)'y^n = Ky^n \) is unique. Its error covariance is \( P = KCK' \).
A.3 Proof of Theorem 3

Since any symmetric matrix can be diagonalized by orthogonal transformation, one has $P_0^{-1} = V\text{diag}(\Lambda_1^{-1}, \Lambda_2)V'$, where $V$ is an orthogonal matrix ($V^{-1} = V'$), and $\Lambda_1 = \text{diag}(\lambda_1, \ldots, \lambda_r)>0$, $r = \text{rank}(P_0^{-1})$ and $\Lambda_2 = 0$. Let

$$
\begin{bmatrix}
  u_1 \\
  u_2 \\
  \bar{u}_1 \\
  \bar{u}_2
\end{bmatrix} = u = V'x = \begin{bmatrix} V'_1 \\
  V'_2 \end{bmatrix}x
$$

where $u_1$ and $\bar{u}_1$ are the subvectors of $u$ and $\bar{u}$ that correspond to $\Lambda_1$ in the sense that $\text{cov}(u_1 - \bar{u}_1) = \Lambda_1$. Note that $\Lambda_2 = 0$ is practically equivalent\(^3\) to no prior knowledge about $u_2$, and thus the information about $x$ contained in $\bar{u}$, $P_0^{-1}$, and $\text{cov}(x, v^n) = A$ are equivalent to that of $u_1$ contained in $\bar{u}$, $\Lambda_1$, and $\text{cov}(u_1, v^n) = V'_1A$.

Treating $\bar{u}_1$ as an observation $y_0$ of $u$ leads to the following data model

$$y_0 = \bar{u}_1 = u_1 + (\bar{u}_1 - u_1) = [I, 0]u + v_0$$

The original data model (5) of $x$ can be converted to a model of observing $u$:

$$y^n = Hx + v^n = (HV)u + v^n$$

Combining these two model yields

$$\tilde{y}^n = \begin{bmatrix} \bar{u}_1 \\
  y^n \end{bmatrix} = \begin{bmatrix} [I, 0] \\
  H'V \end{bmatrix}u + \begin{bmatrix} \bar{u}_1 - u_1 \\
  v^n \end{bmatrix} = \tilde{H}u + \tilde{v}^n$$

(46)

The covariance $\tilde{C} = \text{cov}(\tilde{v}^n)$ is given by

$$\tilde{C} = \text{cov} \left( \begin{bmatrix} \bar{u}_1 - u_1 \\
  v^n \end{bmatrix} \right) = \begin{bmatrix} \Lambda_1 & -V'_1A \\
  -(V'_1A)' & C \end{bmatrix}$$

Once $\bar{u}_1$ is treated as an observation, there is no prior information about $u$ at all. This means that Theorem 2 is applicable now. Therefore, all formulas in this theorem follows from Theorem 2 for the data model (46) and the relationship

$$x = Vu, \quad \dot{x} = V\dot{u}, \quad P = V\text{cov}(u - \bar{u})[\tilde{y}^n]V'$$

A.4 Proof of Theorem 5

In the case of standard distributed fusion, $H'C^{-1}H$ is non-singular since $C$ is always assumed to have full rank for the optimal WLS fusion and $H = [I, \ldots, I]$ for the standard distributed fusion. Hence, the theorem follows from direct manipulations.

\[^{3}\text{\Lambda_2 = 0 is not rigorously equivalent to no prior knowledge about u_2 in theory. In reality, however, P_0^{-1} is chosen in such a way it represents the fact that there is no prior knowledge about u_2 in the first place. This equivalence is thus rigorous provided P_0^{-1} is chosen this way non- rigorously. A rigorous approach here will never be exactly equivalent to this real fact and thus leads to inferior results than this theorem for this practical problem. Note also that cov(u_2, v^n) = V'_2A contains no useful information for BLUE estimation at all since there is no prior knowledge about u_2 anyway.}]

References


