MEMOIRS OF THE SCHOOL OF ENGINEERING OKAYAMA UNIVERSITY Vol. 7, No. 1, JUNE 1972

A Consideration on Sub-Optimal Weighting in Parameter Estimation ★

Masahiro KANEDA *

Synopsis

This paper deals with a parameter estimation method which yields the more suitable estimate of the parameter using noisy data or measured values. The estimation method is one that uses a kind of a weighted mean, and weighting at taking a weighted mean is interested in particularly. That is to say, as the grade of 'more suitable' depends upon the weighting, we can obtain the more suitable estimate by choosing the weighting coefficients suitablly. When the function which yields the estimate using finite measured values, i.e., the estimator is a particular form, sub-optimal weighting in the practical sense is discussed. Here, the concept of 'optimal' implies that the variance of the final estimate is minimum. And the particular form is one that both the denominator and the numerator of the estimator are first order formulas or second order formulas of finite measured values. And two theorems in relation to this problem are proposed and proved.

Moreover, for an exsample of application of these theorems, a parameter estimation method is dealt with, which estimates the parameters of the pulse transfer function of a control system using the sampled measured values of the impulse response of that system.

1. Introduction

It often occurs when we wish to estimate a parameter using finite noisy measured values. And we often come to the case in which the parameter is a function of these measured values and the form of the function is given or can be derived. If measured values do not involve any noises, the parameter is evaluated precisely from the function. But those measured values have noises more or less practically. Therefore, in order to obtain the more suitable estimate of the parameter, we have to measure several times and average the estimates of the parameter corresponding to these several measurements in any sense. Let this estimate corresponding to each measurement be called 'the first estimate' and let the function which yields this first estimate be called 'the first estimator'. Let the more suitable estimate of the parameter which is obtained as a weighted mean over these first estimates be called 'the final estimate' of the parameter.

Now, the grade of 'more suitable' depends upon weighting coefficients at taking a weighted mean over first estimates.

In this paper, let the concept of 'most suitable' imply that the variance of the final estimate is minimum. Then, we would like to seek weighting coefficients which yield the most suitable estimate. But, it is very difficult that we seek such a weighting precisely in general. But, we can obtain such a weighting approximately when the function which yields the first estimate, i.e., the first estimator has a particular form. Here, we deal with the case when the particular form is one that both the denominator and the numerator

* Department of Electronics

★ Received May 13, 1972.

are first order formulas and second order formulas of finite measured values. In relation to these case, we propose two theorems and prove them.

Moreover, for an example of application of these theorems, a parameter estimation method is dealt with, which estimates the parameters of the pulse transfer function of a control system using the sampled measured values of the impulse response of that system.

2. Theorem 1

Consider an estimation method in which both the denominator and the numerator of an estimator for obtaining a first estimate are given as first order formulas of measured values, as shown in (1), and the final estimate is calculated by weighting these several first estimates based on each measurement suitably and averaging them.

In this case, the sub-optimal weight on each first estimate that we can use practically in the sense of the minimum variance of the final estimate is one that is proportional to the square of the denominator of each first estimate respectively under the assumptions A and B.

$$\hat{\theta}_i = \sum_{j=1}^{m} b_j (y_{ji} + \delta_{ji}) \Big/ \sum_{j=1}^{m} a_j (x_{ji} + \varepsilon_{ji}) \qquad (1)$$

Where

$\hat{\theta}_i$: the first estimate of the i-th measurement

x_{ji}, y_{ji} : the true value of the measured value of the i-th measurement respectively

a_j, b_j : a constant which is unrelated to the i-th measurement i respectively

$\varepsilon_{ji}, \delta_{ji}$: a noise involved in the measured value of the i-th measurement respectively

Assumption A

The noises involved in measured values of the denominator and the numerator of the estimator have the same variances σ_ε^2 and σ_δ^2 ,respectively, zero means and are uncorrelated each other.

