IEMS Vol. 1, No. 1, pp. 64-72, December 2002.

On a Multiple Data Handling Method under Online Parameter Estimation

Kazuhiro Takeyasu†
Department of Industrial Engineering
The Japan Research Institute, Ltd., 16 Ichiban-cho, Chiyoda-ku, Tokyo, 102-0082, JAPAN
Tel: +81-3-3288-4171, Fax: +81-3-3288-4339, E-mail: takeyasu.kazuhiro@jri.co.jp

Takashi Amemiya
Department of Industrial Engineering
Tokyo Metropolitan Institute of Technology, 6-6 Asahigaoka, Hino, Tokyo, 191-0065, JAPAN
Tel & Fax: +81-42-585-8666, E-mail: tamemiya@tmit.ac.jp

Katsuhiro Iino
Department of Industrial Engineering
Tokyo Metropolitan Institute of Technology, 6-6 Asahigaoka, Hino, Tokyo, 191-0065, JAPAN
Tel & Fax: +81-42-585-8666, E-mail: katsu@krmgams6.tmit.ac.jp

Shiro Masuda
Department of Industrial Engineering
Tokyo Metropolitan Institute of Technology, 6-6 Asahigaoka, Hino, Tokyo, 191-0065, JAPAN
Tel & Fax: +81-42-585-8666, E-mail: smasuda@cc.tmit.ac.jp

Abstract. In the field of plant maintenance, data that are gathered by sensors on multiple machines are handled and analyzed. Online or pseudo online data handling is required on such fields. When the data occurrence speed exceeds the data handling speed, multiple data should be handled at a time (batch data handling or pseudo online data handling). If l amount of data are received at one time following N amount of data, how to estimate the new parameters effectively is a great concern. A new simplified calculation method, which calculates the N data's weights, is introduced. Numerical examples show that this new method has a fairly good estimation accuracy and the calculation time is less than 1/10 compared with the case when the whole data are re-calculated. Even under the restriction calculation ability in the apparatus is limited, this proposed method makes the failure detection of equipments possible in early stages with a few new coming data. This method would be applicable in many data handling fields.

Keywords: time series analysis, AR model, Yule-Walker equation, autocorrelation function

1. INTRODUCTION

In the field of plant maintenance, the dominant maintenance method is shifting from Time Based Maintenance (TBM) to Condition Based Maintenance (CBM). In CBM, machines are watched continuously and data are gathered by sensors and analyzed. Online or pseudo online data handling is required on such field. Calculating Root mean square of data, Kurtosis, Bicoherence or distance of system parameters, we use them as the method for diagnosis of machines.

In this paper, we discuss the failure detection by inspecting the changes of the parameter distance. For CBM, these detections should be done continuously. Here

† : Corresponding Author

the detection is done by calculating the change of system parameters. It is well known how to estimate the system parameters sequentially when one new datum is added to N amount of data already obtained (Tokumaru *et al.*, 1982; Sagara *et al.*, 1994; Katayama, 1994). However, in the case that the data occurrence speed exceeds the data handling speed, multiple data should be handled at a time (batch data handling or pseudo online data handling). For example, we have to estimate the system parameters in such case that N say 5000, data are already given and on the top of this, l say 100, data are newly added.

In this paper, we discuss these matters and propose a new method to calculate the parameters effectively. So far, Nakamura *et al.* (1984) proposed a method to shorten the calculation time. They calculated parameters of whitening filters in the generalized least square method and utilized the memory for the autocorrelation function and the cross correlation function of input and output data. Yamagata *et al.* (2001) proposed a method for real time diagnosis. They estimated the volume of bias of measurement data utilizing dynamic responses of plant and then estimated the condition, which is not affected by the breakdown of measurement apparatus. Although some related papers have been published, as above mentioned, the theme in this paper is concerned with multiple data handling newly got and the proposed method is a unique one.

We introduce the objective model in section 2. A new method is proposed in section 3. Other approaches are referred in section 4 and 5. In section 6, numerical examples are exhibited.

2. FACTORS FOR VIBRATION CALCULATION

In the analysis of time series data, Autoregressive (AR) model or Autoregressive Moving Average (ARMA) model are frequently adopted. In this paper we adopt AR model, because it has a good estimation property (unbiased estimation) and is easy to identify.

Consider the p-th order AR model expressed as

$$x_n + \sum_{i=1}^{p} a_i x_{n-i} = e_n \tag{1}$$

Here,

$\{x_n\}$: Sample process of a stationary ergodic gaussian process $x(t)$ $(n=1, 2, 3, \cdots, N, \cdots)$

$\{e_n\}$: Gaussian noise of mean 0, variance σ_e^2

Assume that (1) satisfies the stationarity condition.

Let

$$\begin{aligned} \mathbf{Z}_n &= [-x_{n-1}, \cdots, -x_{n-p}]^T \\ \boldsymbol{\theta} &= [a_1, a_2, \cdots, a_p]^T \end{aligned}$$

then, x_n can be expressed as

$$x_n = \boldsymbol{\theta}^T \mathbf{Z}_n + e_n \tag{2}$$

If we set the criteria function as

$$G_N = \sum_{n=1}^{N} \left[x_n - \boldsymbol{\theta}^T \mathbf{Z}_n \right]^2, \tag{3}$$

Then the estimation of θ denoted by $\hat{\theta}_N$, which gives the minimum least square of G_N, is given by

$$\hat{\boldsymbol{\theta}}_N = \left[\sum_{n=1}^{N} \mathbf{Z}_n \mathbf{Z}_n^T \right]^{-1} \sum_{n=1}^{N} \mathbf{Z}_n x_n \tag{4}$$

If l amount of data are newly added to N amount of data (typical case is that this sequential data are added $x_{N+1}, x_{N+2}, \cdots, x_{N+l}$ to $x_1, x_2, \cdots, x_N$), we can detect the irregular condition by calculating

$$J = \|\hat{\boldsymbol{\theta}}_{N+l} - \hat{\boldsymbol{\theta}}_N\|^2 \tag{5}$$

or J in which $\hat{\theta}_N$ is replaced by $\hat{\theta}_{N_0}$ which is the estimated parameter vector with N_0 data obtained under normal condition.

$$J = \|\hat{\boldsymbol{\theta}}_{N+l} - \hat{\boldsymbol{\theta}}_{N_0}\|^2 \tag{6}$$

As an irregular condition appeares in the machine operation, system parameters change usually, and the distance of parameters grows big. We may consider that a failure condition arise if the score of J exceeds a certain value J_0. J_0 may be determined by machines. This kind of method is generally used in failure detection. For example, in the case of using the root mean square (RMS) of the original time series, the distance of RMS from normal condition is calculated. When the distance value exceeds certain value, the corresponding condition is judged as a failure. System parameters distance method is sophisticated than that of RMS because the first is equal to the system identification process while the latter is only raw data handling.

