Optimal Estimation and Precision Analysis of Measuring Data Fusion Model

Xuanying Zhou, Jiongqi Wang, Zhengming Wang and Zhangming He

College of Science, National University of Defense Technology, Changsha, Hunan, China.

Abstract

Data fusion is an effective method to improve the data processing accuracy, and the fusion weight has a great influence to the accuracy of the fusion estimation. In this paper, we study the problem of optimal weight and parameter estimation for the linear fusion model with unequal precision measuring data. The properties of the unbiased estimation are discussed for linear model, and then the optimal estimation for the fusion model with unequal precision measuring data is given. It is proved that there exists the optimal fusion weight and it is unique. Besides, the accuracy of the optimal estimation for the multivariate linear model is analyzed and some conclusions suitable for practical application are obtained, which can provide the theory foundation for the experiment design and the data selection. Finally, two simulations are offered to validate the theories conclusions in this paper.

Keywords Data fusion; Regression model; Fusion weight; Parameter estimation; Precision analysis.

1 Introduction

The main purpose of data fusion is to improve the accuracy of measuring data, to establish a proper processing model, and to give an effective and reliable fusion algorithm [1, 2]. Fusion with different types and unequal precision data is the most typical situation in the data fusion processing [3-5]. After the data are modeled in a parametric model, the data fusion problem can be transferred into the parameter estimation problem [6,7].

To evaluate the performance of the parameter estimation result, we need an evaluation standard, i.e., evaluation criterion or optimal criterion. Such criteria include Minimum Mean-square Error (MSE) criterion, Maximum Likelihood (ML) criterion, Maximum a Posteriori (MAP) criterion, Best Linear Unbiased Estimation (BLUE) criterion and Least Squares Estimation (LSE) criterion, etc. Whether the estimated parameter satisfies the need of the application depends on the estimation criteria as well as the data accuracy and the model properties [8]. Obviously, the selection of the evaluation criterion is affected by the characteristic of the estimation parameter, the demand of the estimation accuracy and the complexity of the estimation algorithm. Specifically, the parameters, estimated according to LSE criterion, will lead to the minimal norm of the observation residual, i.e. the differencebetween the observed value and the calculated value. Usually, LSE does not involve the dynamic and statistical information of the parameter to be estimated. Therefore, LSE is easy implemented but with low estimation accuracy. Nevertheless, when we are short of the error information about the measuring data, LSE also can provide us with an acceptable solution. MSE criterion is the best in terms of that MSE has the minimal mean square error. However, this method needs some statistical prior information, such as the first and the second moment of the data and parameter. The MAP and the ML estimation are both related to the conditional probability density functions, and the estimation is hard to be obtained except for some special cases.

Therefore, in the actual application, the efficient and reliable data fusion algorithm for data fusion should be selected according to the specific situation.

Although most of the fusion systems are nonlinear, they can be linearized into some linear regression model when proper base functions are selected or the nonlinear iterative means are adopted. That is to say, nonlinear fusion problem can be approximated to process with the linear fusion problem.

Furthermore, when the parameters and the measuring data are with the normal distribution, some optimal estimation methods, such as BLUE, MSE and LSE, are equivalence to each other [10]. Therefore, the minimal linear variance criterion is usually applied for the actual application.

Following the introduction in Section 1, the structure of the paper is organized as follows. The form of unbiased estimation for linear model is given and the estimation characters are discussed in Section 2. In Section 3, the optimal weight and parameter estimation of unequal-precision data fusion are researched. Besides, it is proved that there exists the optimal fusion weight and it is unique, and the accuracy of multivariable optimal estimation for linear fusion model is analyzed. Section 4 provides two numerical examples to validate the proposed theory and method. Finally, the paper is concluded in Section 5.

2 Unbiased Estimate of Linear Model

Consider the linear measuring regression model as follow:

$$Y = X\beta + \varepsilon, \quad \varepsilon \sim (0, \sigma^2 \mathrm{I}) \tag{1}$$

where $Y = Y_{m\times 1}$ is the measuring data, $X = X_{m\times n}$ is the design matrix and $\mathrm{rank}(X) = n$, $\beta = \beta_{n\times 1}$ is the estimated parameter vector, and $\varepsilon = \varepsilon_{m\times 1}$ is the measuring random error vector with the zero expectation and diagonal covariance, i.e. $\varepsilon \sim (0, \sigma^2 \mathrm{I})$.

For the parameter estimation problem in model (1), the LSE
$\hat{\beta}_{LS} = \left(X^{\mathrm{T}} X\right)^{-1} X^{\mathrm{T}} Y$ has some good properties as follows:

Property 1: $\hat{\beta}_{LS}$ is the UMVUE (Uniform Minimum Variance Unbiased Estimation) of the parameter β, and moreover, $\forall c \in R^n, c^{\mathrm{T}}\hat{\beta}_{LS}$ is the linear unbiased estimation of the parameter $c^{\mathrm{T}}\beta$.