Assumption B

More than second order terms of the noises in the denominator and the numerator can be neglected compared with the sum of the square of true values of the measured values forming the denominator and the numerator respectively.

3. Proof of Theorem 1

First, let the measurement be repeated n times.

Let us have an eye to the denominator of the first estimator, and consider the weighting coefficients w_i. Where i denotes i-th measurement.

Now, let the final estimate be obtained by weighting w_i given by (2) on given by (1), and averaging them.

$$w_i \triangleq \lambda_i \cdot \sum_{j=1}^{m} a_j (x_{ji} + \varepsilon_{ji}) \qquad (2)$$

$$\hat{\theta} = \left[\sum_{i=1}^{n} \lambda_i \cdot \sum_{j=1}^{m} b_j (y_{ji} + \delta_{ji}) \right] \Big/ \left[\sum_{i=1}^{n} \lambda_i \cdot \sum_{j=1}^{m} a_j (x_{ji} + \varepsilon_{ji}) \right] \qquad (3)$$

Using the assumption B, (3) is transformed into the following equation approximately.

$$\hat{\theta} \simeq \frac{\sum_{i=1}^{n} \lambda_i \cdot \sum_{j=1}^{m} b_j \cdot y_{ji}}{\sum_{i=1}^{n} \lambda_i \cdot \sum_{j=1}^{m} a_j \cdot x_{ji}} \left\{ \left(1 + \frac{\sum_{i=1}^{n} \lambda_i \cdot \sum_{j=1}^{m} b_j \cdot \delta_{ji}}{\sum_{i=1}^{n} \lambda_i \cdot \sum_{j=1}^{m} b_j \cdot y_{ji}} \right) \cdot \left(1 - \frac{\sum_{i=1}^{n} \lambda_i \cdot \sum_{j=1}^{m} a_j \cdot \varepsilon_{ji}}{\sum_{i=1}^{n} \lambda_i \cdot \sum_{j=1}^{m} a_j x_{ji}} \right) \right\} \quad (4)$$

Then, using the assumption A and introducing some adequite letters, we can obtain the variance σ^2 of $\hat{\theta}$ approximately as follows.

$$\sigma^2 \simeq \left(K^2 \sigma_\delta^2 + \theta^2 \sigma_\varepsilon^2 \right) \cdot \left(\sum_{j=1}^{m} a_j^2 \right) \left\{ \left(\sum_{i=1}^{n} \lambda_i^2 \right) \Big/ \left(\sum_{i=1}^{n} \lambda_i \cdot \sum_{j=1}^{m} a_j x_{ji} \right)^2 \right\} \quad (5)$$

where

$$\theta \triangleq \sum_{j=1}^{m} b_j \cdot y_{ji} \Big/ \sum_{j=1}^{m} a_j \cdot x_{ji} \quad , \quad (6)$$

$$K^2 \triangleq \sum_{j=1}^{m} b_j^2 \Big/ \sum_{j=1}^{m} a_j^2 \quad . \quad (7)$$

Let us consider the term involving a variable λ_i in (5) and put it P.

$$P = \left(\sum_{i=1}^{n} \lambda_i^2 \right) \Big/ \left(\sum_{i=1}^{n} \lambda_i \cdot \sum_{j=1}^{m} a_j \cdot x_{ji} \right)^2 \quad (8)$$

When in (8), λ_i is the value satisfying (9), let P be denoted as P_k.

$$\lambda_i = K \cdot \sum_{j=1}^{m} a_j \cdot x_{ji} \quad (9)$$

where K is a constant unrelated with i-th measurement i.

$$P_K = 1 \Big/ \sum_{i=1}^{n} \left(\sum_{j=1}^{m} a_j \cdot x_{ji} \right)^2 \quad (10)$$

Now, from Schwarz's Inequality, we can obtain the following relation.