In this paper, we examine three cases in estimating θ_N.

1. The case of using an autocorrelation function

2. The case of using the canonical equation

$$\left[\sum_{n=1}^{N} \mathbf{Z}_n \mathbf{Z}_n^T \right] \hat{\boldsymbol{\theta}}_n = \sum_{n=1}^{N} \mathbf{Z}_n x_n \tag{7}$$

3. The case of repeating recursive estimation

The third one is only for a reference, because we try to deal newly got l data, not one datum, which is the situation of case 3.

3. THE CASE OF USING AUTOCORRELATION FUNCTION

The autocorrelation function of $\{x_n\}$ is stated as

$$R_j = E[x_n x_{n+j}] \qquad (8)$$
$$R_{-j} = R_j$$

by definition.

For AR process, the following Yule-Walker equation holds true.

$$\left.\begin{aligned} R_1 + a_1 R_0 + a_2 R_1 + \cdots + a_p R_{p-1} = 0 \\ R_2 + a_1 R_1 + a_2 R_0 + \cdots + a_p R_{p-2} = 0 \\ \vdots \qquad\qquad\qquad \\ R_p + a_1 R_{p-1} + a_2 R_{p-2} + \cdots + a_p R_0 = 0 \end{aligned}\right\} \qquad (9)$$

Now, assume that we get N amount data $\{x_n : n = 1, 2, \cdots, N\}$. In this case $\hat{R}_{N\cdot j}$, the estimation of R_j, is stated as

$$\hat{R}_{N\cdot j} = \frac{1}{N-j} \sum_{i=1}^{N-j} x_i x_{i+j} \qquad (10)$$

Assume that we receive l data $\{x_n : n = N+1, N+2, \cdots, N+l\}$ on the top of this, then

$$\begin{aligned} \hat{R}_{N+l\cdot j} &= \frac{1}{N+l-j} \sum_{i=1}^{N+l-j} x_i x_{i+j} \\ &= \frac{N-j}{N+l-j} \hat{R}_{N\cdot j} + \frac{l}{N+l-j} \hat{R}_{N/l\cdot j} \end{aligned} \qquad (11)$$

where $\hat{R}_{N/l\cdot j}$ denote the estimated autocorrelation function by using only l amount of data from $N+1$ to $N+l$. (In detail, see Appendix 1)

We may safely assume that

$$N \gg l \gg p$$

because the order p of AR model is usually within the range of several order to scores of order. Since j satisfies

$$j \leq p-1$$

(11) can be expressed as

$$\hat{R}_{N+l\cdot j} \simeq \frac{N}{N+l} \hat{R}_{N\cdot j} + \frac{l}{N+l} \hat{R}_{N/l\cdot j} \qquad (12)$$

Now, let α and β be defined as

$$\alpha = \frac{N}{N+l}, \quad \beta = \frac{l}{N+l} \qquad (13)$$

Replacing the autocorrelation functions of (9) by the estimation, we obtain

$$\hat{\theta}_N = -\begin{bmatrix} \hat{R}_{N\,0} & \hat{R}_{N\,1} & \cdots & \hat{R}_{N\cdot p-1} \\ \hat{R}_{N\cdot 1} & \hat{R}_{N\,0} & \cdots & \hat{R}_{N\,p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{R}_{N\,p-1} & \hat{R}_{N\,p-2} & \cdots & \hat{R}_{N\,0} \end{bmatrix}^{-1} \begin{bmatrix} \hat{R}_{N\,1} \\ \hat{R}_{N\,2} \\ \vdots \\ \hat{R}_{N\,p} \end{bmatrix} \qquad (14)$$

($\hat{\theta}_N$ is used as the estimation of θ. In this paper, different methods of estimation are introduced such as section 4 and 5, so contents are different by method.)

If l amount of data are added, then

$$\hat{\theta}_{N+l} = -\begin{bmatrix} \hat{R}_{L\,0} & \hat{R}_{L\,1} & \cdots & \hat{R}_{L\,p-1} \\ \hat{R}_{L\,1} & \hat{R}_{L\,0} & \cdots & \hat{R}_{L\,p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{R}_{L\,p-1} & \hat{R}_{L\cdot p-2} & \cdots & \hat{R}_{L\,0} \end{bmatrix}^{-1} \begin{bmatrix} \hat{R}_{L\,1} \\ \hat{R}_{L\,2} \\ \vdots \\ \hat{R}_{L\,p} \end{bmatrix}$$

$$= -\begin{bmatrix} \alpha\hat{R}_{N\,0} + \beta\hat{R}_{N/l\,0} & \cdots & \alpha\hat{R}_{N\,p-1} + \beta\hat{R}_{N/l\,p-1} \\ \alpha\hat{R}_{N\,1} + \beta\hat{R}_{N/l\,1} & \cdots & \alpha\hat{R}_{N\,p-2} + \beta\hat{R}_{N/l\,p-2} \\ \vdots & & \vdots \\ \alpha\hat{R}_{N\,p-1} + \beta\hat{R}_{N/l\,p-1} & \cdots & \alpha\hat{R}_{N\,0} + \beta\hat{R}_{N/l\,0} \end{bmatrix}^{-1} \cdot \begin{bmatrix} \alpha\hat{R}_{N\cdot 1} + \beta\hat{R}_{N/l\,1} \\ \alpha\hat{R}_{N\,2} + \beta\hat{R}_{N/l\cdot 2} \\ \vdots \\ \alpha\hat{R}_{N\,p} + \beta\hat{R}_{N/l\,p} \end{bmatrix} \qquad (15)$$

Here, $L = N + l$

Let

$$\mathbf{A} = \begin{bmatrix} \hat{R}_{N\,0} & \hat{R}_{N\,1} & \cdots & \hat{R}_{N\,p-1} \\ \hat{R}_{N\,1} & \hat{R}_{N\,0} & \cdots & \hat{R}_{N\,p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{R}_{N\,p-1} & \hat{R}_{N\cdot p-2} & \cdots & \hat{R}_{N\,0} \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} \hat{R}_{N/l\cdot 0} & \hat{R}_{N/l\,1} & \cdots & \hat{R}_{N/l\,p-1} \\ \hat{R}_{N/l\,1} & \hat{R}_{N/l\,0} & \cdots & \hat{R}_{N/l\,p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{R}_{N/l\,p-1} & \hat{R}_{N/l\cdot p-2} & \cdots & \hat{R}_{N/l\,0} \end{bmatrix}$$

$$\mathbf{C} = \begin{bmatrix} \hat{R}_{N\,1} \\ \hat{R}_{N\cdot 2} \\ \vdots \\ \hat{R}_{N\cdot p} \end{bmatrix}$$

$$\mathbf{D} = \begin{bmatrix} \hat{R}_{N/l\cdot 1} \\ \hat{R}_{N/l\cdot 2} \\ \vdots \\ \hat{R}_{N/l\cdot p} \end{bmatrix}$$

