Application of Maximum Likelihood Method to Boiler System Identification

By

Tohru KATAYAMA*, Keiichi AKIMOTO** and Yoshikazu SAWARAGI**

(Received September 30, 1976)

Abstract

A maximum likelihood (ML) method is applied to a boiler system identification. The mathematical model used in this paper is a discrete-time, single-input and single-output(SISO), constant, linear system excited by an "innovation" process. Since the ML identification is reduced to a nonlinear optimization problem with equality constraints, the Davidon's conjugate gradient method is employed for numerical solutions. By using the given input/output data, the dynamics of the governor/steam pressure and the governor/steam temperature relations are identified as an SISO system, respectively. AIC and a test based on the innovation process are also applied for selecting an appropriate order of the assumed model.

1. Introduction

A design of an optimal control system requires a complete state-space description of system dynamics. In practice, however, this is rarely the case for complex systems or processes. The construction of a dynamic model based on the observable input/output data is referred to as a system identification. Hence, the system identification is one of the most important problems in systems and control theory. To identify actual processes such as a chemical plant, flying object, or economic system, it may be very important to make use of *a priori* knowledge from chemisty, physics, economics, and so forth. The real processes are, however, so complex that the present status of engineering and science may not be sufficiently able to analyze the real systems in detail. Therefore, in this paper we wish to identify the unknown system dynamics based on the observable input/output data.

* Department of Applied Mathematics and Physics.
** Present Address: The Kawasaki Steel Industry, Mizushima, Okayama.

The identification problem is mainly characterized by the class of the assumed model, the criterion of parameter estimation, and the class of inputs to be applied[1]. In this paper, we consider an ML identification method[2),3)] with its application to a boiler system identification. The mathematical model used here is a discrete-time, SISO, constant, linear system called the "innovation" representation, or the steady state Kalman filter[4]. Assuming that the innovation process is gaussian white, the ML method yields a state-space model suitable for the application of the modern control theory. Since the ML identification is reduced to a nonlinear optimization problem with equality constraints, the standard optimization techniques can be applied for the numerical solutions. Although there are other possible identification methods, such as the least square (LS) method, the instrumental variable method, or the stochastic approximation method, it has been shown that the ML method gives asymptotically unbiased and consistent estimates under certain conditions on the input signal[2]. Thus, the ML method may have advantages over other methods, although it is computationally rather expensive.

The organization of this paper is as follows. In section 2, we describe the mathematical model used and the assumptions on the system structure. In section 3, the identification based on the LS method is shown briefly. Section 4 is devoted to the development of the ML method of system identification. Possible approaches to the estimation of the order of the assumed model are examined in section 5, and the overall algorithm of identification is summarized in section 6. Finally, in section 7 we will show the numerical results of the identification of the once-through type boiler system in the Kainan electric power plant of the Kansai Electric Company[5),6)].

2. Mathematical Model

Consider a discrete-time, SISO, constant, linear system described by

$$\begin{aligned} x(t+1) &= Fx(t) + Gu(t) + Lw(t) \\ y(t) &= Hx(t) + v(t) \end{aligned} \tag{2.1}$$

where $x(t)$ is an $n\times 1$ state vector at sample time t, $u(t)$ is a scalar input to the system, and $y(t)$ is a scalar output. The F is an $n\times n$ state transition matrix, G is an $n\times 1$ driving matrix, H is a $1\times n$ observation matrix, and L is an $n\times 1$ matrix. The plant and observation noises $w(t)$ and $v(t)$ are zero mean, white gaussian sequences with finite variances.

It is well known that only the innovation representation[7] of system (2. 1) is identifiable. In other words, the above system (2. 1) has a redundancy in the noise model as well as in system parameters, so that not all the parameters are identifiable. In order to obtain an identifiable form, we define the one-step predicted estimate $\hat{x}(t)$ of the state $x(t)$ based on the observations $U^{t-1}=\{u(0), \ldots, u(t-1)\}$ and $Y^{t-1}=$

$\{y(0), \ldots, y(t-1)\}$ as

$$\hat{x}(t) = E\{x(t) \mid U^{t-1},\ Y^{t-1}\} \tag{2.2}$$

where $E\{\cdot \mid \cdot\}$ denotes the conditional expectation. Then, by using the theory of the Kalman filter[8], we have the innovation representation of (2. 1):

$$\begin{aligned} \hat{x}(t+1) &= F\hat{x}(t) + Gu(t) + Ke(t) \\ y(t) &= H\hat{x}(t) + e(t) \end{aligned} \tag{2.3}$$

where $e(t)$ is a scalar white noise called the innovation that denotes the one-step prediction error for $y(t)$, given the data up to $t-1$:

$$e(t) = y(t) - H\hat{x}(t) = y(t) - \hat{y}(t \mid t-1) \tag{2.4}$$

and K is an $n \times 1$ matrix that denotes the steady state Kalman filter gain. It should be noted that in (2. 3), the state vector is the one-step predicted estimate of the original state vector of (2. 1), and that system (2. 3) is equivalent in the sense of quadratic mean to system (2. 1) as far as the input/output relation is concerned. In what follows, we consider system (2. 3) as our model of the process to be identified.

It is well known that an SISO system has canonical representations[9]. In this paper we use the following observable canonical form:

$$F = \begin{pmatrix} 0 & 0 \ldots & -a_n \\ 1 & 0 \ldots & -a_{n-1} \\ & 1 & \\ & \ddots & \vdots \\ 0 & & 1 \quad -a_1 \end{pmatrix} \qquad G = \begin{pmatrix} b_n \\ b_{n-1} \\ \vdots \\ b_1 \end{pmatrix} \qquad K = \begin{pmatrix} k_n \\ k_{n-1} \\ \vdots \\ k_1 \end{pmatrix} \tag{2.5}$$

$$H = [0 \ldots 0 \ 1]$$

The pair (H, F) is necessarily observable; however, the pairs (F, G) and (F, K) are not always controllable. Therefore, to obtain a minimal realization that contains the smallest possible number of parameters, we should delete the uncontrollable part of the state vector $\hat{x}(t)$. There exist several algorithms for the minimal realization[10),11].

From (2. 3) and (2. 5), we have the following input/output relation:

$$y(t) + \sum_{i=1}^{n} a_i y(t-i) = \sum_{i=1}^{n} b_i u(t-i) + e(t) + \sum_{i=1}^{n} c_i e(t-i) \tag{2.6}$$

where $c_i = a_i + k_i (i = 1, \ldots, n)$, and it is assumed that the initial conditions

$$\begin{aligned} &y(0),\ y(-1), \ldots,\ y(-n+1) \\ &u(0),\ u(-1), \ldots,\ u(-n+1) \end{aligned} \tag{2.7}$$

are known. Equation (2. 6) is a version of the ARMA model that is popular in the field of time-series analysis[12]. Taking z-transform of (2. 6) yields

$$A(z^{-1})y(t) = B(z^{-1})u(t) + C(z^{-1})e(t) \tag{2.8}$$

where

$$\begin{aligned} A(z^{-1}) &= 1+a_1z^{-1}+\cdots+a_nz^{-n} \\ B(z^{-1}) &= \quad b_1z^{-1}+\cdots+b_nz^{-n} \\ C(z^{-1}) &= 1+c_1z^{-1}+\cdots+c_nz^{-n} \end{aligned} \tag{2.9}$$

and z^{-m} is the shift operator defined as $z^{-m}y(t)=y(t-m)$. From (2. 8), we can readily obtain the pulse transfer function representation of (2. 6).