$$\sum_{i=1}^{n} \lambda_i^2 \cdot \sum_{i=1}^{n} \left(\sum_{j=1}^{m} a_j \cdot x_{ji} \right)^2 \geq \left(\sum_{i=1}^{n} \lambda_i \cdot \sum_{j=1}^{m} a_j \cdot x_{ji} \right)^2 \quad (11)$$

From (9), (10) and (11), the relation $P \geq P_k$ can be obtained. In other words, the weighting function w_i given by (12) minimizes the variance σ^2 given by (5).

$$w_i = K \cdot \left(\sum_{j=1}^{m} a_j \cdot x_{ji} \right) \cdot \left(\sum_{j=1}^{m} a_j \cdot \left(x_{ji} + \varepsilon_{ji} \right) \right) \quad (12)$$

But, as ($\sum_{j=1}^{m} a_j \cdot x_{ji}$) is a sum of true values of measured values, we can use it practically.

Therefore, we must use, in place of it, the sum of the measured values $(\sum_{j=1}^{m} a_j \cdot (x_{ji} + \varepsilon_{ji}))$.

Thus, it can be said that the practical weighting which is sub-optimal in the mean of the minimum variance is a value which is proportional to the square of the denominator of the first estimator

(Q.E.D)

4. Theorem 2

Consider an estimating method in which both the denominator and the numerator of an estimator for obtaining a first estimate are given as second order formulas of measured values, as shown in (13), and the final estimate is calculated by weighting these several first estimates based on each measurement suitably and averaging them.

In this case, the sub-optimal weight that we can use practically in the sense of the minimum variance of the final estimate is one that is proportional to the magnitude of the denominator of the first estimator under the assumptions A and B.

$$\hat{\theta}_i=\left\{\sum_{j,k}^{m} b_{jk}(y_{ji}+\delta_{ji})(y_{ki}+\delta_{ki})\right\}\Big/\left\{\sum_{j,k}^{m} a_{jk}(x_{ji}+\varepsilon_{ji})(x_{ki}+\varepsilon_{ki})\right\} \quad (13)$$

Where

$\hat{\theta}_i$: the first estimate of the i-th measurement

x_{ji}, y_{ji}: the true value of the measured value of the i-th measurement respectively

a_{jk}, b_{jk}: a constant which is unrelated to the i-th measurement i respectively

$\varepsilon_{ji}, \delta_{ji}$: a noise involved in the measured value of the i-th measurement respectively

5. Proof of Theorem 2

This is almost similar to the proof of Theorem 1.

Let the measurement be repeated n times. Let us have an eye to the denominator of the first estimator, and consider the weighting coefficients w_i as shown in (14). Where i denotes i-th measurement.

Then, let the final estimate be obtained by weighting w_i given by (14) on given by (13), and averaging them.

$$w_i = \lambda_i \cdot \sum_{j,k}^{m} a_{jk}\ (x_{ji}+\varepsilon_{ji})(x_{ki}+\varepsilon_{ki}) \quad (14)$$

$$\hat{\theta} = \left\{\sum_{i=1}^{n}\lambda_i\sum_{j,k}^{m} b_{jk}(y_{ji}+\delta_{ji})(y_{ki}+\delta_{ki})\right\}\Big/\left\{\sum_{i=1}^{n}\lambda_i\sum_{j,k}^{m} a_{jk}(x_{ji}+\varepsilon_{ji})(x_{ki}+\varepsilon_{ki})\right\} \quad (15)$$

Using the assumption B, (15) is transformed into the following equation approximately.

$$\hat{\theta} \simeq \frac{\sum_{i=1}^{n}\lambda_i\sum_{j,k}^{m} b_{jk}y_{ji}\cdot y_{ki}}{\sum_{i=1}^{n}\lambda_i\sum_{j,k}^{m} a_{jk}x_{ji}\cdot x_{ki}}\left\{\left(1-\frac{\sum_{i=1}^{n}\lambda_i\sum_{j,k}^{m} b_{jk}(y_{ji}\delta_{ki}+y_{ki}\delta_{ji})}{\sum_{i=1}^{n}\lambda_i\sum_{j,k}^{m} b_{jk}\,y_{ji}\cdot y_{ki}}\right)\left(1-\frac{\sum_{i=1}^{n}\lambda_i\sum_{j,k}^{m} a_{jk}(x_{ji}\varepsilon_{ki}+x_{ki}\varepsilon_{ji})}{\sum_{i=1}^{n}\lambda_i\sum_{j,k}^{m} a_{jk}\cdot x_{ji}\cdot x_{ki}}\right)\right\} \quad (16)$$