Property 2: Assuming $\mathrm{E}\left\|X\hat{\beta}_{LS} - X\beta\right\|^2 = (X\hat{\beta}_{LS} - X\beta)^{\mathrm{T}}(X\hat{\beta}_{LS} - X\beta)$ and then $\mathrm{E}\left\|X\hat{\beta}_{LS} - X\beta\right\|^2 = n\sigma^2$.

Property 3: Assuming $\mathrm{Cov}(\hat{\beta}_{LS}) = \mathrm{E}(\hat{\beta}_{LS} - \beta)(\hat{\beta}_{LS} - \beta)^{\mathrm{T}}$, $\mathrm{MSE}(\hat{\beta}_{LS}) = \mathrm{E}(\hat{\beta}_{LS} - \beta)^{\mathrm{T}}(\hat{\beta}_{LS} - \beta)$ and then $\mathrm{Cov}(\hat{\beta}_{LS}) = \sigma^2 S^{-1}, \mathrm{MSE}(\hat{\beta}_{LS}) = \sigma^2 \sum\limits_{i=1}^{n} \lambda_i^{-1}$, where $S = X^{\mathrm{T}}X$ and $\lambda_i, i = 1, \cdots, n$ are all eigenvalues of the matrix S;

Remark 1: Property 1 shows that in the actual engineering application, d-ifferent constant vector c can be chosen for estimating some components or the linear combination of the parameter β.

Remark 2: Property 2 shows that the estimation precision of $X\hat{\beta}_{LS}$ is directly proportional to n. That is to say, the more the number of the estimated parameters, the lower estimation precision will be obtained. Therefore, the proper base function and parameter model should be chosen when modeling the measuring data to let the parameter number as few as possible. The sparse parameter modeling methods are often used in the real application [11, 12].

Remark 3:Property 3 shows that when the model (1) is multi-collinearity, i.e. the matrix $X^{\mathrm{T}}X$ has some extremely small eigenvalues λ_i, the large $\mathrm{MSE}(\hat{\beta}_{LS})$ or $\mathrm{MSE}(c^{\mathrm{T}}\hat{\beta}_{LS})$ will lead to the bad estimation accuracy. The regularizing methods, a series of biased estimation methods, were proposed to handle the multi-collinear problem in application [13-15]. By choosing the proper regularizing factor μ and the regularizing matrix D with full column rank to solve the following optimization problem

$$\min_{\beta \in \mathrm{R}^n, \mu > 0} \|Y - X\beta\|_2^2 + \mu \|D\beta\|_2^2 \tag{2}$$

The solution of (2) can be easily calculated as follows

$$\hat{\beta}_R = (X^{\mathrm{T}}X + \mu D^{\mathrm{T}}D)^{-1}X^{\mathrm{T}}Y \tag{3}$$

Compared to LSE, the regularizing estimation has properties as follows:

Property 4: $\mathrm{E}\hat{\beta}_R = (X^{\mathrm{T}}X + \mu D^{\mathrm{T}}D)^{-1}X^{\mathrm{T}}X\beta$. i.e. the regularizing parameter estimation is biased;

Property 5: There exists μ, D, and make $\mathrm{MSE}(\hat{\beta}_R) < \mathrm{MSE}(\hat{\beta}_{LS})$, i.e. the regularizing estimation can better than LSE by choosing some proper regularization parameter and matrix.

In the linear measuring regression model, $\varepsilon \sim (0, \sigma^2\mathrm{I})$ means the measures are irrelevant and the precision are equal. In actual, if the measures are relevant and have unequal precision, the model (1) can be transferred to

$$Y = X\beta + \varepsilon, \varepsilon \sim (0, \sigma^2 G) \tag{4}$$

where σ^2 is known or unknown, and G is a known positive definite matrix. For model (4), its UMVUE is the weighted least squares estimation (WLSE)$\tilde{\beta}_{WLS} = \left(X^{\mathrm{T}}G^{-1}X\right)^{-1}X^{\mathrm{T}}G^{-1}Y$, and $\mathrm{MSE}(\tilde{\beta}_{WLS}) = \mathrm{tr}\left(X^{\mathrm{T}}G^{-1}X\right)^{-1}$.

In real application, in order to get the LSE for the linear fusion model, the measuring data should be parametric modeling to make it satisfy the model (1) or (4). Actually, a typical application of model (4) is the unequal precision data fusion processing problem with several kinds of measuring equipment. Although the measuring equations of different equipment are non-linear, the proper basis function can be chosen or the nonlinear iterative means can be adopted to linearize the measuring equations. Therefore, the LSE or WLSE can be an important theory foundation for the measuring data fusion.