In the following we assume that

(A1) All the roots of $A(z^{-1})=0$ lie in the unit circle.

(A2) All the roots of $C(z^{-1})=0$ lie in the unit circle.

Assumption (A1) implies that the homogeneous equation corresponding to (2. 6) is asymptotically stable, and (A2) implies that $C(z^{-1})$ is invertible. We define

$$\begin{aligned} \boldsymbol{a} &= (a_1a_2\ldots a_n)' \\ \boldsymbol{b} &= (b_1b_2\ldots b_n)' \\ \boldsymbol{c} &= (c_1c_2\ldots c_n)' \end{aligned}$$

where the prime(') denotes the transposition of a matrix or vector.

The problem of this paper is to determine the optimal order n_0 of the system and to estimate the parameters $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ based on the available input/output data U^N and Y^N. The determination of the optimal order is much more difficult than the parameter estimation for a fixed n. We will discuss this problem in section 5.

3. Least Square Identification

Before considering the ML identification method, we state the LS method. The estimated parameter values by the LS method will be used as the initial conditions for a parametric optimization algorithm in the ML identification.

In general, there exist three assumptions that may be regarded as *a priori* knowledge in the parameter estimation or system identification[1]:

a) $p(\boldsymbol{e})$: probability density function (pdf) of the noise process $e(t)$. From this we have the conditional pdf $p(\boldsymbol{y}|\boldsymbol{\theta})$ of the output $y(t)$ when the parameter value is $\boldsymbol{\theta}$.

b) $p(\boldsymbol{\theta})$: *a priori* pdf of the parameter $\boldsymbol{\theta}$.

c) $l(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$: loss function when we take $\hat{\boldsymbol{\theta}}$ as an estimate of $\boldsymbol{\theta}$.

If all the conditions a), b) and c) are available as *a priori* information, we can apply the Bayesian approach. If only a) is available, the ML method will be suitable, and if no *a priori* knowledge is given, the LS method will be applied.

Assuming that $\boldsymbol{c}=0$, we apply the LS method to the estimation of $\boldsymbol{a}$ and $\boldsymbol{b}$. From (2. 6) we have

$$y(t)+\sum_{i=1}^{n} a_iy(t-i)=\sum_{i=1}^{n} b_iu(t-i)+e(t) \tag{3.1}$$

where $t=-n+1, \ldots, -1, 0, \ldots, N$. For simplicity, we define

$$\boldsymbol{y}=[y(1)y(2)\cdots y(N)]'$$

$$\boldsymbol{e}=[e(1)e(2)\cdots e(N)]'$$

$$\boldsymbol{A}=\begin{pmatrix} y(0) & y(-1) & \cdots y(-n+1) & u(0) & u(-1) & \cdots u(-n+1) \\ y(1) & y(0) & \cdots y(-n+2) & u(1) & u(0) & \cdots u(-n+2) \\ \vdots & \vdots & & \vdots & \vdots & \vdots \\ y(N-1) & y(N-2) & \cdots y(N-n) & u(N-1) & u(N-2) & \cdots u(N-n) \end{pmatrix}$$

$$\boldsymbol{\alpha}=\begin{pmatrix} -\boldsymbol{a} \\ \boldsymbol{b} \end{pmatrix}$$

Then (3. 1) becomes

$$\boldsymbol{y}=\boldsymbol{A}\boldsymbol{\alpha}+\boldsymbol{e} \tag{3.2}$$

Assuming that the noise term $\boldsymbol{e}$ is independent of the coefficients matrix, the LS estimate $\hat{\boldsymbol{\alpha}}$ of $\boldsymbol{\alpha}$ is given by

$$\hat{\boldsymbol{\alpha}}=[\boldsymbol{A}'\boldsymbol{A}]^{-1}\boldsymbol{A}'\boldsymbol{y} \tag{3.3}$$

where it is assumed that $\boldsymbol{A}'\boldsymbol{A}$ is nonsingular. Actually, in (3. 1) the serially correlated noise $\{e(t)+\sum_{i=1}^{n} c_i e(t-i), t=1, \ldots, N\}$ was replaced by an uncorrelated noise $e(t)$, so that the LS estimate $\hat{\boldsymbol{\alpha}}$ will be biased. However, the LS estimation method requires fewer assumptions compared with other estimation methods, and also requires less computer time. Moreover, the LS method sometimes gives reasonable results from a practical point of view.

4. Maximum Likelihood Identification

4.1 Maximum likelihood method

Given the conditional pdf $p(\boldsymbol{y}|\boldsymbol{\theta})$ of the outputs $y(t)$ relative to $\boldsymbol{\theta}$, we define the likelihood function of $\boldsymbol{\theta}$ as

$$L(\boldsymbol{\theta})=p(\boldsymbol{y}|\boldsymbol{\theta}) \tag{4.1}$$

For a given sample value of $\boldsymbol{y}$, $L(\boldsymbol{\theta})$ is merely a function of the parameter. An ML estimate is the value of $\boldsymbol{\theta}$, which will be denoted by $\hat{\boldsymbol{\theta}}_{ML}$, that maximizes $L(\boldsymbol{\theta})$. In other words, the ML estimate $\hat{\boldsymbol{\theta}}_{ML}$ is determined so that the sample value of $\boldsymbol{y}$ is most likely.

It is well known that the MLE $\hat{\boldsymbol{\theta}}_{ML}$ has some important statistical properties[2),22)].

a) Asymptotic normality, that is, the convergence of $\sqrt{N}(\hat{\boldsymbol{\theta}}^N_{ML}-\boldsymbol{\theta})$ in distribution to a random vector where $\hat{\boldsymbol{\theta}}^N_{ML}$ is the MLE of $\boldsymbol{\theta}$ from N observations.

b) Asymptotic efficiency, that is, the covariance matrix of $\hat{\boldsymbol{\theta}}^N_{ML}-\boldsymbol{\theta}$ attains the Cramer-Rao lower bound for $N\to\infty$.

c) Consistency, that is, $\hat{\boldsymbol{\theta}}_{ML}^{N}$ converges in probability to $\boldsymbol{\theta}$ for $N\to\infty$.

