Proceedings of the 5th International Conf. Information Fusion
Annapolis, MD, USA, July 2002, pp. 497–504

Optimal Linear Estimation Fusion—Part V: Relationships*

X. Rong Li **Keshu Zhang**
Department of Electrical Engineering, University of New Orleans, New Orleans, LA 70148
Phone: 504-280-7416, Email: xli@uno.edu
Juan Zhao **Yunmin Zhu**†
Department of Mathematics, Sichuan University, Chengdu, Sichuan 610064, P.R. China

Abstract – *In this paper, we continue our study of optimal linear estimation fusion in a unified, general, and systematic setting. We clarify relationships among various BLUE and WLS fusion rules with complete, incomplete, and no prior information presented in Part I before; and we quantify the effect of prior information and data on fusion performance, including conditions under which prior information or data are redundant.*

Keywords: Distributed fusion, centralized fusion, track fusion, prior information, BLUE, least squares

1 Introduction

In the previous parts of this series of papers [9, 6, 8, 7], we have tackled the problem of optimal linear estimation fusion in a general, systematic, and unified setting. More specifically, Part I [9] presents several unified optimal fusion rules in the framework of best linear unbiased estimation (BLUE) and weighted least squares (WLS) and their underlying unified data model for estimation fusion; Part II [6] provides a general discussion and examples of the fusion rules presented in Part I; Part III [8] handles the cross correlation of local estimation errors and presents formulas for its computation that account for all sources of cross correlation, including those that have been overlooked before in the literature; Part IV [7] presents theoretical results on performance efficiency of distributed fusion relative to centralized fusion and conditions under which distributed and centralized fusers are identical.

In this paper (Part V), we continue our investigation on optimal estimation fusion. More specifically, we clarify relationships among various fusion rules in the case of complete, incomplete, and no prior information; we quantify the effects or contributions of prior information and data, respectively, on estimation fusion; we present conditions under which prior information or data are redundant in that they do not improve optimal estimation fusion performance.

The results presented here will enable better understanding of many fusion results available in the literature, such as those in [4, 1, 2].

*Research supported in part by ONR grant N00014-00-1-0677, NSF grant ECS-9734285, and NASA/LEQSF grant (2001-4)-01.
†Supported in part by the National Key Project and NNSF of China.

2 BLUE and WLS Fusers

2.1 Unified Linear Data Model

Many results presented in this paper and other parts of this series assume the following unified data model

$$y = Hx + v \tag{1}$$

where for the centralized fusion, y, H, and v are the actual observations, its matrix, and observation noise, respectively; for (standard) distributed fusion, they are actually stacked local estimates $[\hat{x}_1', \ldots, \hat{x}_n']'$, $[I, \ldots, I]'$, and stacked local estimation errors $-[\tilde{x}_1', \ldots, \tilde{x}_n']'$, respectively. See Part I [9] for more details. The first two moments are denoted as

$$\bar{x} = E[x], \quad C_x = \text{cov}(x), \quad C_{xv} = \text{cov}(x, v)$$
$$\bar{v} = E[v], \quad C = \text{cov}(v), \quad \bar{y} = E[y], \quad C_y = \text{cov}(y)$$

2.2 Basic Assumptions and Conventions

The following assumptions and conventions are used throughout the paper. (a) The weight matrix C for optimal WLS is symmetric and positive definite; otherwise the least squares approach is questionable — zero fitting error does not imply zero estimation error. (b) Both $\bar{v}$ and C for optimal WLS are known for the data model (1). For the standard distributed fusion, this amounts to assuming knowing all biases (if any) and cross covariances of the estimation errors of local estimates, although we do not explicitly assume that local estimates are unbiased. (c) For the data model (1), by **prior information**, we mean that related to the first two moments of the estimatee (i.e., quantity to be estimated) x directly — $\bar{x}$, C_x, and C_{xv} — required in the Bayesian approach; **with complete prior information** means that $\bar{x}$, C_x, and C_{xv} are known exactly; **with incomplete prior information** means that not all $\bar{x}$, C_x, and C_{xv} are known or exist; **without prior information** means none of $\bar{x}$, C_x, and C_{xv} are known or exist, as in the case of the classical approach. (c) The H matrix in the unified data model (1) is nonrandom and known.

2.3 Unified Fusion Rules

We summarize the optimal fusers presented in Part I [9] for (1) below.

Theorem 2.1 (BLUE fusion given complete prior). Using data model (1), the (almost surely, i.e., with probability 1) unique BLUE fuser with prior information $\bar{x}$, C_x, and C_{xv} of x is

$$\begin{aligned} \hat{x} &= \bar{x} + K[y - H\bar{x} - \bar{v}] && (2)\\ P &= \text{cov}(x-\hat{x}) = C_x - KSK' && (3)\\ C_y &= HC_xH' + C + HC_{xv} + (HC_{xv})' && (4)\\ K &= (C_xH' + C_{xv})C_y^+ && (5)\end{aligned}$$

In the above, S^+ stands for the unique *Moore-Penrose pseudoinverse* (*MP inverse* in short) of S.

Theorem 2.2 (BLUE fusion without prior). For data model (1) with known $\bar{v}$ and C, a BLUE fuser without prior information of x exits if and only if H has full column rank (i.e., $H^+ = (H'H)^{-1}H'$). If exists, it is unique (almost surely) and given by

$$\hat{x} = K(y-\bar{v}), \quad P = KCK' \qquad (6)$$

where $K = H^+[I - C(TCT)^+]$ and $T = I - HH^+$.