Then, using the assumption A and introducing some adequite letters, we can obtain the variance σ^2 of $\hat{\theta}$ approximately as follows.

$$\sigma^2 \cong \left(K^2\cdot\sigma_\delta^2+\theta^2\cdot\sigma_\varepsilon^2\right)\left(\sum_{i=1}^{n}\lambda_i^2\cdot B_i^2\right)\Big/\left(\sum_{i=1}^{n}\lambda_i\cdot A_i^2\right)^2 \quad (17)$$

where

$$\theta \triangleq \sum_{j,k}^{m} b_{jk}\cdot y_{ji}\cdot y_{ki}\Big/\sum_{j,k}^{m} a_{jk}\cdot x_{ji}\cdot x_{ki} \quad (18)$$

$$K^2 \triangleq \sum_{j,k}^{m} b_{jk}^2(y_{ji}^2+y_{ki}^2)\Big/\sum_{j,k}^{m} a_{jk}^2(x_{ji}^2+x_{ki}^2) \quad (19)$$

$$A_i^2 \triangleq \sum_{j,k}^{m} a_{jk}\cdot x_{ji}\cdot x_{ki} \quad (20)$$

$$B_i^2 \triangleq \sum_{j,k}^{m} a_{jk}^2(x_{ji}^2+x_{ki}^2) \quad (21)$$

Let us consider the term involving a variable λ_i in (17) and put it Q.

$$Q = \left(\sum_{i=1}^{n} \lambda_i^2 \cdot B_i^2\right) \Big/ \left(\sum_{i=1}^{n} \lambda_i \cdot A_i^2\right)^2 \tag{22}$$

When in (22) λ_i is a constant K , let Q be denoted as Q_K.

$$Q_K = \left(\sum_{i=1}^{n} B_i^2\right) \Big/ \left(\sum_{i=1}^{n} A_i^2\right)^2 \tag{23}$$

Moreover, by introducing a letter H given by (24), (22) and (23) are written as follows.

$$H \triangleq B_i^2 / A_i^2 \quad , \tag{24}$$

where H is a constant which is unrelated to i.

$$Q = \left(\sum_{i=1}^{n} \lambda_i^2 \cdot B_i^2\right) \Big/ \left(H^2 \cdot \sum_{i=1}^{n} \lambda_i \cdot B_i^2\right) \tag{25}$$

$$Q_K = 1 \Big/ \left(H^2 \cdot \sum_{i=1}^{n} B_i^2\right) \tag{26}$$

Now, from Schwarz's Inequality, we can obtain the following relation.

$$\left(\sum_{i=1}^{n} \lambda_i^2 \cdot B_i^2\right) \cdot \left(\sum_{i=1}^{n} B_i^2\right) \geqq \left(\sum_{i=1}^{n} \lambda_i \cdot B_i^2\right)^2 \tag{27}$$

From (25), (26) and (27), the relation $Q \geqq Q_K$ can be obtained.

That is to say, in order to minimize the variance σ^2 of $\hat{\theta}$, we must make λ_i a constant. In other words, it can be said that the practical weighting which is sub-optimal in the sense of the minimum variance is a value which is proportional to the magnitude of the denominator of the first estimator $\hat{\theta}_i$.

(Q.E.D)

6. Application of Theorem

Next, consider an application of these theorems.