In the following, $\boldsymbol{\theta}$ denotes the aggregation of the unknown parameters $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$. In (2. 6), assume that $\{e(t),\ t=1, \ldots, N\}$ are gaussian white noises with a mean zero and variance σ^2. Then the *conditional* pdf of Y^N given U^{N-1}, $\boldsymbol{\theta}$, σ^2 and the initial conditions can be written as

$$p(Y^N|U^{N-1}, \boldsymbol{\theta}, \sigma^2, \text{ ICs}) = \prod_{t=1}^{N} p(y(t)\,|\,u(i),\ y(i),\ i\leqslant t-1,\ \boldsymbol{\theta},\ \sigma^2)$$
$$= \prod_{t=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}e^2(t)\right\} \tag{4.2}$$

where ICs denotes the initial conditions (2. 7), and $e(t)$ is defined by (2. 4). Thus the *conditional* log-likelihood function becomes

$$L_N(\boldsymbol{\theta},\ \sigma^2) = \log p(Y^N|U^{N-1},\ \boldsymbol{\theta},\ \sigma^2,\ \text{ICs})$$
$$= -\frac{N}{2}\log\,(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{N} e^2(t) \tag{4.3}$$

Since the logarithm is monotone increasing, the MLE's $\hat{\boldsymbol{\theta}}$ and σ^2 maximize (4. 3). Differentiating (4. 3) with respect to σ^2 and setting the resulting equation at zero, we have

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{t=1}^{N} e^2(t) \tag{4.4}$$

Substituting (4. 4) into (4. 3) yields

$$-L_N(\boldsymbol{\theta},\ \hat{\sigma}^2) = \frac{N}{2}(1+\log\,(2\pi)) + \frac{N}{2}\log\,\hat{\sigma}^2 \tag{4.5}$$

Therefore, the MLE of $\boldsymbol{\theta}$ can be obtained by minimizing

$$L_N^*(\boldsymbol{\theta}) = \frac{1}{N}\sum_{t=1}^{N} e^2(t) \tag{4.6}$$

subject to equality constraints (2. 6). Hence, the ML identification is reduced to a parametric optimization problem with equality constraints that minimize the sample variance of the innovation process. This result appeals to our physical intuition, although the entire development is dependent on the assumption that the innovation is a gaussian white noise.

4.2 Optimization method

In this section, we derive the algorithm to obtain the minimum point of $L_N^*(\boldsymbol{\theta})$ subject to equality constraints (2. 6). It should be noted that, since $e(t)$, defined by (2. 6), is linear with respect to $\boldsymbol{a}$ and $\boldsymbol{b}$, $L_N^*(\boldsymbol{\theta})$ is quadratic in $\boldsymbol{a}$ and $\boldsymbol{b}$, but highly nonlinear in $\boldsymbol{c}$. Thus, we use the Davidon's method[13),14)] for numerical solutions. Application of the Newton's method is rather difficult, because it requires the com-

putation of the Hessian matrix[2].

First, we compute the gradients of $L_N^*(\boldsymbol{\theta})$ under constraints (2. 6). The following method is based on the use of adjoint variables[3]. Define the Hamiltonian

$$H(\boldsymbol{e},\ \boldsymbol{\lambda},\ \boldsymbol{\theta}) = \frac{1}{N}\sum_{t=1}^{N} e^2(t) + \frac{1}{N}\sum_{t=1}^{N}\lambda(t)\Big[e(t) + \sum_{i=1}^{n} c_i e(t-i) + \sum_{i=1}^{n} b_i u(t-i) - y(t) - \sum_{i=1}^{n} a_i y(t-i)\Big] \tag{4.7}$$

where $\{\lambda(t),\ t=1, \ldots, N\}$ are adjoint variables. The difference equations satisfied by the adjoint variables are obtained from

$$\frac{\partial}{\partial e(t)} H(\boldsymbol{e},\ \boldsymbol{\lambda},\ \boldsymbol{\theta}) = 0, \quad (t=1, \ldots, N) \tag{4.8}$$

so that we have

$$\lambda(t) + \sum_{i=1}^{n} c_i \lambda(t+i) + 2e(t) = 0 \tag{4.9}$$

with the terminal conditions

$$\lambda(t) = 0, \quad t = N+1, \ldots, \ N+n \tag{4.10}$$

Then the constrained partial derivatives can be obtained as

$$\begin{aligned} \frac{\partial L_N^*(\boldsymbol{\theta})}{\partial a_i} &= \frac{\partial H}{\partial a_i} = -\frac{1}{N}\sum_{t=1}^{N}\lambda(t) y(t-i) \\ \frac{\partial L_N^*(\boldsymbol{\theta})}{\partial b_i} &= \frac{\partial H}{\partial b_i} = \ \ \frac{1}{N}\sum_{t=1}^{N}\lambda(t) u(t-i) \\ \frac{\partial L_N^*(\boldsymbol{\theta})}{\partial c_i} &= \frac{\partial H}{\partial c_i} = \ \ \frac{1}{N}\sum_{t=1}^{N}\lambda(t) e(t-i) \end{aligned} \tag{4.11}$$

where $i=1, \ldots, \ n$.

Therefore, the procedure of computing the gradients are summarized as follows.

(1) Assuming values of the parameters $\boldsymbol{\theta}$, compute $\{e(t),\ t=1, \ldots, N\}$ according to the relation:

$$e(t) + \sum_{i=1}^{n} c_i e(t-i) = y(t) + \sum_{i=1}^{n} a_i y(t-i) - \sum_{i=1}^{n} b_i u(t-i) \tag{4.12}$$

where the initial conditions for $e(t)$ are given by

$$e(t) = y(t), \quad t = 0, \ -1, \ \ldots, \ -n+1$$

(2) Integrate (4. 9) backward using the terminal conditions (4. 10) to obtain the adjoint variables $\{\lambda(t),\ t=1, \ldots, N\}$.

(3) Compute the gradients by (4. 11).

It should be noted that assumption (A2) ensures the stability of the integrations in steps (1) and (2).

Now we are in position to describe the Davidon's algorithm[13),14]. Define a

$3n\times 1$ vector

$$g_\nu=\frac{\partial L_N^*(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_\nu}$$

where the subscript ν denotes the number of iterations. Then the algorithm can be summarized as follows.

Step 1: Set $\hat{\boldsymbol{\theta}}_1=\begin{bmatrix}\hat{\boldsymbol{a}}\\0\end{bmatrix}$ and $d_1=-g_1$.

Step 2: For $\nu=1, 2, \ldots$, find the values $\lambda=\lambda_\nu$ such that $\phi_\nu(\lambda)=L_N^*(\hat{\boldsymbol{\theta}}_\nu+\lambda d_\nu)$ is minimized.

Step 3: Define $\hat{\boldsymbol{\theta}}_{\nu+1}=\hat{\boldsymbol{\theta}}_\nu+\lambda_\nu d$ and $p_\nu=g_{\nu+1}-g_\nu$

Step 4: Compute the $3n\times 3n$ matrix

$$H_\nu=H_{\nu-1}+\lambda_\nu\frac{d_\nu d_\nu'}{g_\nu' H_{\nu-1} g_\nu}-\frac{H_{\nu-1}p_\nu p_\nu' H_{\nu-1}'}{p_\nu' H_{\nu-1} p_\nu}$$

with $H_0=I_{3n}$, and set $d_{\nu+1}=-H_\nu g_\nu$.