No matter the parameters estimated by model (1) or (4), the statistic properties of the measuring random error, including the correlation, meaning, variance, and covariance and so on, need to be estimated at first. σ^2 in the model (1) or (4) reflects the accuracy of the measuring data. Therefore, as the base of unequal precision data fusion, the estimation of the parameter σ^2 is very important. Besides, when the estimation performance of the LSE (WLSE) is worse, the information of the parameter σ^2 is also needed to be used in order to build the biased estimation of the parameter β. There is the property about the estimation of σ^2:

Property 6:Assume the observation error in model (1) satisfy the normal distribution, i.e. $\varepsilon \sim \mathrm{N}(0, \sigma^2 \mathrm{I})$, then $\hat{\sigma}^2 = RSS/(m-n)$,where $RSS = \sum\limits_{i=1}^{m} \mu_i^2 = \left\| Y - X\hat{\beta}_{LS} \right\|^2$ is the measuring residual square sum, $\mu_i = y_i - X_i\hat{\beta}_{LS}, i = 1, \cdots, m$ is the i^{th} residual between of the measuring data and the calculated value and $\mathrm{E}\hat{\sigma}^2 = \sigma^2$, $\mathrm{MSE}(\hat{\sigma}^2) = 2\sigma^4 / (m-n)$.

Remark 4:Property 6 shows that the parameters σ^2 can be estimated by the residual if the precision of the actual measuring data is unknown. The estimation value and its accuracy of the parameter σ^2 are related to the number of the measuring data as well as the dimensional of the estimated parameter. As a result, in the actual application, the estimation variance can be decreased by increasing the sampling number of the measuring data.

3 Optimal Fusion Estimation of Linear Model with Unequal Precision Data

In many measuring processing problem, like trajectory tracking, the unequal precision data fusion processing often need to be considered. Obviously, the weighting methods for different measuring data have a great influence to the accuracy of the fusion estimation.

3.1 Optimal Estimation in Unequal Precision Linear Fusion Model

Considering s kinds of unequal precision linear data fusion model:

$$\begin{cases} Y_1 = X_1\beta + \varepsilon_1, \varepsilon_1 \sim (0, \sigma_1^2 \mathrm{I}_{m_1}), \\ \cdots\cdots\cdots \\ Y_s = X_s\beta + \varepsilon_s, \varepsilon_s \sim (0, \sigma_s^2 \mathrm{I}_{m_s}), \\ \mathrm{E}\varepsilon_i {\varepsilon_j}^{\mathrm{T}} = O_{m_i \times m_j}, i, j = 1, \cdots, s, i \neq j \end{cases} \tag{5}$$

The definitions of the parameter in (5) are same to that in model (1), and the optimal estimation can be given by the follow theorem.

Theorem 1: For model (5), $\forall c \in \mathrm{R}^n$, the uniformly minimum variance estimation of $c^{\mathrm{T}}\beta$ is $c^{\mathrm{T}}\tilde{\beta}_f$ where

$$\tilde{\beta}_f = \Big(\sum_{i=1}^{s} \frac{\sigma_i^{-2}}{\sum\limits_{i=1}^{s} \sigma_i^{-2}} X_i^{\mathrm{T}} X_i\Big)^{-1} \Big(\sum_{i=1}^{s} \frac{\sigma_i^{-2}}{\sum\limits_{i=1}^{s} \sigma_i^{-2}} X_i^{\mathrm{T}} Y_i\Big) \tag{6}$$

Proof: Assuming $t = 1 \Big/ \sqrt{\sum\limits_{i=1}^{s} \sigma_i^{-2}}$, then model (6) can be rewritten as:

$$\begin{cases} t\sigma_1^{-1} Y_1 = t\sigma_1^{-1} X_1\beta + t\sigma_1^{-1}\varepsilon_1, t\sigma_1^{-1}\varepsilon_1 \sim (0, t^2 \mathrm{I}_{m_1}), \\ \cdots\cdots\cdots \\ t\sigma_s^{-1} Y_s = t\sigma_s^{-1} X_s\beta + t\sigma_s^{-1}\varepsilon_s, t\sigma_s^{-1}\varepsilon_s \sim (0, t^2 \mathrm{I}_{m_s}), \end{cases} \tag{7}$$

Assuming $Y = [t\sigma_1^{-1} Y_1^{\mathrm{T}}, \cdots, t\sigma_s^{-1} Y_s^{\mathrm{T}}]^{\mathrm{T}}, X = [t\sigma_1^{-1} X_1^{\mathrm{T}}, \cdots, t\sigma_s^{-1} X_s^{\mathrm{T}}]^{\mathrm{T}}, \varepsilon = [t\sigma_1^{-1}\varepsilon_1^{\mathrm{T}}, \cdots, t\sigma_s^{-1}\varepsilon_s^{\mathrm{T}}]^{\mathrm{T}}$, combining with (6) and (7), the fusion model (5) can be written as follows:

$$Y = X\beta + \varepsilon, \varepsilon \sim (0, t^2 \mathrm{I}_m), m = \sum_{i=1}^{s} m_i \tag{8}$$

Using the LSE, the theorem can be proved that $\forall c \in \mathrm{R}^n$, the uniformly minimum variance estimation of $c^{\mathrm{T}}\beta$ is $c^{\mathrm{T}}\tilde{\beta}_f$, where

$$\tilde{\beta}_f = (X^{\mathrm{T}} X)^{-1} X^{\mathrm{T}} Y = \Big(\sum_{i=1}^{s} t^2 \sigma_i^{-2} X_i^{\mathrm{T}} X_i\Big)^{-1} \Big(\sum_{i=1}^{s} t^2 \sigma_i^{-2} X_i^{\mathrm{T}} Y_i\Big) \tag{9}$$

The proof is completed.