Theorem 2.3 (BLUE fusion given incomplete prior). Given partial prior information: $\bar{x}$, a positive semidefinite symmetric but singular matrix C_x^{-1}, and C_{xv}, the BLUE fuser for data model (1) with known $\bar{v}$ and C exists if and only if $\bar{H}^+\bar{H} = I$. If exists, it is given by

$$\hat{x} = VK[(V_1'\bar{x})', (y-\bar{v})']', \quad P = VK\bar{C}K'V'$$

where $V = [V_1, V_2]$ is the orthogonal matrix that diagonalizes C_x^{-1}: $C_x^{-1} = V\text{diag}(\Lambda_1, 0)V'$ such that V_1 corresponds to $\Lambda_1 = \text{diag}(\lambda_1, \ldots, \lambda_r) > 0, r = \text{rank}(C_x^{-1})$, the optimal gain matrix K is as given in Theorem 2.2 with H and C replaced by $\bar{H}$ and $\bar{C}$, respectively, given by

$$\bar{H} = \begin{bmatrix} [I_{r\times r}, 0] \\ HV \end{bmatrix}, \qquad \bar{C} = \begin{bmatrix} \Lambda_1^{-1} & -V_1'C_{xv} \\ -(V_1'C_{xv})' & C \end{bmatrix}$$

The same uniqueness of $\hat{x}$ and K as in Theorem 2.2 holds with H and C replaced by $\bar{H}$ and $\bar{C}$, respectively.

Theorem 2.4 (WLS fusion). The unique optimal WLS fuser having minimum norm using data model (1) with known $\bar{v}$ and C is

$$\hat{x}_* = K_*(y-\bar{v}), \quad P_* = K_*CK_*' = (H'C^{-1}H)^+$$

where the gain matrix is given by

$$K_* = (H'C^{-1}H)^+H'C^{-1} = P_*H'C^{-1}$$

This minimum-norm fuser becomes the unique **optimal** WLS fuser if and only if $H'C^{-1}H$ is nonsingular, in which case $P_* = (H'C^{-1}H)^{-1}$.

Theorem 2.5 (Optimal generalized WLS fusion). Given data model (1) with known $\bar{v}$ and C, and prior information $\bar{x}$, C_x, and C_{xv} such that

$$\tilde{C} = \begin{bmatrix} C_x & -C_{xv} \\ -C_{xv}' & C \end{bmatrix}$$

is nonsingular, the unique optimal generalized WLS fuser having minimum norm is

$$\hat{x} = K[\bar{x}', (y-\bar{v})']', \quad P = K\tilde{C}K' = (\tilde{H}'\tilde{C}^{-1}\tilde{H})^+$$

where $\tilde{H} = [I, H']'$ and the gain matrix is given by $K = P\tilde{H}'\tilde{C}^{-1}$; this minimum-norm fuser becomes the unique optimal generalized WLS fuser if and only if $\tilde{H}'\tilde{C}^{-1}\tilde{H}$ is nonsingular, in which case $P = (\tilde{H}'\tilde{C}^{-1}\tilde{H})^{-1}$.

Note that nonsingularity (and positive definiteness) of $\tilde{C}$ is equivalent to $\det(C_x) \neq 0$ and $\det(C - C_{xv}'C_x^{-1}C_{xv}) \neq 0$ (or $\det(C) \neq 0$ and $\det(C_x - C_{xv}C^{-1}C_{xv}') \neq 0$).

3 BLUE vs. Least Squares

The following proposition presents their equivalent and familiar forms of the gain and MSE matrices of the BLUE without prior (Theorem 2.2) when C is nonsingular.

Proposition 3.1 (Alternative form of gain). The gain matrix and the MSE matrix of the BLUE fuser without prior given by Theorem 2.2 have the following alternative forms

$$K = PH'C^{-1}, \quad P = KCK' = (H'C^{-1}H)^+$$

if and only if C is nonsingular.

This follows directly from the following lemma.

Lemma 3.1 For any H and any symmetric and positive semidefinite C, the following holds

$$H^+[I - C(TCT)^+] = (H'C^{-1}H)^+H'C^{-1} \qquad (7)$$

if and only if C is nonsingular, where $T = I - HH^+$.

This proposition establishes an intimate relationship between BLUE without prior information and optimal WLS.

3.1 BLUE without Prior vs. Optimal WLS

Both BLUE fuser without prior and WLS fuser use data to make estimation without prior information. We are interested in their relationship. Clearly, even the optimal WLS fuser cannot have an MSE matrix smaller than that of the BLUE fuser without prior. By definition, the optimal WLS fuser is the best fuser in the class of all WLS fusers that has the smallest MSE matrix. As a corollary of Proposition 3.1, the following theorem states that if exists, it is actually the best among all linear unbiased fusers.

Theorem 3.1 (Equivalence of optimal WLS and BLUE without prior). Consider linear data model (1) with known $\bar{v}$ and C. Assume that C is nonsingular and H has full column rank. Then the optimal WLS fuser (given by Theorem 2.4) and BLUE fuser without prior (given by Theorem 2.2) exist and are identical.

For linear data model (1) with known $\bar{v}$ and C but without prior information, a linear unbiased fuser exists if and only if H has full column rank. When H does not have full column rank, the WLS criterion with weight $W = C^{-1}$ is still meaningful. Its solution, however, is not unique any more. The one with minimum norm is given by

$$\hat{x} = Ky, \quad P = KCK', \quad K = (H'C^{-1}H)^{+}H'C^{-1}$$

By Proposition 3.1, the formulas given in Theorem 2.2 when C^{-1} exists reduce to this minimum-norm solution. As a result, these formulas can be thought of giving a generalization of the minimum-norm optimal WLS fuser, which is biased when H does not have full column rank.

3.2 BLUE with Prior vs. Optimal Generalized WLS

Similarly, we are interested in the relationship of BLUE and WLS fusers that use prior information. As with the case without prior, while any WLS fuser using prior cannot have an MSE matrix smaller than that of the BLUE fuser, the optimal WLS fuser (i.e., the best fuser in the class of all WLS fusers) is actually the same as the BLUE fuser and thus is optimal among all linear unbiased fusers, as the following theorem states.