Using the following formula,

$$[\mathbf{Q}+\mathbf{R}]^{-1} = \mathbf{Q}^{-1}-\mathbf{Q}^{-1}\mathbf{R}[\mathbf{I}+\mathbf{Q}^{-1}\mathbf{R}]^{-1}\mathbf{Q}^{-1}$$
$$[\mathbf{I}+\mathbf{Q}]^{-1} = \mathbf{I}-[\mathbf{I}+\mathbf{Q}]^{-1}\mathbf{Q}$$

we get

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{N+l} &= \hat{\boldsymbol{\theta}}_N - \frac{l}{N}\mathbf{A}^{-1}\mathbf{D} - \frac{l}{N}\mathbf{A}^{-1}\mathbf{B}\hat{\boldsymbol{\theta}}_N \\ &+\left(\frac{l}{N}\right)^2(\mathbf{A}^{-1}\mathbf{B})^2\hat{\boldsymbol{\theta}}_N + \left(\frac{l}{N}\right)^2\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{D} \\ &-\left(\frac{l}{N}\right)^3\mathbf{A}^{-1}\mathbf{B}\left[\mathbf{I}+\frac{l}{N}\mathbf{A}^{-1}\mathbf{B}\right]^{-1}(\mathbf{A}^{-1}\mathbf{B})^2\hat{\boldsymbol{\theta}}_N \\ &-\left(\frac{l}{N}\right)^3(\mathbf{A}^{-1}\mathbf{B})^2\mathbf{A}^{-1}\mathbf{D} \\ &+\left(\frac{l}{N}\right)^4\mathbf{A}^{-1}\mathbf{B}\left[\mathbf{I}+\frac{l}{N}\mathbf{A}^{-1}\mathbf{B}\right]^{-1}(\mathbf{A}^{-1}\mathbf{B})^2\mathbf{A}^{-1}\mathbf{D} \quad (16)\end{aligned}$$

(In detail, see Appendix 2)

Since $N \gg l$, $\hat{\theta}_N$ can be approximated as

$$\hat{\boldsymbol{\theta}}_{N+l} \simeq \hat{\boldsymbol{\theta}}_N - \frac{l}{N}\mathbf{A}^{-1}\mathbf{D} - \frac{l}{N}\mathbf{A}^{-1}\mathbf{B}\hat{\boldsymbol{\theta}}_N \quad (17)$$

or

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{N+l} &\simeq \hat{\boldsymbol{\theta}}_N - \frac{l}{N}\mathbf{A}^{-1}\mathbf{D} - \frac{l}{N}\mathbf{A}^{-1}\mathbf{B}\hat{\boldsymbol{\theta}}_N \\ &+\left(\frac{l}{N}\right)^2(\mathbf{A}^{-1}\mathbf{B})^2\hat{\boldsymbol{\theta}}_N + \left(\frac{l}{N}\right)^2\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{D} \quad (18)\end{aligned}$$

If we adopt (17), we can calculate simply by partially utilizing matrices in $\hat{\theta}_N$'s calculation.

Here, we examine the meaning of (15)~(18). Let

$$\bar{\mathbf{A}} = \alpha\mathbf{A}$$
$$\bar{\mathbf{B}} = \beta\mathbf{B}$$
$$\bar{\mathbf{C}} = \alpha\mathbf{C}$$
$$\bar{\mathbf{D}} = \beta\mathbf{D}$$

we can get

$$\hat{\theta}_{N+l} = -\left[\mathbf{I}+\bar{\mathbf{A}}^{-1}\bar{\mathbf{B}}\right]^{-1}\left[\bar{\mathbf{A}}^{-1}\bar{\mathbf{C}}+\left(\bar{\mathbf{A}}^{-1}\bar{\mathbf{B}}\right)\left(\bar{\mathbf{B}}^{-1}\bar{\mathbf{D}}\right)\right] \quad (19)$$

(In detail, see Appendix 3)

$\bar{A}$, $\bar{C}$ are calculated by using $\{1, \cdots, N\}$ data, and $\bar{B}$, $\bar{D}$ are calculated by using $\{N+1, \cdots, N+1\}$ data. $\hat{\theta}_N$ can be obtained as

$$\hat{\boldsymbol{\theta}}_N = -\bar{\mathbf{A}}^{-1}\bar{\mathbf{C}} \quad (20)$$

from $\{1, \cdots, N\}$ data, and the corresponding value $\hat{\theta}_{\frac{N+l}{N}}$ derived from $\{N+1, \cdots, N+l\}$ data is given as

$$\hat{\boldsymbol{\theta}}_{\frac{N+l}{N}} = -\bar{\mathbf{B}}^{-1}\bar{\mathbf{D}} \quad (21)$$

In (19), $\bar{A}^{-1}\bar{B}$ can be considered to be a weight calculated from $\{1, \cdots, N\}$ data, attaching to $\hat{\theta}_{\frac{N+l}{N}}$.

Let

$$\bar{\mathbf{A}}^{-1}\bar{\mathbf{B}} = \mathbf{F} \quad (22)$$

then, since all elements of $\mathbf{F}$ are of order $O(l/N)$, we can get the following (23), (24).

$$\hat{\boldsymbol{\theta}}_{N+l} = [\mathbf{I}+\mathbf{F}]^{-1}\left[\hat{\boldsymbol{\theta}}_N + \mathbf{F}\hat{\boldsymbol{\theta}}_{\frac{N+l}{N}}\right] \quad (23)$$

$$= \left[\mathbf{I}+\sum_{k=1}^{\infty}(-\mathbf{F})^k\right]\left[\hat{\boldsymbol{\theta}}_N + \mathbf{F}\hat{\boldsymbol{\theta}}_{\frac{N+l}{N}}\right] \quad (24)$$

The 1st order approximation of $\hat{\theta}_{N+l}$ is

$$\hat{\boldsymbol{\theta}}_{N+l} = [\mathbf{I}-\mathbf{F}][\hat{\boldsymbol{\theta}}_N + \mathbf{F}\hat{\boldsymbol{\theta}}_{\frac{N+l}{N}}] \quad (25)$$

Neglecting $O(F^2)$, we get

$$\hat{\boldsymbol{\theta}}_{N+l} \simeq \hat{\boldsymbol{\theta}}_N - \mathbf{F}\hat{\boldsymbol{\theta}}_N + \mathbf{F}\hat{\boldsymbol{\theta}}_{\frac{N+l}{N}} \quad (26)$$

which coincide with (17).