Step 5: Repeat Steps 2-4 until an appropriate convergence condition will be satisfied.

It should be noted that the determination of λ_ν in Step 2 is performed by using linear search techniques such as the golden section method, the Fibonaccian method, or the polynomial interpolation method. Here, we applied a quadratic interpolation method, which is computationally feasible. In fact, let $\phi_\nu(\lambda)$ be a function to be minimized, and find those three points $a<b<c$ which yield $\phi_\nu(a)>\phi_\nu(b)<\phi_\nu(c)$ by a linear search. Then approximate $\phi_\nu(\lambda)$ by the quadratic polynomial passing three points: $(a, \phi_\nu(a))$, $(b, \phi_\nu(b))$ and $(c, \phi_\nu(c))$. Finally, we take λ_ν as

$$\lambda_\nu=\frac{(b^2-c^2)\phi_\nu(a)+(c^2-a^2)\phi_\nu(b)+(a^2-b^2)\phi_\nu(c)}{(b-c)\phi_\nu(a)+(c-a)\phi_\nu(b)+(a-b)\phi_\nu(c)} \tag{4.13}$$

5. Determination of Order of Assumed Model

The problem of determining the system order has received much attention in recent years. For deterministic systems, the order determination problem has already been solved as the minimal realization algorithm[10),11)]. But for stochastic systems, the problem remains to be resolved and requires extensive studies and numerical evidence.

Until now, several methods for determining the order of stochastic systems have been proposed (see references 15), 16), 17)). The algorithms in these papers have shown good results for computer simulation studies. However, for real processes such as power plants, chemical processes, or socio-economic systems, these algorithms do not necessarily give us reasonable results, because the real systems are subject to various nuisances and external disturbances. In other words, it is very difficult to uniquely

determine the system order from available input/output data, because the real data will be contaminated by measurement noises and external disturbances. Thus, when applied to actual data, all the methods to date may require, at least partly, trial and error coupled with a decision by a human being. Therefore, a possible and practical approach to a statistical identification with order determination will be the following. (1) Assuming the order of the assumed model, identify the unknown parameters, and (2) evaluate an appropriate error measure and compute the impulse and step responses. (3) Comparing the results for $n=1, \ldots, n_L$, we take such n_0 as a possible order of the assumed model that yields the least value of error measure. If there is *a priori* information about the structure of the system, it should be also taken into account.

In this paper, we use the following methods.

(1) AIC criterion[18),19)]

The idea that the optimal order of the assumed model is a parameter to be determined was introduced by Akaike who proposed an information theoretic criterion[18),19)]. For any maximum likelihood situation, $AIC(k)$ is defined as

$$AIC(k) = -2\log(\text{maximum likelihood}) + 2k \tag{5. 1}$$

where k is the number of arbitrarily adjustable parameters contained in the assumed model. The order selected is the value of k for which $AIC(k)$ is minimized. Some good results brought about by the use of AIC are presented for both artificial and real data[19)].

(2) Intuitive and experimental methods

(a) polynomial test

This testing method is mainly used for examining whether the polynomials $A(z^{-1})$, $B(z^{-1})$ and $C(z^{-1})$ defined by (2. 9) have common roots or not. It is well known that a necessary and sufficient condition for (F, G) to be controllable is that $A(z^{-1})$ and $B(z^{-1})$ have no common zeros, and that the identifiable transfer function represents only a controllable and observable part of the system. Therefore, if the order of the assumed model is too high, there will be an uncontrollable part in the state vector $\hat{x}(t)$ so that $A(z^{-1})$ and $B(z^{-1})$ will have common zeros. Eliminating the common zeros, we have a lower order model. It is reported that this method has shown good results for computer simulation studies[20),21)].

(b) test for independence of $e(t)$

This method is to test the assumption that the innovation process $\{e(t), t=1, \ldots, N\}$ is a white gaussian noise. The autocorrelation of the innovation process

$$\rho_{ee}(\tau) = \frac{1}{N\hat{\sigma}_N^2}\sum_{t=1}^{N} e(t)e(t+\tau), \quad \tau=0, 1, \ldots \tag{5. 2}$$

must be zero for $\tau \neq 0$, if the assumed model is true. Actually, if $e(t)$ is a gaussian white noise, then $\rho_{ee}(\tau)$ is asymptotically a normal distribution with a mean zero and variance $1/N$[22]. Thus, by computing (5. 2), we can apply a statistical test to evaluate the accuracy of the assumed model.

(c) data checking

A simple but effective method for testing the structure of a model is to use a part of the given data as checking data. First we divide the given data into two parts, D_1 and D_2, where D_1 is the training data, and D_2 is the checking data. Then, the model is identified by D_1 and checked by D_2. In general, as the order of model n increases, the variance $\hat{\sigma}_n^2$ defined by (4. 4) for training data D_1 decreases. However, the variance of the prediction error $y(t)-\hat{y}(t|t-1)$ for checking data D_2 decreases first and then increases when n increases. The probable order of the model is the value of n, for which the variance of the prediction error for D_2 is minimized. In a sense, $AIC(k)$ is a theoretical estimate of the logarithm of the variance of the prediction error for checking data D_2.

6. Overall Identification Algorithm

We summarize the overall algorithm for the ML identification method.

Step 1: Normalize the given data and compute the correlation functions of the input and output.

Step 2: Divide the data into $D_1=\{u(t);y(t), t=1, \ldots, N_1\}$ and $D_2=\{u(t), y(t);\ t=N_1+1, \ldots, N\}$.

Step 3: Set $n=1$, and then go to Step 4.

Step 4: Let $\boldsymbol{c}=0$, and find the LSE's $\boldsymbol{a}$ and $\boldsymbol{b}$ by using the method in section 3.

Step 5: Based on Davidon's algorithm in section 4. 2, find the MLE's $\hat{\boldsymbol{\theta}}$ and $\hat{\sigma}_n^2$.

Step 6: Using the MLE $\hat{\boldsymbol{\theta}}$ obtained in Step 5, compute a step response from (2. 6) and the zeros of $A(z^{-1})$ and $B(z^{-1})$.

Step 7: Compute $AIC(3n)$ by

$$AIC(3n)=N_1 \log\ (\hat{\sigma}_n^2)+2(3n),\quad \dim(\boldsymbol{\theta})=3n \tag{6.1}$$

where N_1 is the number of data in D_1. It may be noted that under the gaussian assumption, (5. 1) is equivalent to (6. 1) up to constant.

Step 8: By using (4. 12), compute the prediction error and its variance for checking data D_2.

Step 9: If $n<n_L$, set $n \rightarrow n+1$, and go to Step 4; if $n=n_L$, go to the next step.

Step 10: Based on the methods in section 5, we estimate the optimal order of the assumed model.