3.2 Optimal Weight for the Linear Fusion Model

The purpose of data fusion is to find the optimal weight ρ_i and then to optimize the fusion problem and obtain the optimal parameter estimation. Theorem 1 above shows that the optimal weight of the data fusion model with unequal precision measuring data is $\rho_i = \sigma_i^{-1} \Big/ \sqrt{\sum\limits_{i=1}^{s} \sigma_i^{-2}} = t\sigma_i^{-1}$, and moreover, it satisfies $\sum\limits_{i=1}^{s} \rho_i^2 = 1$. This indicates that the optimal only related to the data accuracy σ_i^{-1}.

For convenient, the weight method for two types of unequal-precision linear observed data is discussed firstly:

$$\begin{cases} Y_1 = X_1\beta + \varepsilon_1, \varepsilon_1 \sim (0, \sigma_1^2 \mathrm{I}_{m_1}), \\ Y_2 = X_2\beta + \varepsilon_2, \varepsilon_2 \sim (0, \sigma_2^2 \mathrm{I}_{m_2}), \\ \mathrm{E}\varepsilon_1\varepsilon_2^{\mathrm{T}} = O_{m_1 \times m_2} \end{cases} \tag{10}$$

Considering the following optimization problem:

$$\begin{cases} \arg\min\limits_{\beta} \sum\limits_{i=1}^{2} \rho_i^2 \|Y_i - X_i\beta\|^2 \\ \sum\limits_{i=1}^{2} \rho_i^2 = 1 \end{cases} \tag{11}$$

And the solution can be easily obtained as follow:

$$\hat{\beta}(\rho) = (\sum_{i=1}^{2} \rho_i^2 X_i^{\mathrm{T}} X_i)^{-1} (\sum_{i=1}^{2} \rho_i^2 X_i^{\mathrm{T}} Y_i) \tag{12}$$

Theorem 2: Under the assumption of model (10), the solution of the optimization problem $\arg\min\limits_{\rho} \mathrm{E}\left\|\hat{\beta}(\rho) - \beta\right\|^2 = \arg\min\limits_{\rho} \mathrm{MSE}(\hat{\beta}(\rho))$ is:

$$\rho_i = \sigma_i^{-1} \Big/ \sqrt{\sigma_1^{-2} + \sigma_2^{-2}}, i = 1, 2 \tag{13}$$

Proof: By calculating $\mathrm{E}\left\|\hat{\beta}(\rho) - \beta\right\|^2 = \mathrm{tr}(\sum\limits_{i=1}^{2} \rho_i^2 X_i^{\mathrm{T}} X_i)^{-2} (\sum\limits_{i=1}^{2} \rho_i^4 \sigma_i^2 X_i^{\mathrm{T}} X_i)$ assuming $A = X_1^{\mathrm{T}} X_1$, $B{=}X_2^{\mathrm{T}} X_2$, and then

$$\begin{aligned} f(\rho_1, \rho_2) &= (\rho_1^2 A + \rho_2^2 B)^{-1} (\rho_1^4 \sigma_1^2 A + \rho_2^4 \sigma_2^2 B)(\rho_1^2 A + \rho_2^2 B)^{-1} \\ &= \sigma_2^2 \left(\frac{\rho_1^2}{\rho_2^2} A + B\right)^{-1} \left(\frac{\rho_1^4}{\rho_2^4}\frac{\sigma_1^2}{\sigma_2^2} A + B\right) \left(\frac{\rho_1^2}{\rho_2^2} A + B\right)^{-1} \end{aligned} \tag{14}$$

As both of A and B are real symmetric positive definite matrices, and can be similarity diagonalized simultaneously, that is to say existing an invertible matrix

P, let $A = P^{\mathrm{T}}\Lambda P$, $B = P^{\mathrm{T}}P$, where $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$, therefore

$$\begin{aligned} f(\rho_1,\rho_2) &= \sigma_2^2 P^{\mathrm{T}}\left(\tfrac{\rho_1^2}{\rho_2^2}\Lambda + \mathrm{I}\right)^{-1}\left(\tfrac{\rho_1^4}{\rho_2^4}\tfrac{\sigma_1^2}{\sigma_2^2}\Lambda + \mathrm{I}\right)\left(\tfrac{\rho_1^2}{\rho_2^2}\Lambda + \mathrm{I}\right)^{-1}(P^{-1})^{\mathrm{T}} \\ &= \sigma_2^2 P^{\mathrm{T}}\mathrm{diag}\left[\left(\tfrac{\rho_1^4}{\rho_2^4}\tfrac{\sigma_1^2}{\sigma_2^2}\lambda_i + 1\right)\left(\tfrac{\rho_1^2}{\rho_2^2}\lambda_i + 1\right)^{-2}\right](P^{-1})^{\mathrm{T}} \end{aligned} \tag{15}$$