Theorem 3.2 (Equivalence of optimal WLS and BLUE given prior). Consider linear data model (1) with known $\bar{v}$ and C and complete prior information ($\bar{x}$, C_x, and C_{xv}). Assume that $\tilde{C}$ of Theorem 2.5 is nonsingular. Then the optimal generalized WLS fuser (given by Theorem 2.5) and BLUE fuser with complete prior (given by Theorem 2.1) exist and are identical.

Since optimal generalized WLS is actually optimal WLS with prior mean treated as extra data (see Proposition 4.2), this theorem follows directly from Theorem 4.1, which states that BLUE with complete prior for linear data model can always be converted to BLUE without prior by treating prior mean as extra data.

3.3 Remarks

BLUE without prior exists when optimal WLS exists, but optimal WLS may not exist when BLUE without prior exists. Likewise, BLUE with complete prior exists when optimal generalized WLS exists, but optimal generalized WLS may not exist when BLUE with complete prior exists. These two theorems state that they are identical when they both exist.

4 Effect of Prior Information

4.1 BLUE: With Prior vs. Without Prior

What is the relationship between BLUE fusion with and without prior? Can one be converted to the other? The following theorem states that for a linear data model BLUE fusion without prior is more general than BLUE fusion with complete prior in that any latter problem can be converted to a former problem.

Theorem 4.1 (Conversion of BLUE with prior to BLUE without prior). For the linear data model (1) with known $\bar{v}$ and C, BLUE with complete prior information ($\bar{x}$, C_x, and C_{xv}) can always be converted to BLUE without prior information by treating the prior mean $\bar{x}$ as extra data in the linear model: $\bar{x} = x + (\bar{x} - x)$. More specifically, BLUE with complete prior information ($\bar{x}$, C_x, and C_{xv}) for the linear data model (1) with known $\bar{v}$ and C always coincides (almost surely) with BLUE without prior information for the linear data model $\tilde{y} = \tilde{H}x + \tilde{v}$ with

$$\begin{aligned} \tilde{y} &= \begin{bmatrix} \bar{x} \\ y \end{bmatrix}, \quad \tilde{H} = \begin{bmatrix} I \\ H \end{bmatrix} \\ E[\tilde{v}] &= \begin{bmatrix} 0 \\ \bar{v} \end{bmatrix}, \quad \tilde{C} = \begin{bmatrix} C_x & -C_{xv} \\ -C'_{xv} & C \end{bmatrix} \end{aligned}$$

which is given by

$$\hat{x} = \tilde{K}(\tilde{y} - E[\tilde{v}]), \quad P = \tilde{K}\tilde{C}\tilde{K}', \quad \tilde{K} = [I - KH, K] \tag{8}$$

where K is the gain matrix of the BLUE fuser with complete prior, given by Theorem 2.1.

Note that in this case, the gain matrix of BLUE without prior does not involve MP inverse.

Albeit close, Theorem 4.1 does not imply that Theorem 2.1 can be derived from Theorem 2.2, which should be true but has not been shown explicitly here because we have not derived the specific formulas given in Theorem 2.1 from those of Theorem 2.2. What Theorem 4.1 shows is that prior information can always be completely embedded into a linear data model with prior mean as data. In fact, Theorem 2.2 is valid no matter if the estimatee x is random or not. It can be thus viewed as a unification of classical and Bayesian linear estimation fusion. For random x, it is more precise to state Theorem 2.2 as giving the BLUE fuser for the case with unknown $\bar{x}$ (i.e., regardless if C_x or C_{xv} is known or not), rather than without prior. In other words, the BLUE fusion formulas in this case are invariant with respect to C_x and C_{xv} (even if they are known) provided $\bar{x}$ is unknown—the possible effect of C_x and C_{xv} comes only through other quantities, e.g., C. For example, Theorem 2.2 is valid whether C_{xv} of the data model (1) is known or not.

Theorem 4.1 is the foundation of an approach to optimal linear update with out-of-sequence measurements [11].

The following theorem states that BLUE without prior can also be converted to BLUE with complete prior.

Theorem 4.2 (Conversion of BLUE without prior to BLUE with prior). The problem of BLUE without prior using data $y = Hx + v$ can always be converted to that of BLUE with complete prior $\bar{x} = H_1^{-1}y_1$, $C_x = H_1^{-1}\text{cov}(v_1)(H_1^{-1})'$, and $C_{xv} = -H_1^{-1}\text{cov}(v_1, v_2)$ using data $y_2 = H_2x + v_2$ (i.e., the two BLUE estimators are identical almost surely), where H_1 is nonsingular, formed by proper rows of H such that

$$Ay = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \quad AH = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix}, \quad Av = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

with Au denoting some row-interchanging version of vector u.

The key in the application of this theorem is to find H_1 for the problem at hand.

4.2 Contribution of Prior to BLUE

For the same linear data model, BLUE fuser without prior clearly can never have a smaller MSE matrix than BLUE fuser with complete prior information since the latter uses more information optimally. The questions are: (a) how much worse? (b) can they be the same? Clearly, they are the same if and only if the prior information turns out to be redundant given the data for the problem. The following theoretical results answer these questions.

Lemma 4.1 (Redundancy condition of prior for BLUE). Let $\hat{x}_1$ and $\hat{x}_2 = K_2(y - \bar{v})$ be BLUE with complete prior and without prior information, respectively, using the same data y with known mean error $\bar{v}$. Then a necessary and sufficient condition for $\hat{x}_1 = \hat{x}_2$ almost surely is $K_2C_y = C_{xy}$, where $C_y = \text{cov}(y)$ and $C_{xy} = \text{cov}(x, y)$.

Note that this theorem is valid regardless if y is linear or nonlinear in the estimatee x provided BLUE without prior is given in the form $\hat{x}_2 = K_2(y - \bar{v})$. By Theorem 2.2, $\hat{x}_2$ is always of this form for linear data model (1). For linear data, we have the following stronger results.