7. Identification of Boiler System

In this section, we show some numerical results of an identification of a once-through type boiler system at the Kainan thermoelectric power plant of the Kansai Electric Company[5),6)]. A block diagram of the boiler system is depicted in Fig. 1, where the rated steam generation is 1380 t/hr. The steam pressure is rated at 255 kg/cm^2 and the steam temperature at 568 °C. The steam is supplied to a turbin generator, rated at 450 Mw of electricity. The amount of the steam flow is controlled by a governor valve, and the quality of the steam is mainly characterized by the steam pressure and temperature. For a large scale thermoelectric power plant like this, the deviations of the steam pressure and temperature should be less than one to two percent of the desired values. To meet the stringent requirements under various load variations and disturbances, a design of the control system with a desired performance is needed, based on the modern control theory. Therefore, the identification of the boiler system is very important for obtaining a state-space model of the system on which the modern control theory is based. Here, the object to be identified is the whole boiler system rather than the boiler itself, because analog type controllers have already been supplied.

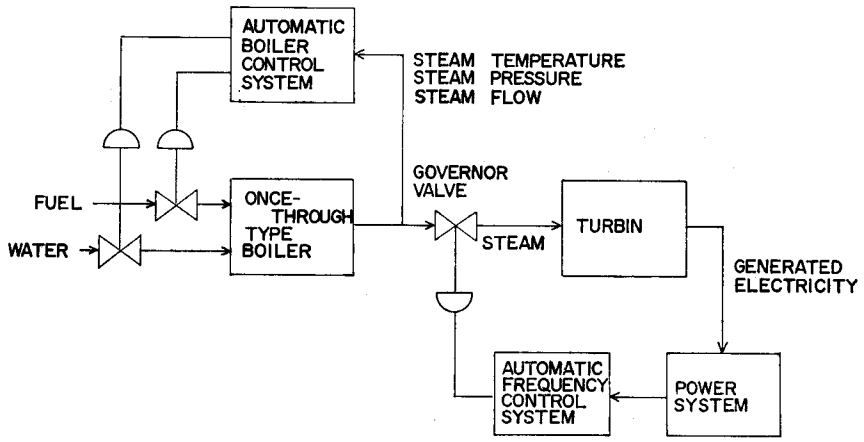

Fig. 1 Block diagram of once-through type boiler system

A part of the logging data of the governor valve position, the steam pressure and the steam temperature is shown in Fig. 2. These data were obtained by intentionally changing the valve position under the normal operating condition, so that the input is the governor valve position and the outputs are the steam pressure and temperature. The total number of the sampling is 1410 and the sampling interval is 5 sec. The normalized correlation functions of the input/output data are shown in Fig. 3, where $u(t)$ is the relative position of the governor valve, and $y_1(t)$ and

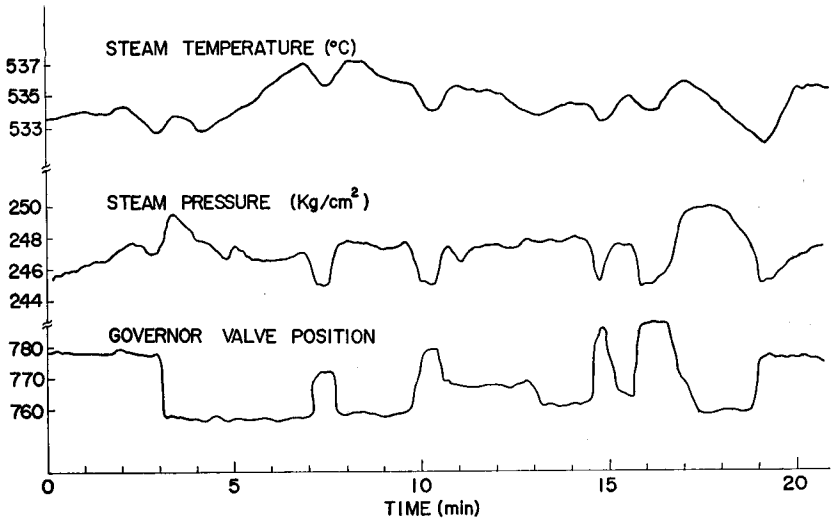

Fig. 2 Input and output data

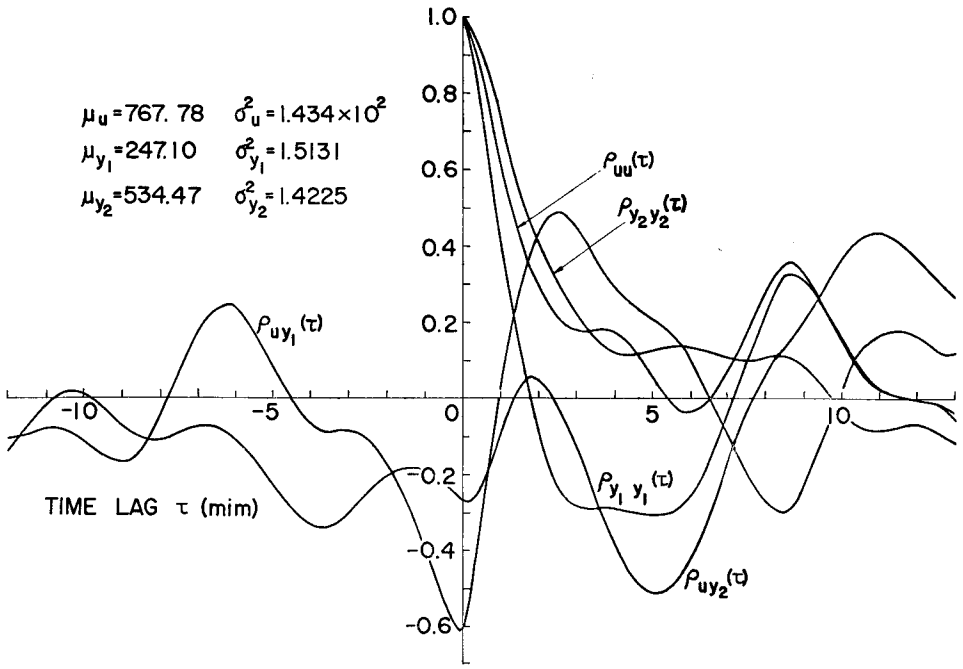

Fig. 3 Normalized correlation functions of input and output data

$y_2(t)$ denote the steam pressure and temperature, respectively. From Figs. 2 and 3, we see that inverse reaction processes exist between input $u(t)$ and outputs $y_1(t)$ and $y_2(t)$, respectively. By opening the valve, the steam pressure goes down, so that a drop of the steam temperature follows, due to the adiabatic expansion. Then the feedback loops work so that both fuel and water are supplied. Hence, the pressure and, then, the temperature will increase again. But the response of the pressure is faster and simpler than that of the temperature. Here, we wish to identify these mechanisms by using the given input/output data. The identification was carried out by using the first 1000 data, and the remainder was used for the data checking or prediction. We will show the results for the governor/pressure relation in section 7. 1, and the governor/temperature relation in section 7. 2.