Assuming that $t = \frac{\rho_1^2}{\rho_2^2}$, $\sigma = \frac{\sigma_1}{\sigma_2}$, $\lambda_i = a$, $g(t) = (t^2\sigma^2 a + 1)\big/(ta+1)^2$, let $\frac{\mathrm{d}g(t)}{\mathrm{d}t} = \frac{2a(t\sigma^2-1)}{(ta+1)^3} = 0$ and then $t = \sigma^{-2}$. Besides,

$$\frac{\mathrm{d}^2 g(t)}{\mathrm{d}t^2}\Big|_{t=\sigma^{-2}} = \frac{2a(\sigma^2 + a)}{(ta+1)^4} > 0 \tag{16}$$

Noticed that $\rho_1^2 + \rho_2^2 = 1$, and the function $g(t)$ gets the minimum value at $\rho_i = \sigma_i^{-1}\Big/\sqrt{\sigma_1^{-2} + \sigma_2^{-2}}, i = 1, 2$, then

$$\begin{aligned} \mathrm{E}\left\|\hat{\beta}(\rho) - \beta\right\|^2 &= \mathrm{tr}\left(\sum_{i=1}^{2}\rho_i^2 X_i^{\mathrm{T}}X_i\right)^{-2}\left(\sum_{i=1}^{2}\rho_i^4\sigma_i^2 X_i^{\mathrm{T}}X_i\right) \\ &= \mathrm{tr}\left[\sigma_2^2 P^{\mathrm{T}}\mathrm{diag}\left[\left(\tfrac{\rho_1^4}{\rho_2^4}\tfrac{\sigma_1^2}{\sigma_2^2}\lambda_i + 1\right)\left(\tfrac{\rho_1^2}{\rho_2^2}\lambda_i + 1\right)^{-2}\right](P^{-1})^{\mathrm{T}}\right] \\ &= \sigma_2^2\sum_{i=1}^{n}\tfrac{t^2\sigma^2\lambda_i + 1}{(t\lambda_i+1)^2} \end{aligned} \tag{17}$$

Since the each item in equation (17) gets the minimum at $\rho_i = \sigma_i^{-1}\Big/\sqrt{\sigma_1^{-2} + \sigma_2^{-2}}, i = 1, 2$, therefore, the solution of $\arg\min_{\rho}\mathrm{E}\left\|\hat{\beta}(\rho) - \beta\right\|^2$ is $\rho_i = \sigma_i^{-1}\Big/\sqrt{\sigma_1^{-2} + \sigma_2^{-2}}, i = 1, 2$. The proof is complete.

Remark 5: The conclusion of Theorem 2 has important application value. In the actual problem, unequal precision data are usually measured, so the data weight has an important effect to the data fusion accuracy. Theorem 2 shows that the unique optimal fusion weigh depends on the data precision in the linear fusion model with unequal precision measuring data. Actually, this is the Gauss-Markov theorem applies to the LSE for the linear model. However, the Gauss-Markov theorem shows the optimal estimation only can be given when the data accuracy is known, and while Theorem 2 shows the optimal estimation can be obtained by solving the optimal problem $\arg\min_{\rho}\mathrm{E}\left\|\hat{\beta}(\rho) - \beta\right\|^2$ when the data precision is unknown. That is to say, for model (10), $\rho_i = \sigma_i^{-1}\Big/\sqrt{\sigma_1^{-2} + \sigma_2^{-2}}, i = 1, 2$ is the necessary and sufficient condition if $\hat{\beta}(\rho) = \left(\sum_{i=1}^{2}\rho_i^2 X_i^{\mathrm{T}}X_i\right)^{-1}\left(\sum_{i=1}^{2}\rho_i^2 X_i^{\mathrm{T}}Y_i\right)$ is the uniformly minimum variance solution (optimal solution) for the parameter β.

From Theorem 2, for the linear fusion model (5), assuming $\rho = [\rho_1, \cdots, \rho_s]$, the parameter estimation and optimal weight determination can be handled by the following two-step minimal problem

$$(1) \begin{cases} \arg\min_{\beta} \sum_{i=1}^{s} \rho_i^2 \|Y_i - X_i\beta\|^2 \\ \sum_{i=1}^{s} \rho_i^2 = 1 \end{cases} \tag{18}$$

$$(2) \quad \arg\min_{\rho_i} \mathrm{E}\left\|\tilde{\beta}(\rho) - \beta\right\|^2 \tag{19}$$

3.3 Accuracy Analysis of Optimal Fusion Estimation

For convenient, consider the parameter estimation accuracy problem of two kinds of unequal precision data fusion model. Suppose that $\hat{\beta}(i), i = 1, 2$ are the estimation by measuring $Y_i, i = 1, 2$, separately, $\hat{\beta}(1, 2)$ is the traditional joint estimation of these two kinds of measuring data, and $\hat{\beta}_f$ is the optimal fusion estimation, i.e.

$$\begin{aligned} &\hat{\beta}(1) = (X_1^{\mathrm{T}} X_1)^{-1} X_1^{\mathrm{T}} Y_1 \\ &\hat{\beta}(2) = (X_2^{\mathrm{T}} X_2)^{-1} X_2^{\mathrm{T}} Y_2 \\ &\hat{\beta}(1, 2) = (X_1^{\mathrm{T}} X_1 + X_2^{\mathrm{T}} X_2)^{-1} (X_1^{\mathrm{T}} Y_1 + X_2^{\mathrm{T}} Y_2) \\ &\hat{\beta}_f = (\sigma_1^{-2} X_1^{\mathrm{T}} X_1 + \sigma_2^{-2} X_2^{\mathrm{T}} X_2)^{-1} (\sigma_1^{-2} X_1^{\mathrm{T}} Y_1 + \sigma_2^{-2} X_2^{\mathrm{T}} Y_2) \end{aligned} \tag{20}$$

Then the follow conclusions can be drawn.