Theorem 4.3 (Redundancy conditions of prior for BLUE). Consider linear data model (1) with known $\bar{v}$ and C. Let $\hat{x}_1$ and $\hat{x}_2$ be BLUE with complete prior and without prior information, respectively, using the same data y. Then the following statements are equivalent.

(a) $\hat{x}_1 = \hat{x}_2$ almost surely (i.e., prior information is redundant for BLUE)

(b) The gain matrix K_2 of $\hat{x}_2$ satisfies $K_2C_y = C_{xy}$.

(c) The gain matrix K_2 of $\hat{x}_2$ satisfies $K_2C_{vy} = 0$.

(d) $(I - C_{vy}C_{vy}^+)H$ has full column rank; that is,

$$[(I - C_{vy}C_{vy}^+)H]^+[(I - C_{vy}C_{vy}^+)H] = I \qquad (9)$$

where C_{vy}^+ is the MP inverse of $C_{vy} = \text{cov}(v, y)$.

(e) The following condition holds

$$\begin{aligned}&[I - C_{xy}C_y^+H][(I - C_yC_y^+)H]^+[(I - C_yC_y^+)H] \\ = \;& I - C_{xy}C_y^+H \end{aligned} \qquad (10)$$

Note that by Theorem 2.2, BLUE without prior exists if and only if H has full column rank.

The redundancy conditions given in Theorem 4.3 are general—it holds whenever BLUE without prior exists for the model (1) with known $\bar{v}$ and C. With additional assumptions, we have stronger results, as stated in the following corollaries.

Corollary 4.2 (Sufficient condition for redundancy of prior for BLUE). Consider linear data model (1) with known $\bar{v}$ and C. Then prior information is redundant for BLUE if $\text{rank}[C_y, H] = \text{rank}(C_y)$ and $I - C_{xy}C_y^+H = 0$.

This corollary follows from Theorem 4.3, in particular, (10), immediately since $\text{rank}[C_y, H] = \text{rank}(C_y)$ is equivalent to $C_yC_y^+H = H$.

Corollary 4.3 (Contribution of prior to BLUE). Consider linear data model (1) with known $\bar{v}$ and C. Assume that $\tilde{C}$ of Theorem 2.5 is nonsingular. Then the contribution of the prior information ($\bar{x}$, C_x, and C_{xv}) to BLUE fusion in the sense of Fisher is given by

$$\begin{aligned} P_1^{-1} - P_2^{-1} = &(I + C_{xv}C^{-1}H)'(C_x - C_{xv}C^{-1}C_{xv}')^{-1} \\ &\cdot(I + C_{xv}C^{-1}H) \end{aligned}$$

where P_1 and P_2 are the MSE matrices of BLUE fusers with complete and without prior information, respectively. In particular, the prior information is redundant for BLUE fusion—BLUE fuser with complete prior information is (almost surely) identical to BLUE fuser without prior information—if and only if

$$I + C_{xv}C^{-1}H = 0 \qquad (11)$$

Corollary 4.4. Consider linear data model (1) with known $\bar{v}$ and C. Assume $C > 0$, $C_x > 0$, and $C_{xv} = 0$. Then the prior information ($\bar{x}$, C_x) is useful for BLUE fusion—BLUE fuser with complete prior information has a smaller MSE matrix than BLUE fuser without prior information—and the contribution of the prior information is $P_1^{-1} - P_2^{-1} = C_x^{-1}$.

This corollary indicates that in the usual case ($C > 0$, $C_x > 0$, and $C_{xv} = 0$), the optimal use of the prior mean and covariance does improve the performance of BLUE fusion; and the prior information and the data information are additive because they are uncorrelated ("independent").

It should be clear from these results that the performance "limits" discussed in [1, 2] are valid only in the sense that prior information is either unavailable or redundant, which is, however, not the usual case for steady-state filtering (e.g., α-β filtering), where $C_x := \lim_{k\to\infty} P_{k|k-1} > 0$.

4.3 BLUE: Partial Prior vs. No Prior

In practice, it is sometimes more desirable to define a singular C_x^{-1} and thus the corresponding covariance C_x does not exist[1]. This would be the case if prior information about some but not all components of $\bar{x}$ were available. For example, when tracking an aircraft just taking off from an airport, it is easy to determine the prior velocity vector (it must be within a certain velocity range) with a covariance, but not the prior position vector (it can be over a much larger region). A practical means of specifying such knowledge is to set the corresponding elements (or eigenvalues) of C_x to infinity (or more appropriately, set C_x^{-1} to zero). Such an incomplete prior problem can be converted to a problem without prior information, as the following lemma states.

Proposition 4.1. Given partial prior information: $\bar{x}$, a positive semidefinite symmetric but singular matrix C_x^{-1}, and cross covariance C_{xv}, the corresponding BLUE fusion for data model (1) with known $\bar{v}$ and C can be converted to BLUE without prior information (Theorem 2.2) with H and C replaced by $\bar{H}$ and $\bar{C}$ of Theorem 2.3.

Theorem 2.3 is actually an immediate consequence of this proposition. In fact, the proof of Proposition 4.1 is embedded in the proof of Theorem 2.3, as found in [9].

4.4 Optimal WLS: Role of Prior

The WLS method is usually limited to the case where the estimatee x is unknown and there exists no prior information. As supported by Theorem 4.1, the optimal generalized WLS fusion treats prior mean as extra data optimally. The following proposition states that the optimal generalized WLS fuser is actually a special case of the optimal WLS fuser by embedding prior information into a linear data model with prior mean as data: $\bar{x} = x + (\bar{x} - x)$.

