Special Lecture

An Outlook on the Identification of MIMO Linear and Nonlinear Stochastic Systems

AKIRA OHSUMI*
Professor Emeritus of Kyoto Institute of Technology

In this lecture I would like to talk about my recent works on the modeling and identification of multi-input, multi-output (MIMO) linear and nonlinear stochastic systems. The lecture is divided into the following two parts:

Part I. Identification of linear stochastic systems

Part II. Nonlinear stochastic system identification.

Part I is concerned with the problem of the identification of such physical matrices as mass, damping and stiffness of the structural systems like high-rise building structures. And, in Part II, an approach to the identification problem is stated for a class of black-box systems which are subjected to random disturbances and involve some nonlinearity in the system state and, possibly, input. Postulating the relevant black-box system is given by a linear state-space model with nonlinear terms, the system quadruple matrices and the covariances of both system and observation noises as well as the system order are roughly identified first using the standard 4SID (subspace-based identification) method as preliminaries.

The lecture is based on my recent works [1]-[6]. Themes concerned with Part I and Part II have been inspired by the papers by Bruschetta, et al. [7] and Westrick, et al. [8], respectively.

Part I. Identification of Linear Stochastic Systems

Problem Statement
Consider an n-story high-rise building structure whose dynamics is described by a vector second-order differential equation [4], [9]:

\[ M \ddot{z}(t) + C \dot{z}(t) + Kz(t) = Gw(t), \]

where \( M, C, K \) are the mass, damping and stiffness matrices; \( z(t) = [z_1(t), z_2(t), \cdots, z_n(t)]^T \) and \( w(t) = [w_1(t), w_2(t), \cdots, w_n(t)]^T \) in which \( z(t) \) and \( w(t) \) are the transverse displacement from its equilibrium state and the external random load like wind force at the \( i \)-th floor of \( n \)-story building, respectively; and \( G \) is a known matrix. The matrices \( M \) and \( K \) are symmetric and positive-definite, while \( C \) is symmetric, positive-semi-definite, but all are unknown.

The problem is to identify the structural matrices \((M, C, K)\) from the sampled observation data on \( z(t), \dot{z}(t) \) or \( \ddot{z}(t) \).

The identification of the structural parameters is nowadays one of the most important problems. Especially, in the architectural fields, it is said that the investigation of dynamic characteristics of an existing building structure by field tests is a necessary and important task in the course of checking the construction quality, validating or improving analytical structural models, or conducting damages. The ambient vibration experimental test is considered as the most popular one because of its portability and ease of setup. Up to the present time, the identification problem of the structural parameters has been investigated by many researchers, for instance, refer [7], [9]-[14].

The identification of the structural matrices has been traditionally investigated using the first-order discrete-time state-space model, \( x(k + 1) = Fx(k) + G_0w(k), \) which is transformed equivalently from the continuous-time state-space model for (1), \( \dot{z}(t) = Ax(t) + Bu(t) \)

\[ x(t) = [\dot{z}(t), \ddot{z}(t)]^T \in \mathbb{R}^{2n}. \]

However, since all structural matrices are shut in the matrices \( A \) and \( B \), in order to recover them from the estimates of the matrices \( F \) and \( G_0 \), we need to invert the relations, \( F = e^{Ah} \) and \( G_0 = [\int_0^he^{At}dT]B \) where \( h(>0) \) is a constant sampling interval. This, as it were, central dogma in the subspace method is the architectonic prop of the \( \text{d2c} \) routine in MATLAB. However, the recovery of \( A \) from the estimated \( F \) involves the computation of the logarithm of \( F \). For a sufficiently small \( h \) the condition number of computing \( A \) from \( F \) tends to infinity, because \( F \rightarrow I \) (unit matrix) and \( G_0 \rightarrow 0 \) as \( h \rightarrow 0 \) so that the discrete-time model becomes degenerated. Hence, the discrete-time model obtained as mentioned above becomes ill-conditioned [15]. Furthermore, the structural matrices \((M, C, K)\) are involved in the forms of \(-M^{-1}K\) and \(-M^{-1}C\) in (2,1)- and (2,2)-blocks of the matrix \( A \). This serious fact means that each structural matrix can not be obtained separately. Furthermore,
even by the subspace-based identification method, the matrix $F$ is identified up to within a similarity transformation $T$ as $F_T = T^{-1} F T$; and $T$ is, in general, not determined uniquely. Therefore, we understand that the use of the state-space model is inadequate for the system identification of structural systems. So, the discrete-time second-order dynamic model is exceedingly expected as a continuous-time counterpart (1) in the identification problem of the mechanical or structural systems.