Theorem 3: For the different estimation for the parameter β there are:

$$(1)\mathrm{E}\left\|\hat{\beta}_f - \beta\right\|^2 \leq \min\{\mathrm{E}\left\|\hat{\beta}(1) - \beta\right\|^2,\ \mathrm{E}\left\|\hat{\beta}(2) - \beta\right\|^2,\ \mathrm{E}\left\|\hat{\beta}(1, 2) - \beta\right\|^2\} \tag{21}$$

$$(2) \quad \mathrm{E}\left\|\hat{\beta}(1, 2) - \beta\right\|^2 < \max\{\mathrm{E}\left\|\hat{\beta}(1) - \beta\right\|^2,\ \mathrm{E}\left\|\hat{\beta}(2) - \beta\right\|^2\} \tag{22}$$

(3)If $\sigma_2^2/\sigma_1^2 \leq 2$, and then

$$\mathrm{E}\left\|\hat{\beta}(1, 2) - \beta\right\|^2 \leq \min\{\mathrm{E}\left\|\hat{\beta}(1) - \beta\right\|^2,\ \mathrm{E}\left\|\hat{\beta}(2) - \beta\right\|^2\} \tag{23}$$

Proof: (1) From the LSE properties:

$$\begin{aligned} &\mathrm{E}\left\|\hat{\beta}(1) - \beta\right\|^2 = \sigma_1^2 \mathrm{tr}(X_1^{\mathrm{T}} X_1)^{-1}, \mathrm{E}\left\|\hat{\beta}(2) - \beta\right\|^2 = \sigma_2^2 \mathrm{tr}(X_2^{\mathrm{T}} X_2)^{-1} \\ &\mathrm{E}\left\|\hat{\beta}(1, 2) - \beta\right\|^2 = \mathrm{tr}(\sigma_1^2 X_1^{\mathrm{T}} X_1 + \sigma_2^2 X_2^{\mathrm{T}} X_2)(X_1^{\mathrm{T}} X_1 + X_2^{\mathrm{T}} X_2)^{-2} \\ &\mathrm{E}\left\|\hat{\beta}_f - \beta\right\|^2 = \mathrm{tr}(\sigma_1^{-2} X_1^{\mathrm{T}} X_1 + \sigma_2^{-2} X_2^{\mathrm{T}} X_2)^{-1} \end{aligned} \tag{24}$$

Obviously, $\mathrm{E}\left\|\hat{\beta}_f-\beta\right\|^2 \le \min\{\mathrm{E}\left\|\hat{\beta}(1)-\beta\right\|^2, \mathrm{E}\left\|\hat{\beta}(2)-\beta\right\|^2\}$ and furthermore, from Theorem 2, $\mathrm{E}\left\|\hat{\beta}_f-\beta\right\|^2 \le \mathrm{E}\left\|\hat{\beta}(1,2)-\beta\right\|^2$.

(2) Assuming $\mathrm{E}\left\|\hat{\beta}(1)-\beta\right\|^2 \le \mathrm{E}\left\|\hat{\beta}(2)-\beta\right\|^2$ and $A = X_1^{\mathrm{T}}X_1,\ B = X_2^{\mathrm{T}}X_2$, $\mathrm{E}\left\|\hat{\beta}(1,2)-\beta\right\|^2 \le \mathrm{E}\left\|\hat{\beta}(2)-\beta\right\|^2$ need to be proved, the follow inequality need to be proved first:

$$\mathrm{tr}(\sigma_1^2A+\sigma_2^2B)(A+B)^{-2} \le \sigma_2^2\mathrm{tr}(B^{-1}) \tag{25}$$

That is

$$\mathrm{tr}[(\sigma_1^2A+\sigma_2^2B)-(A+B)\sigma_2^2B^{-1}(A+B)] \le 0 \tag{26}$$

Noticed that $\sigma_1^2A+\sigma_2^2B-(A+B)\sigma_2^2B^{-1}(A+B) = A(\sigma_1^2A^{-1}-2\sigma_2^2A^{-1}-\sigma_2^2B^{-1})A < 0$.

Then equation (26) is right, and thus equation (22) is proved.