Proposition 4.2. Given prior information $\bar{x}$, C_x, and C_{xv} and data model (1) with known $\bar{v}$ and C such that $\tilde{C}$ of Theorem 2.5 is nonsingular, the corresponding optimal generalized WLS fuser (given by Theorem 2.5) exists and is identical to optimal WLS fuser (given by Theorem 2.4) with H and C replaced by $\tilde{H}$ and $\tilde{C}$ of Theorem 4.1, respectively.

The optimality of this prior information embedding by the model $\bar{x} = x + (\bar{x} - x)$ for WLS follows from the fact that after embedding the optimal WLS is identical to BLUE with complete prior (Proposition 4.2 and Theorem 3.2).

The optimal WLS can never have a smaller MSE matrix than that of the optimal generalized WLS fusion since the latter uses prior information optimally. However, they can have the same MSE matrix when the prior information is redundant. The following theorem presents the improvement of the optimal generalized WLS over optimal WLS in terms of MSE matrix, including a necessary and sufficient condition for them to be the same.

[1] C_x^{-1} is just a symbol here, not the inverse of any matrix, although it is meant to be the inverse of C_x if it exists.

Theorem 4.5 (Contribution of prior to optimal WLS). Consider linear data model (1) with known $\bar{v}$ and C. Assume that $\tilde{C}$ of Theorem 2.5 is nonsingular. Then the contribution of the prior information ($\bar{x}$, C_x, and C_{xv}) to optimal WLS fusion in the sense of Fisher is given by

$$P_{\text{GWLS}}^{-1} - P_{\text{WLS}}^{-1} = (I + C_{xv}C^{-1}H)'(C_x - C_{xv}C^{-1}C'_{xv})^{-1} \cdot (I + C_{xv}C^{-1}H)$$

where P_{GWLS} and P_{WLS} are the MSE matrices of optimal generalized WLS and optimal WLS fusers, respectively. In particular, the prior information is redundant for optimal WLS fusion—optimal generalized WLS and optimal WLS fusers are identical (almost surely)—if and only if

$$I + C_{xv}C^{-1}H = 0 \tag{12}$$

Note that $C_x > C_{xv}C^{-1}C'_{xv}$ whenever $\tilde{C}$ of Theorem 2.5 is nonsingular and thus this theorem follows from Corollary 4.3, Proposition 4.2, and Theorem 3.2.

5 Contribution of Data

Given a particular piece of random data, we may ask similar questions: Will the optimal use of it improve BLUE fusion? If the answer is yes, how much is the improvement? The following theorem answers these questions.

Theorem 5.1 (Contribution of data to BLUE). Consider the linear data model (1) with known $\bar{v}$ and C. Assume that $\tilde{C}$ of Theorem 2.5 is nonsingular. Then the contribution of the data y to BLUE fuser having MSE matrix P in the sense of Fisher is given by

$$P^{-1} - C_x^{-1} = (H + C'_{xv}C_x^{-1})'(C - C'_{xv}C_x^{-1}C_{xv})^{-1} \cdot (H + C'_{xv}C_x^{-1})$$

In particular, data y carries no useful information beyond the prior information if and only if $HC_x + C'_{xv} = 0$.

Note that the above condition becomes

$$P^{-1} - C_x^{-1} = H'C^{-1}H = P_2^{-1}$$

if $C_{xv} = 0$, where P_2 is the MSE matrix of BLUE fuser without prior information.

It is intuitively correct that data would not help BLUE fusion if it is uncorrelated with the estimatee (i.e., $C_{xy} = 0$), which also follows rigorously from $\text{MSE}(\hat{x}) = C_x - C_{xy}C_y^{+}C'_{xy}$. The first part of the theorem thus follows from the fact that $C_{xy} = C_xH' + C_{xv}$ for linear data model (1).

6 Summary

In this paper, we have clarified the following: (a) the relationships between BLUE and WLS fusion; (b) the relationships between BLUE fusion with complete, incomplete, and no prior information; (c) the relationships between WLS fusion with and without prior information; (d) the contribution of prior information to BLUE and WLS fusion, in particular, conditions under which prior information is redundant for BLUE and WLS fusion; (e) the contribution of data to BLUE and WLS fusion. In addition, we have also shown the following: (a) treating prior mean as extra data in a linear model that accounts for the prior covariance does not lose the optimality of BLUE and WLS fusion; (b) BLUE with complete prior and without prior can be mutually converted.

In our opinion, the results presented in this series of papers, along with existing results in the literature, form essential ingredients of a basic theory of linear estimation fusion. These results also have wide application in state estimation fusion, as evidenced by the fact that they include most existing linear fusion formulas for distributed filtering as special cases. Nevertheless, fusion for estimation of a process does have its own problems and emphases, which will be handled in several forthcoming papers.

A Appendix

A.1 Proof of Lemma 3.1

The following lemma is needed in our proof.

Lemma. The following holds for any H and any symmetric and nonsingular R:

$$(R^{-1}H)^+ = H^+R[I-(TR)^+(TR)]$$

where $T = I - HH^+$.

Proof: It suffices to show that $X = H^+R[I-(TR)^+(TR)]$ is the MP inverse of $A = R^{-1}H$. Note first that $TH = 0$ and $(B^+B)' = B^+B$ for any B. Then

$$\begin{aligned}
AXA &= R^{-1}HH^+R[I-(TR)^+(TR)]R^{-1}H \\
&= R^{-1}HH^+H - R^{-1}HH^+R(TR)^+TRR^{-1}H \\
&= R^{-1}HH^+H = R^{-1}H = A \\
XAX &= H^+R[I-(TR)^+(TR)]R^{-1}HH^+R \\
&\quad \cdot[I-(TR)^+(TR)] \\
&= HH^+HR[I-(TR)^+(TR)] \\
&= H^+R[I-(TR)^+(TR)] = X \\
AX &= R^{-1}HH^+R[I-(TR)^+(TR)] \\
&= R^{-1}(I-T)R[I-(TR)^+(TR)] \\
&= (I-R^{-1}TR)[I-(TR)^+(TR)] \\
&= I - R^{-1}TR - (TR)^+(TR) \\
&\quad + R^{-1}TR(TR)^+(TR) \\
&= I-(TR)^+(TR) \\
(AX)' &= [I-(TR)^+(TR)]' = I-(TR)^+(TR) = AX \\
XA &= H^+R[I-(TR)^+(TR)]R^{-1}H \\
&= H^+RR^{-1}H = H^+H \\
(XA)' &= (H^+H)' = H^+H = XA
\end{aligned}$$