The 2nd order approximation is

$$\hat{\boldsymbol{\theta}}_{N+l} = [\mathbf{I}-\mathbf{F}+\mathbf{F}^2][\hat{\boldsymbol{\theta}}_N + \mathbf{F}\hat{\boldsymbol{\theta}}_{\frac{N+l}{N}}] \quad (27)$$

Neglecting $O(F^3)$, we get

$$\hat{\boldsymbol{\theta}}_{N+l} \simeq \hat{\boldsymbol{\theta}}_N - \mathbf{F}\hat{\boldsymbol{\theta}}_N + \mathbf{F}^2\hat{\boldsymbol{\theta}}_N + \mathbf{F}\hat{\boldsymbol{\theta}}_{\frac{N+l}{N}} - \mathbf{F}^2\hat{\boldsymbol{\theta}}_{\frac{N+l}{N}} \quad (28)$$

Which coincide with (18).

In utilizing this method to the defect detection, it is natural to take the following steps.

First get N amount of data under regular condition and calculate A^{-1}, $\hat{\theta}_N$, which appear in (17), and then calculate $\hat{\theta}_{N+l}$ receiving new l amount of data. Then the failure of machines can be detected by (6).

Newly added l amount of data may be those following N amount of data. However, using this method for failure detection, it is often the case that the first N data are those for regular condition and the newly obtained l amount data represent the current situation, that mean time at which l amount of data are received can be separated from the time N data are taken.

4. THE CASE OF USING CANONICAL EQUATION

Using the canonical equation of (7), we can get

$$\hat{\theta}_{N+l} = \left[\sum_{n=1}^{N+l} \mathbf{Z}_n\mathbf{Z}_n^T\right]^{-1} \sum_{n=1}^{N+l} \mathbf{Z}_n x_n$$

$$= \left[\sum_{n=1}^{N} \mathbf{Z}_n\mathbf{Z}_n^T + \sum_{n=N+1}^{N+l} \mathbf{Z}_n\mathbf{Z}_n^T\right]^{-1} \cdot\left(\sum_{n=1}^{N} \mathbf{Z}_n x_n + \sum_{n=N+1}^{N+l} \mathbf{Z}_n x_n\right) \quad (29)$$

Let

$$\mathbf{K} = \sum_{n=1}^{N} \mathbf{Z}_n\mathbf{Z}_n^T$$
$$\mathbf{L} = \sum_{n=N+1}^{N+l} \mathbf{Z}_n\mathbf{Z}_n^T$$
$$\mathbf{M} = \sum_{n=1}^{N} \mathbf{Z}_n x_n$$
$$\mathbf{N} = \sum_{n=N+1}^{N+l} \mathbf{Z}_n x_n$$

then

$$\hat{\theta}_{N+l} = [\mathbf{K}+\mathbf{L}]^{-1}[\mathbf{M}+\mathbf{N}] \quad (30)$$

Although signs may differ because of the difference of definition, this is the same form with (19).

Adopting (12) and (13), (30) can be expressed as

$$\begin{aligned}\hat{\theta}_{N+l} &= -[N\mathbf{A}+l\mathbf{B}]^{-1}[N\mathbf{C}+l\mathbf{D}] \\ &= -[\bar{\mathbf{A}}+\bar{\mathbf{B}}]^{-1}[\bar{\mathbf{C}}+\bar{\mathbf{D}}]\end{aligned} \quad (31)$$

and it is quite the same with (19). Therefore, the 1st order approximation and the 2nd order approximation lead to the same kind of equations of (25) and (27) respectively.

5. THE CASE OF REPEATING RECURSIVE ESTIMATION

Assume we are given N amount of data and we get newly (N+1)-th data, then

$$\hat{\theta}_{N+1} = \left[\sum_{n=1}^{N+1} \mathbf{Z}_n\mathbf{Z}_n^T\right]^{-1} \sum_{n=1}^{N+1} \mathbf{Z}_n x_n$$

$$= \left[\sum_{n=1}^{N} \mathbf{Z}_n\mathbf{Z}_n^T + \mathbf{Z}_{N+1}\mathbf{Z}_{N+1}^T\right]^{-1} \cdot\left(\sum_{n=1}^{N} \mathbf{Z}_n x_n + \mathbf{Z}_{N+1} x_{N+1}\right) \quad (32)$$

Let

$$\sum_{n=1}^{N} \mathbf{Z}_n\mathbf{Z}_n^T = \mathbf{A}_N$$

Using the formula

$$(\mathbf{Q}+\mathbf{R}\mathbf{R}')^{-1} = \mathbf{Q}^{-1} - \frac{\mathbf{Q}^{-1}\mathbf{R}\mathbf{R}'\mathbf{Q}^{-1}}{1+\mathbf{R}'\mathbf{Q}^{-1}\mathbf{R}}$$
$$(\mathbf{Q}: n\times n \ , \ \mathbf{R}: n\times 1 \ , \ \mathbf{R}': 1\times n)$$

we get

$$\mathbf{A}_{N+1}^{-1} = \mathbf{A}_N^{-1} - \frac{\mathbf{A}_N^{-1}\mathbf{Z}_{N+1}\mathbf{Z}_{N+1}^T\mathbf{A}_N^{-1}}{1+\mathbf{Z}_{N+1}^T\mathbf{A}_N^{-1}\mathbf{Z}_{N+1}} \quad (33)$$

$$\hat{\theta}_{N+1} = \left(\mathbf{I} - \frac{\mathbf{A}_N^{-1}\mathbf{Z}_{N+1}\mathbf{Z}_{N+1}^T}{1+\mathbf{Z}_{N+1}^T\mathbf{A}_N^{-1}\mathbf{Z}_{N+1}}\right)\hat{\theta}_N + \frac{\mathbf{A}_N^{-1}\mathbf{Z}_{N+1}x_{N+1}}{1+\mathbf{Z}_{N+1}^T\mathbf{A}_N^{-1}\mathbf{Z}_{N+1}} \quad (34)$$

It can be re-stated as

$$\hat{\theta}_{N+1} = (\mathbf{I}-\mathbf{K}_{N+1}\mathbf{Z}_{N+1}^T)\hat{\theta}_N + \mathbf{K}_{N+1}x_{N+1} \quad (35)$$

$$\mathbf{K}_{N+1} = \mathbf{P}_N\mathbf{Z}_{N+1}(1+\mathbf{Z}_{N+1}^T\mathbf{P}_N\mathbf{Z}_{N+1})^{-1} \quad (36)$$

$$\mathbf{P}_{N+1} = (\mathbf{I}-\mathbf{K}_{N+1}\mathbf{Z}_{N+1}^T)\mathbf{P}_N \quad (37)$$

Where $A^{-1}_N = P_N$.