Let us consider the application to the estimation method which has been proposed.[1] This method is one with which we estimate the parameters α_i of the pulse transfer function of a control system which is given as a second order equation, as shown in (28), approximately using finite noisy measured values of the impulse response of that system. The counterparts of the first estimates for α_i are given by equations (29) and (30).[1),2)]

$$G_\alpha(z) = \alpha_3 \cdot z^{-1} \Big/ \left(1 + \alpha_1 \cdot z^{-1} + \alpha_2 \cdot z^{-2}\right) \tag{28}$$

$$\hat{\alpha}_1^{(k)} = \left(h_1 \cdot h_{2k} - 2 \cdot h_k \cdot h_{k+1}\right) \Big/ h_k^2 \tag{29}$$

$$\hat{\alpha}_2^{(k)} = \left(-h_1 \cdot h_{2k} + h_k \cdot h_{k+1}\right) \Big/ \left(h_k \cdot h_{k-1}\right) \tag{30}$$

where k=2,3,$\cdots$,[n/2] , and n denotes the number of the total sampled values of the impulse response. And h_k denotes the measured value of the sampled value of the impulse response at time kT and consists of the true value g_k and the noise .

From (29) and (30), as it is obvious that the noises involved in the denominator and the numerator are not uncorrelated each other, we can not apply these theorems to this problem in the rigorous sense.

But, by applying Theorem 2 to this estimation method formally, and averaging over the first estimates which are weighted h_k^2 on $\hat{\alpha}_1^{(k)}$ and $|h_k \cdot h_{k-1}|$ on $\hat{\alpha}_2^{(k)}$, we can obtain the final estimates of $\hat{\alpha}_1$ and $\hat{\alpha}_2$, as shown in (31) and (32).

This estimation method is INM[2] (Improved New Method) which has been proposed.

$$\hat{\alpha}_1 = \sum_{k=2}^{[N/2]} (h_1 \cdot h_{2k} - 2h_k \cdot h_{k+1}) \Big/ \sum_{k=2}^{[N/2]} h_k^2 \tag{31}$$

$$\hat{\alpha}_2 = \sum_{k=2}^{[N/2]} (-h_1 \cdot h_{2k} + h_k \cdot h_{k+1}) \cdot \mathrm{sgn}(h_k \cdot h_{k-1}) \Big/ \sum_{k=1}^{[N/2]} |h_k \cdot h_{k-1}| \tag{32}$$

where $|\cdot|$ denotes the absolute value of $\cdot$ and [ ] denotes Gaussian sign.

Now, it is important that in any measured value treated in Theorem 1 and 2, its true value is not necessary larger than its noise in the point of the absolute value, but the former has only to be larger than the latter in the point of the power or the energy-like.

Practically, it has been sured with some computer simulations[2] that INM given by (31) and (32) seems to be enough good method even when some of the true values of measured values are much smaller than the corresponding noises in the point of the absolute value.

7. Conclusion

When an estimator is a particular form of which both the denominator and the numerator are first order formulas or second order formulas, sub-optimal weighting has been obtained in the practical sense. Where, the concept of 'optimal' implies that the variance of the final estimate is minimum.

Whereas, as these theorems are concerned with the case when the estimator is a particular form as above discussed, they may seem to be impractical. However, in practice, these forms are pretty probable.

Moreover, it seems that these theorems are practically useful, as they are available even when the true values of some measured values are much smaller than the noises of the corresponding measured values in the point of their absolute values respectively.

Acknowledgement

The author wishes to express his thanks to Dr. Kazuaki Ando (Assistant Professor of Kyoto University) for his valuable discussion. Especially, the author is indebted to him for some pertinent suggestions in the proofs of the theorems.

Reference

(1) Kaneda, M. and Ando, K. : J.Institute of Electronics and Communication Eng. Japan, 54-C (1971), 1168

(2) Kaneda, M. and Ando, K. : 3-th Simposium on Stochastic Control Theory, Japan Assoc.n of Automatic Control Engineers, Oct. 1971, 37-40