Thus, the lemma follows since all the four Penrose conditions are satisfied. □

The proposition is meaningless if C is singular. Assume that C is nonsingular. Let $C^{1/2}$ be the positive-definite symmetric square-root matrix of C, which always exists and is in fact unique since $C > 0$. Since $B^+ = (B'B)^+B' = B'(BB')^+$ and it can be shown that $(TCT)^+ = T(TCT)^+ = (TCT)^+T$, we have

$$\begin{aligned}
(C^{1/2}T)^+ &= (TC^{1/2}C^{1/2}T)^+TC^{1/2} = (TCT)^+TC^{1/2} \\
&= T(TCT)^+C^{1/2} = T(C^{1/2}T)^+
\end{aligned}$$

Taking transpose of the last equation above leads to

$$(TC^{1/2})^+ = (TC^{1/2})^+T = C^{1/2}(TCT)^+$$

It thus follows that

$$\begin{aligned}
&(H'C^{-1}H)^+H'C^{-1} \\
&= (H'C^{-1/2}C^{-1/2}H)^+(H'C^{-1/2})C^{-1/2} \\
&= (C^{-1/2}H)^+C^{-1/2} \\
&= H^+C^{1/2}[I-(TC^{1/2})^+TC^{1/2}]C^{-1/2} \\
&= H^+C^{1/2}[I-(TC^{1/2})^+C^{1/2}]C^{-1/2} \\
&= H^+[I-C^{1/2}(TC^{1/2})^+] = H^+[I-C(TCT)^+]
\end{aligned}$$

where use has been made of the above lemma. When C is nonsingular, K is unique and the above shows that

$$K = H^+[I-C(TCT)^+] = (H'C^{-1}H)^+H'C^{-1}$$

Finally,

$$\begin{aligned}
P &= KCK' \\
&= [(H'C^{-1}H)^+H'C^{-1}]C[(H'C^{-1}H)^+H'C^{-1}]' \\
&= (H'C^{-1}H)^+
\end{aligned}$$

A.2 Proof of Theorem 4.1

Treat the prior mean $\bar{x}$ as data, that is,

$$\tilde{y} = \begin{bmatrix} \bar{x} \\ y \end{bmatrix} = \begin{bmatrix} I \\ H \end{bmatrix} x + \begin{bmatrix} \bar{x}-x \\ v \end{bmatrix} = \tilde{H}x + \tilde{v}$$

Clearly, $\tilde{H}$ has full column rank and thus BLUE fuser without prior always exists. Consider a particular gain matrix $\tilde{K} = [I-KH, K]$, where K is the gain matrix of the BLUE fuser with complete prior, given by Theorem 2.1. Since

$$\tilde{K}\tilde{H} = [I-KH, K]\begin{bmatrix} I \\ H \end{bmatrix} = I$$

$\hat{x} = \tilde{K}(\tilde{y} - E[\tilde{v}])$ is a linear unbiased fuser without prior, which is not necessarily BLUE. On the other hand, however, we have

$$\begin{aligned}
\hat{x} &= \tilde{K}(\tilde{y} - E[\tilde{v}]) = [I-KH, K]\begin{bmatrix} \bar{x} \\ y-\bar{v} \end{bmatrix} \\
&= (I-KH)\bar{x} + K(y-\bar{v})
\end{aligned}$$

which is equal to the BLUE fuser with complete prior (Theorem 2.1) and thus it minimizes MSE matrix. Therefore, it must be the BLUE fuser without prior due to the uniqueness of this BLUE fuser (Theorem 2.2), although the gain matrix is not necessarily unique.

A.3 Proof of Theorem 4.2

Note first that by Theorem 2.2, BLUE without prior exists iff H has full column rank and thus the nonsingular H_1 always exists. As a result, $y_1 = H_1 x + v_1$ and $H_1^{-1} y_1 = x + H_1^{-1} v_1$ are equivalent. Treating $y_2 = H_2 x + v_2$ as the data model for BLUE with prior and $H_1^{-1} y_1$ as the corresponding prior mean $\bar{x}$, we have $\bar{x} = x + H_1^{-1} v_1 = x + (\bar{x} - x)$ and thus $C_x = H_1^{-1}\text{cov}(v_1)(H_1^{-1})'$, and $C_{xv} := C_{xv_2} = -H_1^{-1}\text{cov}(v_1, v_2)$. Note that $\bar{x}$ so defined is actually random and possibly correlated with y_2 or v_2 but BLUE with complete prior (Theorem 2.1) is still valid in these cases because more fundamentally it is valid for estimating the random variable $\tilde{x} = x - \bar{x}$. By Theorem 4.1, this BLUE with prior coincides almost surely with the BLUE without prior using data $y = Hx + v$.