$\hat{\theta}_{N+l}$ is stated as

$$\hat{\theta}_{N+l} = \prod_{i=1}^{l}(\mathbf{I}-\mathbf{K}_{N+i}\mathbf{Z}_{N+i}^T)\hat{\theta}_N + \sum_{j=1}^{l-1}\left\{\prod_{i=1}^{l-j}(\mathbf{I}-\mathbf{K}_{N+l+1-i}\mathbf{Z}_{N+l+1-i}^T)\right\}\cdot\mathbf{K}_{N+j}x_{N+j} + \mathbf{K}_{N+l}x_{N+l} \quad (38)$$

(In detail, see Appendix 4)

Although this mathematical expression is simple, the volume of calculation does not differ from l times of the calculation for (35)~(37), because it is only a recursive calculation itself. Introducing our method reduces the amount of the calculations substantially.

If N is sufficiently large, using the relation shown in section 3,

$$1+\mathbf{Z}_{N+1}^T\mathbf{A}_N^{-1}\mathbf{Z}_{N+1} = 1+O\left(\frac{1}{N}\right) \rightarrow 1 \ (N\rightarrow\infty) \quad (39)$$

Therefore, $\hat{\theta}_{N+l}$ can be approximated as

$$\hat{\theta}_{N+1} \simeq (\mathbf{I} - \mathbf{A}_N^{-1}\mathbf{Z}_{N+1}\mathbf{Z}_{N+1}^T)\hat{\theta}_N + \mathbf{A}_N^{-1}\mathbf{Z}_{N+1}x_{N+1} \tag{40}$$

This is equivalent to $l = 1$ in (26). As (40) is the same form with (35), using the relationship of (39), (36) can be stated as

$$\mathbf{K}_{N+1} = \mathbf{P}_N\mathbf{Z}_{N+1}(1 + \mathbf{Z}_{N+1}^T\mathbf{P}_N\mathbf{Z}_{N+1})^{-1} \simeq \mathbf{A}_N^{-1}\mathbf{Z}_{N+1} \tag{41}$$

Therefore, we can get

$$\hat{\theta}_{N+l} \simeq \prod_{i=1}^{l}(\mathbf{I} - \mathbf{A}_{N-1+i}^{-1}\mathbf{Z}_{N+i}\mathbf{Z}_{N+i}^T)\hat{\theta}_N + \sum_{j=1}^{l-1}\left\{\prod_{i=1}^{l-j}(\mathbf{I} - \mathbf{A}_{N+l-i}^{-1}\mathbf{Z}_{N+l+1-i}\mathbf{Z}_{N+l+1-i}^T)\right\} \cdot \mathbf{A}_{N-1+j}^{-1}\mathbf{Z}_{N+j}x_{N+j} + \mathbf{A}_{N-1+l}^{-1}\mathbf{Z}_{N+l}x_{N+l} \tag{42}$$

(In detail, see Appendix 5)

The volume of calculation decreases considerably for the above equation compared with those for (38).

If data occurrence speed is faster than the calculation time for the above mentioned recursive algorithm, these approaches are not appropriate for the failure detection. These are the motivation of our considering new method handling.

Here, we briefly consider the comparison with Residual-Based Approach (RBA). As for the RBA, see Appendix 6. Even for RBA, the derived situation is the same as the original approach, considering the case that l data are newly obtained on the top of N data already received.

RBA have to calculate from (53) to (57) recursively for l times. This is the same calculation from (35) to (38). On the other hand, our approach takes less time by simple calculation method, stated from (39) to (42) and the reason for this is also stated below (40) and (42). We may state that our approach is much better in calculating multiple data receiving online with limited calculation ability.

6. NUMELICAL EXAMPLE

6.1 Numerical Calculation

We consider examples of the following three cases.

[1] $a_1 = -1.5,\ a_2 = 0.7$

[2] $a_1 = -1.4,\ a_2 = 0.6$

[3] $a_1 = -1.2,\ a_2 = 0.4$

On a 2nd order AR model of (1), we examine the case

$$N = 3000, 5000 \quad l = N/50, N/25, N/10$$

1000 times simulations were accomplished for each case with different initial points and the mean and the variance were examined. Calculation time in the table is summary of 1000 times calculation. The first 100 data are ignored and we used MATLAB for calculation.

Hereafter, we state the result mainly for section 3. Table 1 to Table 6 are their results.

6.2 Remarks

Numerical examples show that the new method has a fairly good estimation result and the calculation time is less than 1/10 compared with the case when the whole data are re-calculated. The level of the mean and the variance of each Table can safely be said sufficiently satisfactory.

Table 1. Case [1], $N = 3000$

	l	a_1 mean	a_1 variance ($\times 10^{-4}$)	a_2 mean	a_2 variance ($\times 10^{-4}$)	Calculation Time (sec)
(20)	60	-1.4997	1.7319	0.7001	1.6717	13.3890
$-\bar{\mathbf{A}}^{-1}\bar{\mathbf{C}}$	120	-1.4998	1.7202	0.7000	1 6715	13.3790
	300	-1.4998	1.7836	0.7000	1 7308	13 3790
(21)	60	-1 4821	235	0 7002	188	1.1420
$-\bar{\mathbf{B}}^{-1}\bar{\mathbf{D}}$	120	-1.4845	82	0 6927	70	1.2420
	300	-1.4935	23	0.6973	21	1.2420
(17)	60	-1.4997	1.7657	0.7002	1.6882	1 0820
1st order	120	-1 4996	1.7555	0.7000	1.6863	1.1620
approximation	300	-1.4996	1 6672	0.7000	1.5799	1.1620
(18)	60	-1.4997	1 7625	0.7002	1 6858	1.1610
2nd order	120	-1.4996	1.7480	0.7000	1.6801	1.2420
approximation	300	-1.4996	1.6559	0.7000	1.5742	1.2420
(19)	60	-1.4996	1.6523	0.7000	1.6112	13.4090
	120	-1.4996	1.6507	0 6998	1.6012	13.4000
	300	-1.4997	1.5777	0.6999	1.5065	13.4000

Table 2. Case [1], $N = 5000$

	l	a_1 mean	a_1 variance ($\times 10^{-4}$)	a_2 mean	a_2 variance ($\times 10^{-4}$)	Calculation Time (sec)
(20)	100	-1.4998	9.7090	0.7001	9.6703	28 8610
$-\bar{\mathbf{A}}^{-1}\bar{\mathbf{C}}$	200	-1.5002	10.619	0 7004	10.054	29.7630
	500	-1.5001	10.258	0.7004	9.9791	30.1340
(21)	100	-1 4833	116	0.6942	96	1.2320
$-\bar{\mathbf{B}}^{-1}\bar{\mathbf{D}}$	200	-1 4920	42	0 6968	36	1 4520
	500	-1 4958	12	0 6978	12	2.1730
(17)	100	-1 4997	9.8708	0.7001	9.8395	1.1510
1st order	200	-1.5001	10 683	0 7004	9.8656	1 3520
approximation	500	-1.5000	9 9346	0 7004	9.6123	1.9430
(18)	100	-1.4997	9 8552	0.7001	9 8250	1 2220
2nd order	200	-1 5001	10 653	0.7004	9 8485	1.4320
approximation	500	-1.5000	9.8530	0.7004	9.5323	1.9520
(19)	100	-1.4999	9.4092	0.7002	9.4313	29.9130
	200	-1.5001	10.198	0.7003	9.4714	30.3940
	500	-1 5001	9 5457	0.7004	9.3146	30.7040