7.1 Governor/steam pressure

The numerical results are shown in Table 1 and in Fig. 4, where $\hat{\sigma}_n^2$ denotes the minimum value of the variance of the innovation process (or residual) defined by (4. 6). We observe that $\hat{\sigma}_n^2$ does not indicate a significant decrease as n in-

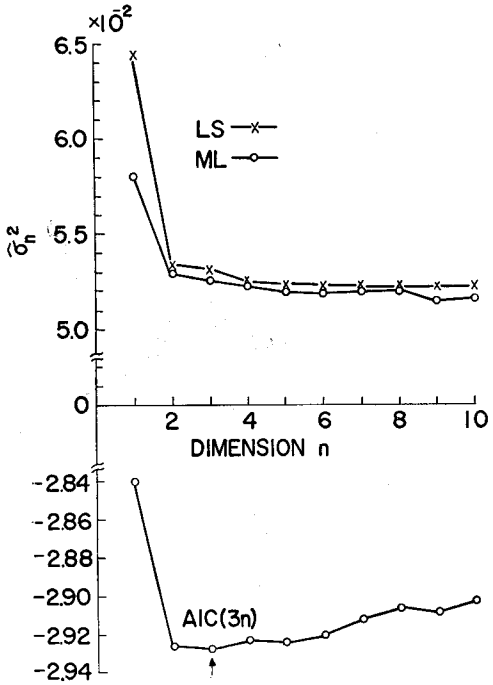

Fig. 4 Numerical result for the steam pressure: Variance of innovation and AIC vs. n

Table 1 Numerical results for the steam pressure

n	LS		ML		PREDICTION
	$\hat{\sigma}_n^2$	FPE	$\hat{\sigma}_n^2$	AIC	
1	6.44623×10^{-2}	6.47209×10^{-2}	5.80444×10^{-2}	−2.84054	4.13054×10^{-2}
2	5.33308	5.37600	5.29345	−2.92667	3.41396
3	5.31096	5.37517	5.25217	−2.92847*	3.40559
4	5.24415	5.32907*	5.23922	−2.92490	3.37654
5	5.23406	5.34033	5.20326	−2.92573	3.35024
6	5.21605	5.34352	5.19629	−2.92100	3.34714*
7	5.21048	5.35950	5.20195	−2.91384	3.36242
8	5.21306	5.38396	5.20566	−2.90704	3.38103
9	5.20412	5.39665	5.15907*	−2.90992	3.39557
10	5.17329*	5.38660	5.16004	−2.90354	3.40940

Identification period: $t=1$ to 1000.
Prediction period: $t=1001$ to 1410.
* denotes the minimum values for $n\leqslant10$.

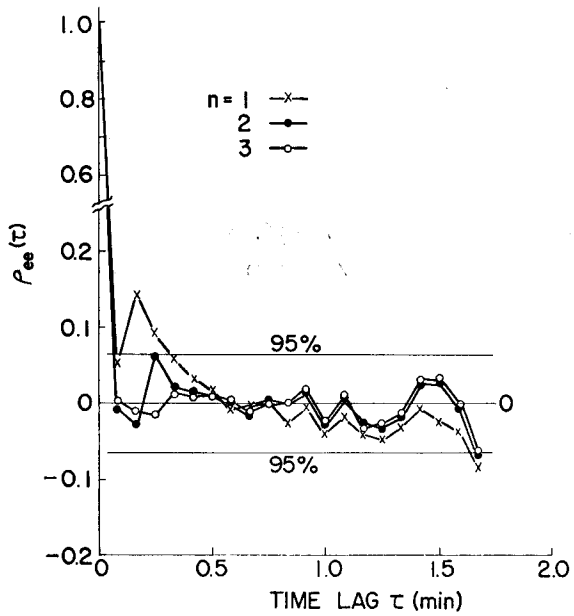

Fig. 5 Autocorrelation functions of $e(t)$ for the steam pressure

creases. The AIC and FPE criteria are also computed[19], and the respective minima are obtained for $n=3$ and $n=4$ as indicated by the asterisk in Table 1. The normalized autocorrelation functions of the innovation process are depicted in Fig. 5. Numerical studies show that the "whiteness" is improved for the larger value of n,

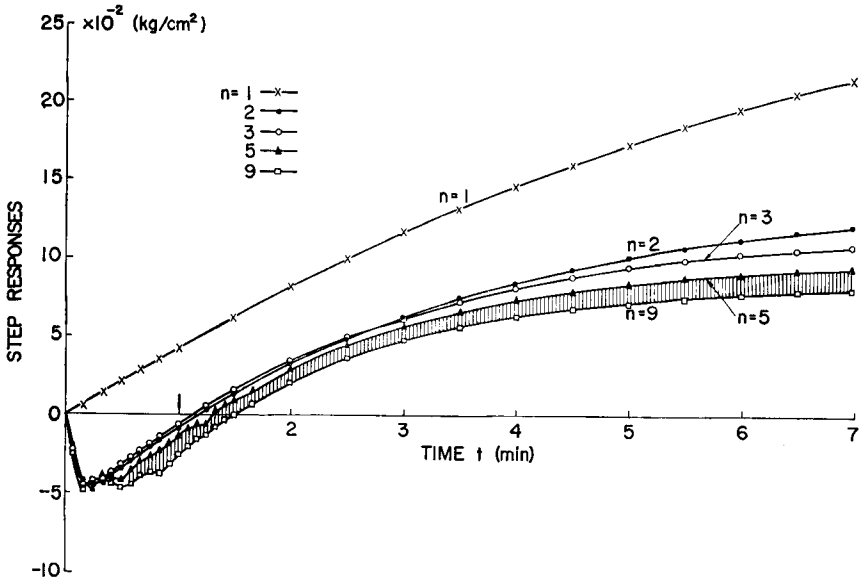

Fig. 6 Step responses of the steam pressure

and that for $n \geqslant 3$ the autocorrelation functions are within the 95% confidence intervals. Thus, the innovation processes are practically whitened for $n \geqslant 3$. Fig. 6 shows the step responses computed from the MLE's of the system parameters. The step responses for $n=4$ to 10 are all between the curves for $n=5$ and $n=9$. The inverse reaction process is clearly identified for $n \geqslant 2$. The models thus obtained are applied to the one-step ahead prediction of the steam pressure for $t=1001$ to 1410. The result is shown in Table 1, where the best MSE performance was obtained for $n=6$. Based on the above numerical studies, we conclude that low order models with $n=2$ or 3 will be sufficient for the description of the governor/pressure relation. Table 2 shows the estimated values of the parameters and zeros of $A(z^{-1})$, $B(z^{-1})$ and $C(z^{-1})$ for $n=2$ and 3.