(3) Following the symbol and assumption in (2), if want to prove $\mathrm{E}\left\|\hat{\beta}(1,2)-\beta\right\|^2 \le \mathrm{E}\left\|\hat{\beta}(1)-\beta\right\|^2$, i.e. to prove $\mathrm{tr}(\sigma_1^2A+\sigma_2^2B)(A+B)^{-2} \le \sigma_1^2\mathrm{tr}(A^{-1})$

That is

$$\mathrm{tr}[(\sigma_1^2A+\sigma_2^2B)-(A+B)\sigma_1^2A^{-1}(A+B)] \le 0 \tag{27}$$

Noticed that when $\sigma_2^2/\sigma_1^2 \le 2$, and then $\sigma_1^2A+\sigma_2^2B-(A+B)\sigma_1^2A^{-1}(A+B) = (\sigma_2^2-2\sigma_1^2)B^{-1}-\sigma_1^2BA^{-1}B < 0$, then equation (27) as well as equation (23) is proved. The proof is complete.

Obviously, the conclusion in Theorem 3 can also be adapted to s kinds of unequal precision data fusion processing problem. It has the great effect to experiment designed and data selection scheme optimization problem in the actual application.

Equation (21) shows that the accuracy of several sensors optimal fusion estimation is the best comparing to the any single or any combination sensors joint estimation. And Equation (22) shows that the precision of several sensors joint (traditional weighted scheme) estimation is better than the worst single sensors estimation, but the estimation precision of several sensors joint can better than the best single sensors if each sensors measuring accuracy reaches some certain conditions.

4 Numerical Examples

In this section, two calculation examples are given to validate the proposed theory and algorithm for optimal weight and parameter estimation of unequal-precision data fusion.

Example 1: Fusion processing of static measuring data

Assuming two unequal-precision equipment measure the physical signal β. Suppose the real value of the signal β is 10, and randomly create 100 high-precision measuring data (the root mean square error, RMS, is 3), 100 medium-precision measuring data (the RMS is 4) and 100 low-precision measuring data (the RMS is 6), and simulate 100 times. The estimation of β and its variance is get. Seen in the Table 1 as follow (the root variance is come from the 100 simulate data statistic)

the true value of the physical quantity is $\beta = 10$. We have 100 groups of data. In each group, there are 100 high-precision data samples (the standard deviation of is 3), 100 medium-precision data samples (the standard deviation of is 4) and 100 low-precision data samples (the standard deviation of is 6). The parameter estimated and its MSE is shown in Table 1 below. (The estimated variance is obtained based on statistics of 100 groups of observed data.)

Table 1 Parameter Estimate Result in Different Weighted Methods

Method / Result	High-precision only	Medium-precision only	Low-precision only	Traditional weighted jointestimation with high- and medium-precision data	Traditional weighted joint estimation with high- and low-precision data	Optimal fusion estimation
Truth Parameter	10	10	10	10	10	10
Estimated Value	9.975	10.087	10.114	9.983	10.052	9.995
Mean Square Error	0.087	0.131	0.154	0.071	0.116	0.045

In the linear measuring data fusion processing, the unique fusion weight depends on the measuring data accuracy. Solve the minimum optimization problem (18) and (19), and get the MSE of the estimated parameter is smallest, 0.045. For the high- and medium-precision data fusion, the precision of two data satisfies $\sigma_2^2/\sigma_1^2 < 2$, therefore, the MSE of the parameter with the traditional joint weighted method is 0.071, which is better than that only with the high-precision data, 0.087. For the high- and low-precision data fusion, the precision of two data does not dissatisfy $\sigma_3^2/\sigma_1^2 < 2$, the MSE of the estimated parameter with the traditional weight joint method is 0.116, which is worse than that only with the high-precision data, 0.087, while better than the MSE, 0.154, which only use the low-precision data.

Example 2: Fusion processing of dynamic tracking data

Assuming that GPS and BDS are tracking and measuring a dynamic target, simultaneity, and the measuring data are the single point positioning data. Suppose $(x(t), y(t), z(t))^{\mathrm{T}}$ is the position of the target orbit at time t, the positioning accuracy of GPS in every direction is 1m, and that of BDS is 3m. Simulate 80 groups of measuring data by the theoretical orbit, including $t = 0.05 \times j,\ \ j = 1, \cdots, 600$ GPS and BDS positioning data in each group. In the tracking period, the orbit data is model by the cubic spline function of the optimal node according to the reference [6]. The spline coefficient is estimated first and then the orbit $(\hat{x}^{(k)}(t_j), \hat{y}^{(k)}(t_j), \hat{z}^{(k)}(t_j), \hat{\dot{x}}^{(k)}(t_j), \hat{\dot{y}}^{(k)}(t_j), \hat{\dot{z}}^{(k)}(t_j))(\rho)$, $k = 1, 2, \cdots, 80$, $j = 1, 2, \cdots 600$ can be calculated sequentially. Assume that,

$$\mathrm{MSE}R(\rho_1^2, \rho_2^2) = \\ \frac{1}{80}\frac{1}{600}\sum_{k=1}^{80}\sum_{j=1}^{600}\sqrt{\left(x(t_j) - \hat{x}^{(k)}(t_j)\right)^2 + \left(y(t_j) - \hat{y}^{(k)}(t_j)\right)^2 + \left(z(t_j) - \hat{z}^{(k)}(t_j)\right)^2}(\rho_1^2, \rho_2^2) \tag{28}$$

where

$$(x(t_j), y(t_j), z(t_j))^{\mathrm{T}}$$

is the theoretical orbit and Table 2 gives the estimation precision for the orbit position with different weighted methods.