Table 3. Case [2], $N = 3000$

	l	a_1		a_2		Calculation
		mean	variance ($\times 10^{-4}$)	mean	variance ($\times 10^{-4}$)	Time (sec)
(20)	60	-1.3997	2.1425	0.6002	2.0565	13.3390
$-\hat{A}^{-1}\hat{C}$	120	-1.3998	2.1571	0.6002	2.0926	13.3490
	300	-1.3998	2.2400	0.6001	2.1694	13.3090
(21)	60	-1.3791	222	0.6043	177	1.1310
$-\hat{B}^{-1}\hat{D}$	120	-1.3832	82	0.5947	72	1.2420
	300	-1.3935	24	0.5986	23	1.9130
(17)	60	-1.3996	2.1287	0.6003	2.0396	1.0820
1st order	120	-1.3995	2.1435	0.6001	2.0585	1.1620
approximation	300	-1.3996	2.0371	0.6002	1.9376	1.7620
(18)	60	-1.3996	2.1269	0.6003	2.0383	1.1420
2nd order	120	-1.3995	2.1384	0.6001	2.0545	1.2420
approximation	300	-1.3996	2 0330	0 6001	1.9382	1 8330
(19)	60	-1.3997	2.0445	0.6001	1.9860	13.4000
	120	-1.3997	2.0781	0.6000	2.0078	13.4190
	300	-1.3997	1.9838	0.6000	1.8962	13 4200

Table 4. Case [2], $N = 5000$

	l	a_1		a_2		Calculation
		mean	variance ($\times 10^{-4}$)	mean	variance ($\times 10^{-4}$)	Time (sec)
(20)	100	-1.3998	12.015	0.6002	11.957	28.7210
$-\hat{A}^{-1}\hat{C}$	200	-1.4002	13.368	0.6005	12.721	28.6610
	500	-1.4003	12.821	0.6007	12.597	29.5420
(21)	100	-1.3837	116	0.5987	98	1.2420
$-\hat{B}^{-1}\hat{D}$	200	-1.3911	45	0.5980	39	1.4520
	500	-1.3957	14	0.5986	14	1.9830
(17)	100	-1.3997	12.005	0.6003	12.014	1.1610
1st order	200	-1.4001	13.216	0.6005	12.252	1.3420
approximation	500	-1.4001	12.300	0.6006	12.017	1.8020
(18)	100	-1.3997	11.994	0.6003	12.003	1.2320
2nd order	200	-1.4001	13.193	0.6005	12.245	1.4020
approximation	500	-1.4001	12.218	0.6006	11.935	1.8730
(19)	100	-1.3999	11.693	0.6003	11.725	29.1720
	200	-1.4001	12.868	0.6004	11.987	28.7710
	500	-1.4002	11.955	0.6006	11.748	29.6030

Table 5. Case [3], $N = 3000$

	l	a_1		a_2		Calculation
		mean	variance ($\times 10^{-4}$)	mean	variance ($\times 10^{-4}$)	Time (sec)
(20)	60	-1.1997	2.7749	0.4006	2.6357	13.3490
$-\hat{A}^{-1}\hat{C}$	120	-1.1999	2.8601	0.4006	2.7539	13.3190
	300	-1.1998	2.9480	0.4004	2.8381	13.3190
(21)	60	-1.1751	205	0.4146	164	1.1320
$-\hat{B}^{-1}\hat{D}$	120	-1.1832	83	0.4007	75	1.2520
	300	-1.1936	27	0.4013	27	1.9630
(17)	60	-1.1996	2.7062	0.4008	2.5859	1.0710
1st order	120	-1.1996	2.7776	0.4005	2.6522	1.1820
approximation	300	-1.1996	2.6314	0.4006	2.5036	1.7340
(18)	60	-1.1996	2.7061	0.4008	2.5875	1.1520
2nd order	120	-1.1996	2.7754	0.4005	2.6509	1.2510
approximation	300	-1.1996	2.6360	0.4006	2 5131	1.9430
(19)	60	-1.1998	2.6648	0.4006	2 5654	13.3690
	120	-1.1998	2.7490	0 4004	2.6337	13.3590
	300	-1.1998	2.6306	0.4004	2 5071	13.3990

Table 6. Case [3], $N = 5000$

	l	a_1		a_2		Calculation
		mean	variance ($\times 10^{-4}$)	mean	variance ($\times 10^{-4}$)	Time (sec)
(20)	100	-1.1998	15.569	0.4005	15.250	29.5720
$-\hat{A}^{-1}\hat{C}$	200	-1.2002	17.494	0.4007	16.736	29 6030
	500	-1 2005	16.722	0.4011	16 510	30.1530
(21)	100	-1.1845	119	0.4077	105	1.2520
$-\hat{B}^{-1}\hat{D}$	200	-1.1895	49	0.4006	43	1.4920
	500	-1.1955	18	0.3999	18	1.9630
(17)	100	-1.1998	15.324	0.4006	15.167	1.1720
1st order	200	-1.2000	17.031	0.4007	15.899	1.3820
approximation	500	-1.2002	15 938	0.4010	15.620	1.8020
(18)	100	-1.1998	15 320	0.4006	15.159	1.2520
2nd order	200	-1.2000	17.031	0.4006	15.905	1.4820
approximation	500	-1.2002	15 848	0.4010	15.334	1.9830
(19)	100	-1 1999	15 190	0 4005	15.030	30.6440
	200	-1.2001	16.886	0.4005	15.810	30.2040
	500	-1.2004	15.625	0 4009	15.366	30.2440

Generally, the following results are expected, although they do not come true always.

(a) Accuracy in the parameter estimation is better in large N than in small N. Comparison of estimation accuracy is carried out by the value

$$|\hat{a}_1 - a_1| + |\hat{a}_2 - a_2|$$

If two estimations have the same values, then the case with smaller variance is judged to be better.

(b) Data amount of $N + l$ in (19) would be better than N in (20) in the estimation.

(c) The calculation result by only l data amount is rough in estimation.

(d) 2nd order approximation is better than 1st order approximation in estimation accuracy.