Table 2 Result of parameter identification for the steam pressure for $n=2$ and 3

n	$A(z^{-1})$		$B(z^{-1})$		$C(z^{-1})$	
	$\boldsymbol{a}$	zeros	$\boldsymbol{b}$	zeros	$\boldsymbol{c}$	zeros
2	$a_1=-1.266$	0.975	$b_1=-0.0396$	1.068	$c_1=-0.215$	0.281
	$a_2=0.275$	0.282	$b_2=0.0423$		$c_2=-0.184$	−0.658
3	$a_1=-1.096$	0.969	$b_1=-0.043$	1.07	$c_1=-0.0887$	−0.469
	$a_2=0.0109$	0.404	$b_2=0.048$	0.044	$c_2=-0.0888$	0.278
	$a_3=0.108$	−0.277	$b_3=0.002$		$c_3=0.0808$	±0.308j

7.2 Governor/steam temperature

Since, as is seen from the correlation functions in Fig. 3, the governor/steam temperature relation is very complex, the identification of the dynamics of the steam temperature is much more difficult. The results of the identification are shown in Table 3 and in Fig. 7. We observe that both the minima of AIC and the MSE performance of prediction are attained at $n=11$, where $n_L=12$. The normalized autocorrelation functions of the innovation process are depicted in Fig. 8. The numerical results show that for $n\geqslant 8$, the normalized autocorrelation functions are within the 95% confidence intervals and that for $n\geqslant 4$, the violations of the 95% confidence intervals are less than 10%. Therefore, the innovation processes are practically whitened for $n=4$. The step responses computed from the MLE's are shown in Fig. 9. We observe that the pattern of the step responses drastically changes between $n=9$ and 10, although all the MSE performances of identification do not change significantly for $n\leqslant 12$. For $n\geqslant 10$, the step responses do not come back to the zero level. However, the step responses obtained for $n\geqslant 10$ correspond, rather well to the results given by Suzuki[5),6)], who estimated the impulse and step responses directly from the given data by using the LS method. Incidentally, AIC and the MSE performance of prediction support the choice $n=11$ as the optimal order of the model. We also made other numerical studies by dividing the given data into two parts: $D_1=\{u(2t-1), y(2t-1);\ t=1,\ldots, 705\}$ and $D_2=\{u(2t), y(2t);\ t=1,\ldots, 705\}$.

Table 3 Numerical results for the steam temperature

n	LS		ML		PREDICTION
	$\hat{\sigma}_n^2$	FPE	$\hat{\sigma}_n^2$	AIC	
1	2.73635×10^{-2}	2.74733×10^{-2}	2.73177×10^{-2}	−3.59421	2.59302×10^{-2}
2	2.68722	2.70885	2.53516	−3.66289	2.35316
3	2.59455	2.62597	2.22038	−3.78944	2.04682
4	2.40103	2.43992	2.15131	−3.81500	2.07865
5	2.25369	2.29945	2.15167	−3.80878	2.06956
6	2.19478	2.24942	2.12956	−3.81304	2.06747
7	2.15181	2.21335	2.12634	−3.80847	2.07415
8	2.13208	2.20198	2.10152	−3.81412	2.07142
9	2.12901	2.20777	2.09801	−3.80969	2.05617
10	2.10076	2.18738	2.05685	−3.82339	2.04407
11	2.09099	2.18612	2.04032*	−3.82533*	2.01844*
12	2.08018*	2.18374*	2.04204	−3.81835	2.02977

Identification period: $t=1$ to 1000.
Prediction period: $t=1001$ to 1410.
* denotes the minimum values for $n\leqslant 12$.

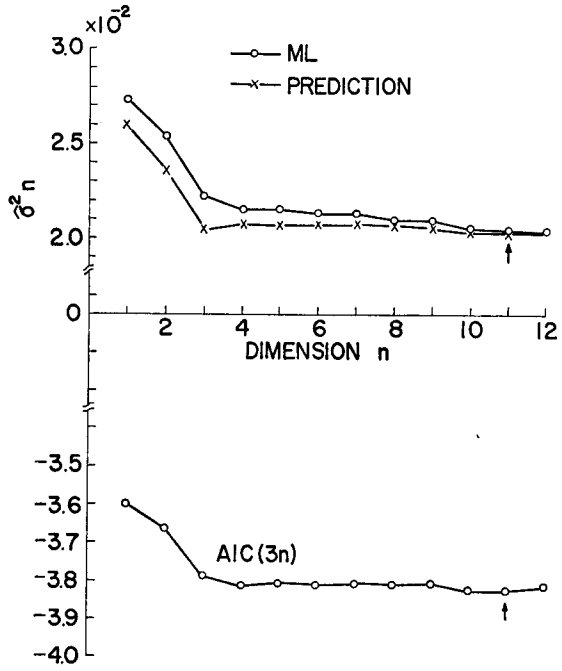

Fig. 7 Numerical result for the steam temperature: MSE of innovation, performance of prediction and AIC vs. n