A.4 Proof of Lemma 4.1

Let $u = \hat{x}_1 - \hat{x}_2$. It is well known that $\hat{x}_1 = \hat{x}_2$ almost surely iff $\bar{u} = 0$ and $C_u = 0$. Here

$$\bar{u} = E[\hat{x}_1 - \hat{x}_2] = E[\hat{x}_1] - E[\hat{x}_2] = \bar{x} - \bar{x} = 0$$

since both $\hat{x}_1$ and $\hat{x}_2$ are unbiased. It is also well known that BLUE with prior is given by $\hat{x}_1 = \bar{x} + K_1(y - \bar{y}), \forall K_1 \in \mathcal{K}$, where $\mathcal{K}$ is the set of $K = C_{xy}C_y^+ + A(I - C_y C_y^+)$ for all compatible matrices A. Since $\hat{x}_2 = K_2(y - \bar{v})$, we have

$$\begin{aligned} C_u &= \text{cov}[\bar{x} + K_1(y - \bar{y}) - K_2(y - \bar{v})] \\ &= \text{cov}[(K_1 - K_2)y] = (K_1 - K_2)C_y(K_1 - K_2)' \end{aligned}$$

Hence, $C_u = 0$ iff $(K_1 - K_2)C_y = 0$, that is, $K_1 C_y = K_2 C_y$ or $[C_{xy}C_y^+ + A(I - C_y C_y^+)]C_y = K_2 C_y$ or $C_{xy}C_y^+ C_y = K_2 C_y$, which is equivalent to $C_{xy} = K_2 C_y$ since $C_{xy}C_y^+ C_y = C_{xy}$.

A.5 Proof of Theorem 4.3

(a) = (b): It follows from Lemma 4.1.

(b) = (c): Since $C_y = HC_{xy} + C_{vy}$, condition (b) (i.e., $K_2 C_y = C_{xy}$) becomes

$$K_2 H C_{xy} + K_2 C_{vy} = C_{xy}$$

or equivalently $C_{xy} + K_2 C_{vy} = C_{xy}$ because $K_2 H = I$ holds for all BLUE without prior, which is $K_2 C_{vy} = 0$.

(c) = (d): $K_2 C_{vy} = 0$ always has a solution and the general solution is

$$K_2 = A(I - C_{vy}C_{vy}^+)$$

where A is any matrix of the same dimension as K_2. Since

$$I = K_2 H = A(I - C_{vy}C_{vy}^+)H$$

the condition becomes

$$A(I - C_{vy}C_{vy}^+)H = I$$

A necessary and sufficient condition for this equation to have a solution for A is

$$[(I - C_{vy}C_{vy}^+)H]^+[(I - C_{vy}C_{vy}^+)H] = I$$

(b) = (e): It is well known that the equation $K_2 C_y = C_{xy}$ has a solution iff $C_{xy}C_y^+ C_y = C_{xy}$, which always holds true. Its general solution is

$$K_2 = C_{xy}C_y^+ + A(I - C_y C_y^+)$$

where A is any matrix of the same dimension as K_2. Since

$$I = K_2 H = C_{xy}C_y^+ H + A(I - C_y C_y^+)H$$

condition $K_2 C_y = C_{xy}$ becomes

$$C_{xy}C_y^+ H + A(I - C_y C_y^+)H = I$$

or

$$A(I - C_y C_y^+)H = I - C_{xy}C_y^+ H$$

A necessary and sufficient condition for this equation to have a solution for A is

$$\begin{aligned} &[I - C_{xy}C_y^+ H][(I - C_y C_y^+)H]^+[(I - C_y C_y^+)H] \\ &= I - C_{xy}C_y^+ H \end{aligned}$$

In essence, the conditions in this theorem are all equivalent to the condition that the gain matrices K_1 and K_2 of BLUE with prior and without both belong to $\mathcal{K}$, where $\mathcal{K}$ is the set of $K = C_{xy}C_y^+ + A(I - C_y C_y^+)$ for all compatible matrices A.

A.6 Proof of Corollary 4.3

It can be shown that Corollary 4.3 follows from Corollary 4.2. However, we provide a more direct proof here. Note that by Proposition 3.1, the MSE matrix of BLUE without prior is $P_2 = (H'C^{-1}H)^{-1}$ iff $C > 0$, while by Theorem 4.1 and Proposition 3.1, the BLUE fuser with complete prior when $\tilde{C}^{-1}$ exists is actually the optimal generalized WLS fuser and thus its MSE matrix is $P_1 = (\tilde{H}'\tilde{C}^{-1}\tilde{H})^{-1}$. Using Schur's identity on the inverse of a partitioned matrix and the fact that $\tilde{C} > 0 \iff \{C > 0, C_x > C_{xv}C^{-1}C_{xv}'\}$ (see, e.g., Theorem 7.7.6 of [5]), it is straightforward to show the following identity

$$\begin{aligned} \tilde{H}'\tilde{C}^{-1}\tilde{H} &= (I + C_{xv}C^{-1}H)'(C_x - C_{xv}C^{-1}C_{xv}')^{-1} \\ &\cdot (I + C_{xv}C^{-1}H) + H'C^{-1}H \end{aligned}$$

Note also that $\tilde{C} > 0 \Longleftrightarrow \det(C) \neq 0$ since $\tilde{C}$ is a covariance matrix. Hence

$$P_1^{-1} - P_2^{-1} = (I + C_{xv}C^{-1}H)'(C_x - C_{xv}C^{-1}C'_{xv})^{-1} \cdot(I + C_{xv}C^{-1}H)$$

The BLUE fuser with complete prior information is (almost surely) identical to BLUE fuser without prior information iff $P_1 = P_2$. Since $C_x - C_{xv}C^{-1}C'_{xv}$ is positive definite under the stated assumption, $P_1 = P_2$ (i.e., $\tilde{H}'\tilde{C}^{-1}\tilde{H} = H'C^{-1}H$) iff $I + C_{xv}C^{-1}H$. This completes the proof.