(e) 1st order, 2nd order approximation are better than (20) in estimation accuracy.

(f) 1st order approximation takes less time than 2nd order approximation in calculation

(g) 1st order, 2nd order approximation is much shorter than (19) in calculation time.

(h) When l become larger, the estimation accuracy becomes better.

(i) When the absolute value of a_1, a_2 become larger, the estimation accuracy become better.

Examining these hypotheses, we get Table 7.

Almost all hypotheses fit with the result. (b) or (e) in the case $N = 3000$ could not find much difference.

Although we find few differences partially, it can be said that all hypotheses from (a) to (i) fit with the result. As the above hypotheses are based on macroscopic viewpoint, we have to examine much further to another cases. From these considerations, it can be said that the

Table 7. The result of comparison

	$N=3000$			$N=5000$		
	$l=60$	$l=120$	$l=300$	$l=100$	$l=200$	$l=500$
(a)	o					
(b)	–			o		
(c)	o	o	o	o	o	o
(d)	o	o	o	o	o	o
(e)	–	–	–	o	o	o
(f)	o	o	o	o	o	o
(g)	o	o	o	o	o	o
(h)	o			–		
(i)	o					

o : be able to see trends – : few difference

estimation accuracy is rough only by the l amount of data. However, by using the 1st order approximation, we can get a good estimation result and the calculation time is less than 1/10 to the case whole data are re-calculated. 1st order and 2nd order approximation do not have many differences in estimation.

We examined $N=1000$, 2000 (small size) and $N=10000$ (large size). When we take smaller N, l becomes much smaller and the estimation result becomes more rough. When we take larger size, parameter estimation result comes quite close to the real values, so that it is hard to find difference. The cases $N=3000$, 5000, which we take, are the suitable cases with proper weighting influence and the results show difference between N amount of data and $N+l$ amount of data.

7. CONCLUSIONS

A new simplified calculation method to estimate parameter when l amount of data are gathered following N amount of data is proposed. Using this method, the calculation time can be reduced to 1/10 of the one needed for the recalculation with the whole data.

In the field of plant maintenance, large amount of data gathered by sensors on multiple machines should be handled and analyzed speedily. However, there are often cases that the analyzing machine has limited calculating ability. In those fields, the proposed method enables failure detection of the equipment in early stages with a few new coming data. This method would be applicable in many data handling fields. Finding and examining such new field would be a great issue in the near future.

APPENDIX

Appendix 1

$$\begin{aligned}
\hat{R}_{N+l,j} &= \frac{1}{N+l-j}\sum_{i=1}^{N+l-j} x_i x_{i+j} \\
&= \frac{1}{N+l-j}\left(\sum_{i=1}^{N-j} x_i x_{i+j} + \sum_{i=N-j+1}^{N+l-j} x_i x_{i+j}\right) \\
&= \frac{N-j}{N+l-j}\frac{1}{N-j}\sum_{i=1}^{N-j} x_i x_{i+j} \\
&\quad + \frac{l}{N+l-j}\frac{1}{l}\sum_{i=N-j+1}^{N+l-j} x_i x_{i+j} \\
&= \frac{N-j}{N+l-j}\hat{R}_{N,j} + \frac{l}{N+l-j}\hat{R}_{N/l,j}
\end{aligned} \tag{43}$$

Appendix 2

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_{N+l} &= -(\alpha\mathbf{A})^{-1}(\alpha\mathbf{C}) - (\alpha\mathbf{A})^{-1}(\beta\mathbf{D}) \\
&+ (\alpha\mathbf{A})^{-1}(\beta\mathbf{B})\left[\mathbf{I} + (\alpha\mathbf{A})^{-1}\beta\mathbf{B}\right]^{-1}(\alpha\mathbf{A})^{-1}(\alpha\mathbf{C}+\beta\mathbf{D}) \\
&= \hat{\boldsymbol{\theta}}_N - \frac{l}{N}\mathbf{A}^{-1}\mathbf{D} - \frac{l}{N}\mathbf{A}^{-1}\mathbf{B}\hat{\boldsymbol{\theta}}_N \\
&+ \left(\frac{l}{N}\right)^2(\mathbf{A}^{-1}\mathbf{B})^2\hat{\boldsymbol{\theta}}_N + \left(\frac{l}{N}\right)^2\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{D} \\
&- \left(\frac{l}{N}\right)^3\mathbf{A}^{-1}\mathbf{B}\left[\mathbf{I}+\frac{l}{N}\mathbf{A}^{-1}\mathbf{B}\right]^{-1}(\mathbf{A}^{-1}\mathbf{B})^2\hat{\boldsymbol{\theta}}_N \\
&- \left(\frac{l}{N}\right)^3(\mathbf{A}^{-1}\mathbf{B})^2\mathbf{A}^{-1}\mathbf{D} \\
&+ \left(\frac{l}{N}\right)^4\mathbf{A}^{-1}\mathbf{B}\left[\mathbf{I}+\frac{l}{N}\mathbf{A}^{-1}\mathbf{B}\right]^{-1}(\mathbf{A}^{-1}\mathbf{B})^2\mathbf{A}^{-1}\mathbf{D}
\end{aligned} \tag{44}$$

Appendix 3

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_{N+l} &= -\left[\bar{\mathbf{A}}+\bar{\mathbf{B}}\right]^{-1}\left[\bar{\mathbf{C}}+\bar{\mathbf{D}}\right] \\
&= -\left[\mathbf{I}+\bar{\mathbf{A}}^{-1}\bar{\mathbf{B}}\right]^{-1}\left[\bar{\mathbf{A}}^{-1}\bar{\mathbf{C}}+\bar{\mathbf{A}}^{-1}\bar{\mathbf{D}}\right] \\
&= -\left[\mathbf{I}+\bar{\mathbf{A}}^{-1}\bar{\mathbf{B}}\right]^{-1}\left[\bar{\mathbf{A}}^{-1}\bar{\mathbf{C}}+\left(\bar{\mathbf{A}}^{-1}\bar{\mathbf{B}}\right)\right]
\end{aligned} \tag{45}$$

Appendix 4

(34) can be re-stated as

$$\hat{\boldsymbol{\theta}}_{N+1} = (\mathbf{I} - \mathbf{K}_{N+1}\mathbf{Z}_{N+1}^T)\hat{\boldsymbol{\theta}}_N + \mathbf{K}_{N+1}x_{N+1} \tag{46}$$