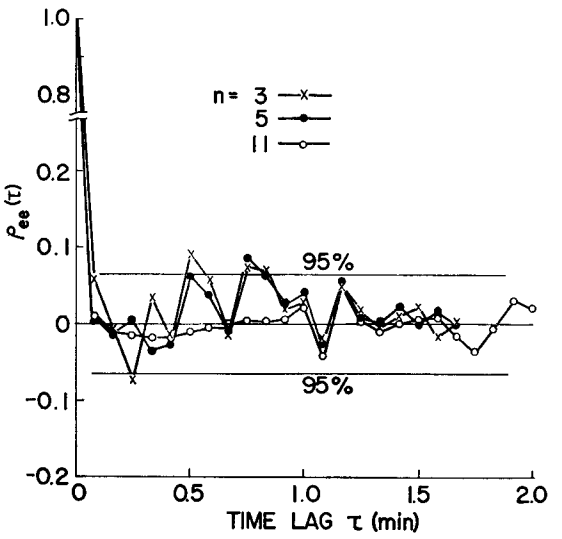

Fig. 8 Autocorrelation functions of $e(t)$ for the steam temperature

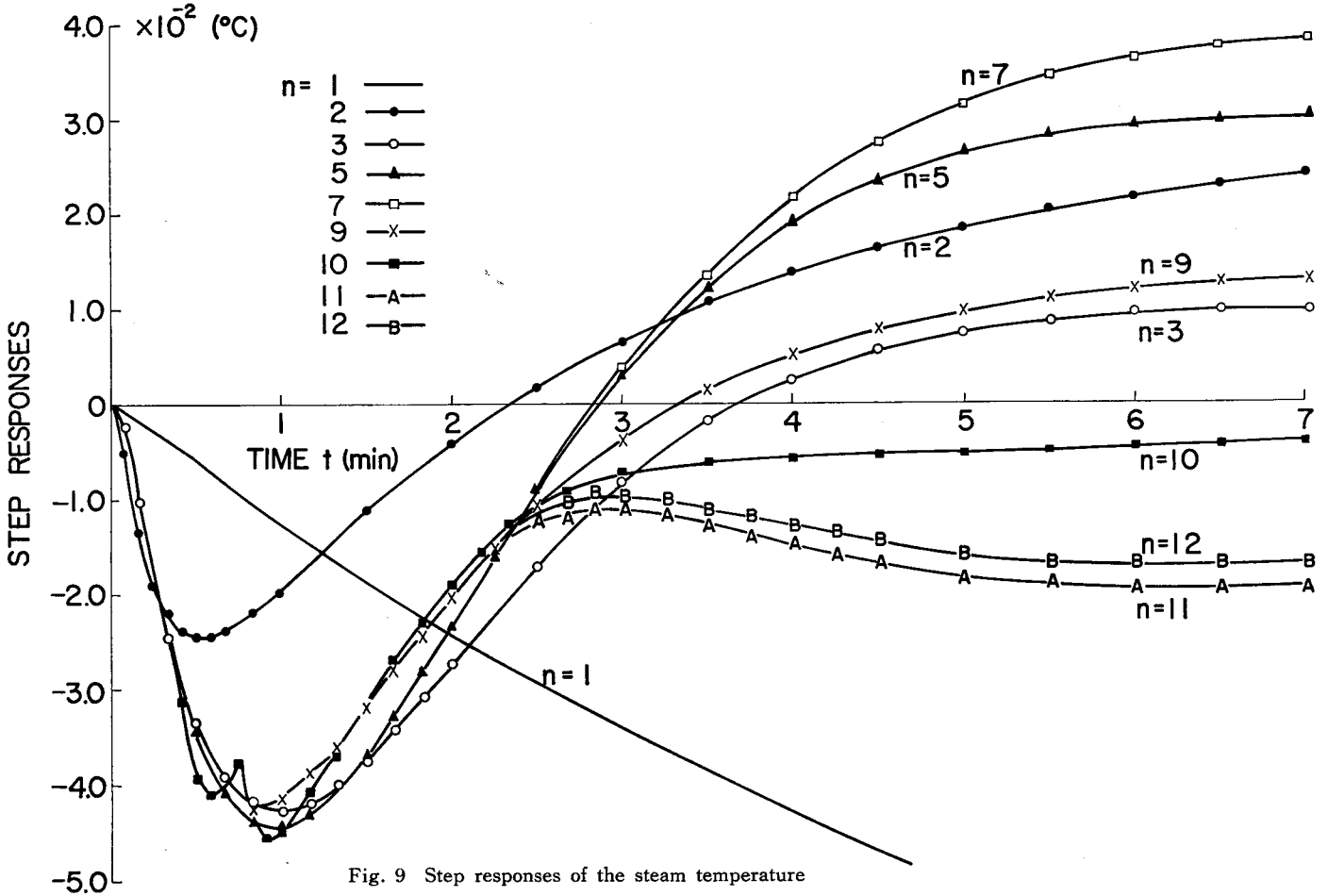

Fig. 9 Step responses of the steam temperature

For Case 1, D_1 is used for training and D_2 for checking. For Case 2, it is vice versa. The results again suggest that the optimal choice of the order is $n=10$ or 11. Indeed, $AIC(3n)$ was minimized at $n=11$ for Case 1 and at $n=8$ for Case 2, respectively. The MSE performance of checking data was minimized at $n=10$ for both cases. Hence, we may conclude that the higher order models with $n=10$ or 11 are superior to the low order models. Besides, the choice of the higher order model agrees with the results by Suzuki.[5),6)] But it should be noted that the accuracy of the individual parameters will be reduced as n increases, because the amount of available data is limited. In other words, the larger the n we assume, the greater is the amount of data required to estimate the parameters in the same accuracy. Therefore, the simpler models with $n=3$ or 4 might be better for deriving the state-space representation, which will be used for the design of optimal controls.

8. Conclusions

We have applied the ML method to a boiler system identification. For the governor/steam pressure relation, the identification has been made very successfully. Numerical results show that low order models with $n=2$ or 3 will be enough for use as the prediction and control. For the governor/steam temperature relation, however, the order of the fitted model is very high compared with that of the model for the pressure. This is due to the fact that the dynamics of the governor/steam temperature relation is quite complex. Further studies will be needed for deriving a simplified state-space representation on which the design of optimal control is based.

All the numerical computations were carried out by using FACOM 230-75 at the Data Processing Center of Kyoto University, Kyoto.

Authors would like to express their thanks to Prof. Y. Suzuki, The University of Osaka, for providing us the data used in this paper.

References

1) K. J. Aström and P. Eykhoff, "System identification-A survey," Automatica, **7**, 123 (1971).
2) K. J. Aström and T. Bohlin, "Numerical identification of linear dynamic systems for normal operating records," Proc. 2nd IFAC Symp. Theory of Self-Adaptive Systems, p. 96 (1965).
3) R. L. Kashyap, "Maximum likelihood identification of stochastic linear systems," IEEE Trans. Auto. Contr., **AC-15**, 25 (1970).
4) R. K. Mehra, "Identification of stochastic linear dynamic system using Kalman filter representation," AIAA J., **9**, 28 (1971).
5) Y. Suzuki and K. Fujii, "Application of least square method to process identification," Trans. SICE, **10**, 563 (1971) (in Japanese).
6) Y. Suzuki, "Application technique in the process control," Preprints Japan United Conference of Electrical Engineers, p. 826 (1972) (in Japanese).
7) T. Kailath, "An innovation approach to least square estimation, Part I: Linear filtering in

additive noise," IEEE Trans. Auto. Contr., **AC-13**, 646 (1968).
8) R. S. Bucy and P. D. Joseph, "Filtering for Stochastic Processes with Applications to Guidance," Interscience, New York (1968).
9) R. E. Kalman, "Mathematical description of linear dynamical systems," SIAM J. Contr., ser. A. **1**, 152 (1963).
10) B. L. Ho and R. E. Kalman, "Effective construction of linear state variable models from input/output functions," Regelungstechnik, **14**, 545 (1966).
11) D. O. Mayne, "Computational procedure for the minimal realization of transfer function matrices," Proc. IEE, **115**, 1363 (1968).
12) G. E. P. Box and G. M. Jenkins, "Time Series Analysis-Forecasting and Control," Holden-Day, San Francisco (1971).
13) J. Kowalik and M. R. Osborne, "Methods for Unconstrained Optimization Problems," American Elsevier, New York (1968).
14) D. G. Luenberger, "Introduction to Linear and Nonlinear Programming," Addison-Wesley (1973).
15) C. M. Woodside, "Estimation of order of linear systems," Preprints 2nd IFAC Symp. Identification and Process Parameter Estimation, Prague (1970).
16) J. Valis, "On-line identification of multivariable linear systems of unknown structure from input/output data," ibid. (1970).
17) J. C. Chow, "On the structure identification and parameter estimation in linear multivariable dynamic systems," Preprints JACC, p. 620 (1972).
18) H. Akaike, "Information theory and an extension of the maximum likelihood principle," Proc. 2nd Int. Symp. Information Theory, p. 267, House of Hungarian Academy of Science (1972).
19) H. Akaike, "A new look at the statistical model identification," IEEE Trans. Auto. Contr., **AC-19**, 716 (1974).
20) H. Unbehauen and B. Göhring, "Application of different statistical tests for the determination of the most accurate order of the model in parameter estimation," Preprints 3rd IFAC Symp. Identification, Hague, p. 917 (1973).
21) A. J. van den Boom and A. W. M. van den Enden, "The determination of the order of process and noise dynamics," ibid. p. 929 (1973).
22) T. W. Anderson, "An Introduction to Multivariate Statistical Analysis," John Wiley, New York (1958).