In fact, the sufficiency that $I + C_{xv}C^{-1}H'$ implies $P_1 = P_2$ can be shown under the weaker assumption that $P_2 = (H'C^{-1}H)^{-1}$ as follows. (4) and (11) imply that

$$\begin{aligned} & (H'C^{-1}H)^{-1}H'C^{-1}C_y \\ \overset{(4)}{=} \; & C_xH' + (H'C^{-1}H)^{-1}H' + C_{xv} \\ & +(H'C^{-1}H)^{-1}H'C^{-1}C'_{xv}H' \\ \overset{(11)}{=} \; & C_xH' + C_{xv} = C_{xy} \end{aligned}$$

Then we have

$$\begin{aligned} P_1 \; &= \; C_x - C_{xy}C_y^{+}C'_{xy} \\ &= \; C_x - (H'C^{-1}H)^{-1}H'C^{-1}C_yC^{-1}H \\ & \quad \cdot(H'C^{-1}H)^{-1} \\ &\overset{(4)}{=} \; C_x - [C_x + (H'C^{-1}H)^{-1} + C_{xv}C^{-1}H \\ & \quad \cdot(H'C^{-1}H)^{-1} + (H'C^{-1}H)^{-1}H'C^{-1}C'_{xv}] \\ &\overset{(11)}{=} -(H'C^{-1}H)^{-1} + 2(H'C^{-1}H)^{-1} = P_2 \end{aligned}$$

A.7 Proof of Proposition 4.2

Treating prior mean $\bar{x}$ as extra data in the model $\bar{x} = x + v_0$ with $\text{cov}(v_0) = \text{cov}(\bar{x} - x) = C_x$. Let $\tilde{y}$, $\tilde{H}$, $E[\tilde{v}] = [0, \bar{v}']'$, and $\tilde{C}$ be as defined in Theorem 4.1. Then, by Theorem 3.1, for data model $\tilde{y} = \tilde{H}x + \tilde{v}$ BLUE without prior becomes optimal WLS if $\tilde{C}$ is nonsingular. In other words, the original optimal generalized WLS problem of estimating a *random* variable with prior information becomes an optimal WLS problem without prior information.

A.8 Proof of Theorem 5.1

It follows from $\text{MSE}(\hat{x}) = C_x - C_{xy}C_y^{+}C'_{xy}$ and $C_{xy} = C_xH' + C_{xv}$ for linear data model $y = Hx + v$ that y is not useful for BLUE fusion if $C_xH' + C_{xv} = 0$. Using Schur's identity on the inverse of a partitioned matrix and the fact that $\tilde{C} > 0 \Longleftrightarrow \{C_x > 0, C > C'_{xv}C_x^{-1}C_{xv}\}$ (see, e.g., Theorem 7.7.6 of [5]), it is straightforward to show the following identity

$$\tilde{H}'\tilde{C}^{-1}\tilde{H} = (H + C'_{xv}C_x^{-1})'(C - C'_{xv}C_x^{-1}C_{xv})^{-1} \cdot(H + C'_{xv}C_x^{-1}) + C_x^{-1}$$

Note also that $\tilde{C} > 0 \Longleftrightarrow \det(C) \neq 0$ since $\tilde{C}$ is a covariance matrix. Hence

$$P^{-1} - C_x^{-1} = (I + C_{xv}C^{-1}H)'(C_x - C_{xv}C^{-1}C'_{xv})^{-1} \cdot(I + C_{xv}C^{-1}H)$$

References

[1] H. Chen, T. Kirubarajan, and Y. Bar-Shalom, "Centralized vs. Distributed Tracking Algorithms for Air to Air Scenarios," in *Proc. 2000 SPIE Conf. on Signal and Data Processing of Small Targets*, vol. 4048, (Orlando, Florida, USA), April 2000.

[2] H. Chen, T. Kirubarajan, and Y. Bar-Shalom, "Performance Limits of Track-to-Track Fusion vs. Centralized Estimation: Theory and Application," *IEEE Trans. Aerospace and Electronic Systems* (to appear), 38, 2002.

[3] C. Y. Chong, "Hierarchical Estimation," in *Proc. Second MIT/ONR Workshop on C3*, (Monterey, CA), July 1979.

[4] C. Y. Chong, K. C. Chang, and S. Mori, "Distributed Tracking in Distributed Sensor Networks," in *Proc. 1986 American Control Conf.*, (Seattle, WA), June 1986.

[5] R. A. Horn and C. Johnson, *Matrix Analysis*. Cambridge, UK: Cambridge University Press, 1985.

[6] X. R. Li and J. Wang, "Unified Optimal Linear Estimation Fusion—Part II: Discussions and Examples," in *Proc. 2000 International Conf. on Information Fusion*, (Paris, France), pp. MoC2.18–MoC2.25, July 2000.

[7] X. R. Li and K. S. Zhang, "Optimal Linear Estimation Fusion—Part IV: Optimality and Efficiency of Distributed Fusion," in *Proc. 2001 International Conf. on Information Fusion*, (Montreal, QC, Canada), pp. WeB1–19–WeB1–26, Aug. 2001.

[8] X. R. Li and P. Zhang, "Optimal Linear Estimation Fusion—Part III: Cross-Correlation of Local Estimation Errors," in *Proc. 2001 International Conf. on Information Fusion*, (Montreal, QC, Canada), pp. WeB1–11–WeB1–18, Aug. 2001.

[9] X. R. Li, Y. M. Zhu, and C. Z. Han, "Unified Optimal Linear Estimation Fusion—Part I: Unified Models and Fusion Rules," in *Proc. 2000 International Conf. on Information Fusion*, (Paris, France), pp. MoC2.10–MoC2.17, July 2000.

[10] J. L. Speyer, "Computation and Transmission Requirements for a Decentralized Linear-Quadratic-Gaussian Control Problem," *IEEE Trans. Automatic Control*, AC-24:266–269, Apr. 1979.

[11] K. S. Zhang, X. R. Li, and Y. M. Zhu, "Optimal Update with Out-of-Sequence Observations for Distributed Filtering," in *Proc. 2002 International Conf. on Information Fusion*, (Annapolis, MD, USA), July 2002.

[12] Y. M. Zhu, Z. S. You, J. Zhao, K.-S. Zhang, and X. R. Li, "The Optimality for the Distributed Kalman Filtering Fusion with Feedback," *Automatica*, 37:1489–1493, 2001.