Repeating iteration, we get

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{N+2} &= (\mathbf{I}-\mathbf{K}_{N+2}\mathbf{Z}_{N+2}^T)(\mathbf{I}-\mathbf{K}_{N+1}\mathbf{Z}_{N+1}^T)\hat{\boldsymbol{\theta}}_N \\ &\quad +(\mathbf{I}-\mathbf{K}_{N+2}\mathbf{Z}_{N+2}^T)\mathbf{K}_{N+1}x_{N+1} \\ &\quad +\mathbf{K}_{N+2}x_{N+2}\end{aligned} \tag{47}$$

$$\begin{aligned}&\hat{\boldsymbol{\theta}}_{N+3} = (\mathbf{I}-\mathbf{K}_{N+3}\mathbf{Z}_{N+3}^T)\hat{\boldsymbol{\theta}}_{N+2} + \mathbf{K}_{N+3}x_{N+3} \\ &= (\mathbf{I}-\mathbf{K}_{N+3}\mathbf{Z}_{N+3}^T)(\mathbf{I}-\mathbf{K}_{N+2}\mathbf{Z}_{N+2}^T)(\mathbf{I}-\mathbf{K}_{N+1}\mathbf{Z}_{N+1}^T)\hat{\boldsymbol{\theta}}_N \\ &+(\mathbf{I}-\mathbf{K}_{N+3}\mathbf{Z}_{N+3}^T)(\mathbf{I}-\mathbf{K}_{N+2}\mathbf{Z}_{N+2}^T)\mathbf{K}_{N+1}x_{N+1} \\ &+(\mathbf{I}-\mathbf{K}_{N+3}\mathbf{Z}_{N+3}^T)\mathbf{K}_{N+2}x_{N+2} \\ &+\mathbf{K}_{N+3}x_{N+3}\end{aligned} \tag{48}$$

These are generalized as

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{N+l} &= \prod_{i=1}^{l}(\mathbf{I}-\mathbf{K}_{N+i}\mathbf{Z}_{N+i}^T)\hat{\boldsymbol{\theta}}_N \\ &+ \sum_{j=1}^{l-1}\left\{\prod_{i=1}^{l-j}(\mathbf{I}-\mathbf{K}_{N+l+1-i}\mathbf{Z}_{N+l+1-i}^T)\right\} \\ &\cdot \mathbf{K}_{N+j}x_{N+j} \\ &+ \mathbf{K}_{N+l}x_{N+l}\end{aligned} \tag{49}$$

Appendix5

Utilizing (41), we can easily obtain

$$\begin{aligned}\mathbf{K}_{N+i} &\simeq \mathbf{A}_{N-1+i}^{-1}\mathbf{Z}_{N+i} \\ \mathbf{K}_{N+l+1-i} &\simeq \mathbf{A}_{N+l-i}^{-1}\mathbf{Z}_{N+l+1-i} \\ \mathbf{K}_{N+j} &\simeq \mathbf{A}_{N-1+j}^{-1}\mathbf{Z}_{N+j} \\ \mathbf{K}_{N+l} &\simeq \mathbf{A}_{N-1+l}^{-1}\mathbf{Z}_{N+l}\end{aligned}$$

Applying these to each corresponding part of (38), we can get

$$\begin{aligned}&\hat{\boldsymbol{\theta}}_{N+l} \simeq \prod_{i=1}^{l}(\mathbf{I}-\mathbf{A}_{N-1+i}^{-1}\mathbf{Z}_{N+i}\mathbf{Z}_{N+i}^T)\hat{\boldsymbol{\theta}}_N \\ &+\sum_{j=1}^{l-1}\left\{\prod_{i=1}^{l-j}(\mathbf{I}-\mathbf{A}_{N+l-i}^{-1}\mathbf{Z}_{N+l+1-i}\mathbf{Z}_{N+l+1-i}^T)\right\} \\ &\cdot\mathbf{A}_{N-1+j}^{-1}\mathbf{Z}_{N+j}x_{N+j} + \mathbf{A}_{N-1+l}^{-1}\mathbf{Z}_{N+l}x_{N+l}\end{aligned} \tag{50}$$

Appendix 6

Suppose that system is described as

$$x(k+1) = Ax(k) + \Gamma\omega(k) \tag{51}$$

$$y(k) = Cx(k) + v(k) \tag{52}$$

Where input $w(k)$ and measurement noise $v(k)$ have covariance Q, R for each and are independent white noise series. Under the normal condition, the optimal estimate of $\hat{x}(k+1|k)$ is given by the following Kalman Filter.

$$\hat{x}(k+1|k) = A\hat{x}(k|k-1) + K(k)\{y(k) - C\hat{x}(k|k-1)\} \tag{53}$$

$$K(k) = AP(k|k-1)C^TV^{-1}(k) \tag{54}$$

$$\begin{aligned}P(k+1|k) &= AP(k|k-1)A^T + \Gamma Q\Gamma^T \\ &\quad - K(k)CP(k|k-1)A^T\end{aligned} \tag{55}$$

$$V(k) = CP(k|k-1)C^T + R \tag{56}$$

Residual series of Kalman Filter $\gamma(k)$ is defined as

$$\gamma(k) = y(k) - C\hat{x}(k|k-1) \tag{57}$$

$\gamma(k)$ becomes gaussian white noise series with mean 0, covariance $V(k)$. For failure detection, we have to make χ^2-test to the Residual series. For the latest l amount of data, we define

$$L(k) = \frac{1}{N}\sum_{j=k-l+1}^{k}\gamma^T(j)V^{-1}(j)\gamma(j) \tag{58}$$

then, $L(k)$ follows χ^2-distribution. When the system become irregular , residual become large, so we can guess the system as

$$L(k) \geq \epsilon : irregular$$

$$L(k) < \epsilon : normal$$

Though there may be may variation of RBA method, the above method is at least a typical one of RBA (Wilsky, 1976; Yamazaki, 1997)

REFERENCES

Wilsky, A. L. (1976) A Survey of Design Methods for Failure Detection in Dynamic Systems. *Automatica* **12**, 601-611.

Katayama, T. (1994) *System Identification* (In Japanese). Asakura-Shoten Publishing.

Nakamura, M. and Y. Oishi (1984) Generalized Least Square Estimation Algorithm Designed to Reduce A Calculation Time (In Japanese). *SICE Trans.* **20**(6), 471-478.

Sagara, S., Akitsuki, K., Nakamizo, T. and Katayama, T. (1994) *System Identification* (In Japanese). The Society of Instrument and Control Engineers.

Tokumaru, H., Soeda, T., Nakamizo, T. and Akitsuki, K. (1982) *Measurement and Analysis-Theory and Application of Random Data Handling* (In Japanese). Baifukan Publishing.

Yamagata, K., Takeda, K., Tsuge, Y. and Matsuyama, H. (2001) Real-time Failure Diagnosis of Instruments and State Estimation Using Plant Dynamics (In Japanese). *J.SOPE Japan* **13**(1), 1-9.

Yamazaki, H. (1977) *Failure Detection and Prediction* (In Japanese). Kogyo Chosakai Publishing.