Table 2 Parameter Estimate Result in Different Weighted Methods

Weighted Method / Estimation Result	Only use GPS data	Only use BDS data	Traditional weighted estimation	Optimal fusion estimation
Weighted Method	(1, 0)	(0, 1)	(1/2, 1/2)	(9/10, 1/10)
MSE	1.512	3.247	1.243	0.716

5 Conclusion

Data fusion is one effective method to improve the precision of data processing. This paper researches the optimal weight and parameter estimation of unequal precision linear data fusion. For the linear fusion model, the optimal weight depends on the precision of the measuring data only, and it is consistent with the classical Gauss-Markov Theory. Furthermore, when the data precision in the fusion model is unknown, the parameter estimation and the optimal weight can be obtained by minimizing the mean square error. The parameter estimation precision of multivariate linear fusion model is given in this paper as well as some conclusions, which can be used in the practical engineering application. The accuracy of the optimal fusion estimation with multi-measuring data is better than that of only using any single measuring data and the traditional weighted estimation. The estimation precision with the traditional weighted method is better than that only with the low-precision data, and if the data precision of each measuring data satisfies some certain condition, the precision with the traditional weighted method can better than that only with the high-precision data. And these conclusions can be the support to experiment design and data selected scheme.

Note that the optimal fusion weight of the unequal precision discussed in this paper is the linear fusion model. The nonlinear model needs to be considered further, although the theory and method in this paper is the basic of nonlinear problem.

Besides, the optimal fusion weight of the multi-structure linear regression model is obtained under certain evaluation criterion. Different evaluation criterion leads to different optimal fusion weight. The evaluation criterion corresponding to the minimal MSE of parameter estimation is used in the paper, i.e. $\arg\min_{\rho} \mathrm{E}\left\|\hat{\beta}(\rho)-\beta\right\|^2$. Nevertheless, this criterion has some limitations, e.g. the contribution of each parameter to the problem is not distinguished. Certainly, other evaluation criteria should be considered for specific issues, which will be studied in the future.

References

[1] Hall D L.(1992). *Mathematical Techniques in Multi-sensor Data Fusion*, Artech House, Boston, London.

[2] Jacqueline Le Moigne and James Smith. (2000). "Image registration and fusion in remote sensing for NASA", *Proceedings of 2000 International Conference on Information Fusion,* Paris, France.

[3] Y. Bar-Shalom, H. Chen and M. Mallick. (2004). "One-step solution for the multi-step out-of-sequence-measurement problem in tracking", *IEEE Transactions on Aerospace and Electronics Systems.*

[4] X. R Li and Vesselin P. (2003), "A survey of maneuvering target tracking-part v: multiple-model methods ", *Proceeding of SPIE Conference on Signal and Data Proceeding of Small Targets, San Diego,* CA, USA.

[5] Prieto, J., Mazuelas, S., Bahillo, A., Fernandez, P., Lorenzo, R. M., and Abril, E. J. (2012), "Adaptive data fusion for wireless localization in harsh environments", *Signal Processing, IEEE Transactions.*

[6] Z. M. Wang and D. Y. Yi. (2011), *Measurement Data Modeling and Parameter Estimation*, CRC Press, China.

[7] Y. Bar- Shalom, X. R. Li and T. Kirubarajan. (2001), *Esitmationwith Applications to Tracking and Navigation*, Wiley, New York.

[8] Simon D. (2006). *Optimal State Estimation*, John Wiley & Sons, Inc. New York.

[9] Haiyin Zhou, Jiongqi Wang and Xiaogang Pan. (2013), *Fusion Theory and Methods for Satellite State Estimation*, Science Press, China.

[10] Jiongqi Wang, Haiyin Zhou and Yi Wu. (2007), “The theory of data fusion based on state optimal estimation”, *Acta Mathematicae Applicatae Sinica*, China.

[11] Cetin M., Malioutov D. M. and Willsky A. S. (2002), “A variational technique for source localization based on a sparse signal reconstruction perspective”, *Proceedings of the 2002 IEEE international conference on Acoustics, Speech, and Signal Processing,* Orlando.

[12] Tenorio L. (2001), “Statistical regularization of inverse problems”, *SIAM Review.*

[13] Tikhonov, A. N., and Arsenin, V. Y. (1977), *Solutions of ill-posed Problems*, John Wiley & Sons, New York.

[14] Voutilainen, A., Stratmann, F., and Kaipio, J. P. (2000), “A nonhomogeneous regularization method for the estimation of narrow aerosol size distributions”, *Journal of aerosol science.*

[15] Aster, R. C., Borchers, B. and Thurber, C. H. (2013), *Parameter estimation and inverse problems* , Academic Press.

Corresponding author
Xuanying Zhou can be contacted at: julia_chow07@163.com