**Discrete Structural Dynamics**

Here, preserving the second-order structural dynamics given by the Langevin-type stochastic equation (1), let us reconsider and reinterpret it.

As far as the external input $w(t)$ is assumed to be white noise, Eq. (1) no longer has its meanings because the “derivatives” $\ddot{z}(t)$ and $\dddot{z}(t)$ do not exist in the ordinary sense [16]. To remedy such an inadequate situation, I have proposed in [1]-[3] to introduce the following two mean derivatives from Nelson’s stochastic mechanics [17]:

$$\begin{align*}
Dz(t) &= \lim_{h \to 0} \mathcal{E} \left\{ \frac{z(t+h) - z(t)}{h} \bigg| z(t) \right\} \quad (2) \\
D_\star z(t) &= \lim_{h \to 0} \mathcal{E} \left\{ \frac{z(t) - z(t-h)}{h} \bigg| z(t) \right\}, \quad (3)
\end{align*}$$

where $\mathcal{E}\{ \cdot | \cdot \}$ denotes the conditional expectation. $Dz(t)$ and $D_\star z(t)$ are both stochastic processes and they are called the mean forward and mean backward derivatives, respectively. The stochastic process $z(t)$ having these two mean derivatives is called the Nelson process.

Let the solution process of a second-order system, $z(t)$, be a Nelson process on the time interval $[0,T]$. For the postiterated system model (1), consider the Lagrangian [1],

$$L(t) = L(z, Dz, D_\star z)$$

$$= \frac{1}{2} \left\{ \frac{1}{2} (Dz)^T M Dz + \frac{1}{2} (D_\star z)^T M D_\star z \right\} - \frac{1}{2} z^T K z,$$  

(4)

which corresponds to ordinary $L(z, \dot{z}) = \frac{1}{2} \dot{z}^T M \dot{z} - \frac{1}{2} \dot{z}^T K \dot{z}$ and let the nonconservative force (Lagrangian force) $f_L(t)$ consist of the dissipation force $f_D(t)$ (which corresponds to $-C \ddot{z}(t)$) and the random external force $Gw(t)$,

$$f_L(t) = -C \frac{1}{2} (Dz(t) + D_\star z(t)) + Gw(t). \quad (5)$$

Then, the Hamilton’s principle,

$$0 = \mathcal{E} \left\{ \int_0^T \delta L[z(t), Dz(t), D_\star z(t)] \, dt \right\} + \int_0^T \delta W(t) \, dt \quad (6)$$

in which the virtual work is $\delta W(t) = f_L^T(t) \delta z(t)$, yields the stochastic equation,

$$M \frac{1}{2} \{DD_\star z(t) + D_\star Dz(t)\} + C \frac{1}{2} \{Dz(t) + D_\star z(t)\} + K \dot{z}(t) = Gw(t). \quad (7)$$

For detail derivation, see [1],[2]. This is the generalized Newton’s equation of motion, so that the expression (1) should be interpreted as (7).

Denote $z_k (k = 0, 1, 2, \cdots)$ as $z(t_k) = z(\cdot k)$, where $h$ is the (constant) sampling interval. By approximating as

$$\begin{align*}
Dz(t_k) &= \frac{z_{k+1} - z_k}{h}, \\
D_\star z(t_k) &= \frac{z_k - z_{k-1}}{h}, \quad (8)
\end{align*}$$

the discrete form of (7) is given as follows:

$$M_d z_{k+2} + C_d z_{k+1} + K_d z_k = Gw_{k+1}, \quad (9)$$

where $w_{k+1} = h w((k+1)h)$ and

$$\begin{align*}
M_d &= \frac{M}{h} + \frac{C}{2}, \\
C_d &= -\frac{2M}{h} + hK, \\
K_d &= \frac{M}{h} - \frac{C}{2} \quad (10)
\end{align*}$$

Equation (9) is the discrete structural dynamics and this is a discrete-time counterpart of continuous-time system (1). The matrices $M_d$,$C_d$ and $K_d$ in (10) have no their meanings of mass, damping and stiffness.

It is a simple exercise to show the relations

$$\begin{align*}
M &= \frac{h}{2} (M_d + K_d), \\
C &= M_d - K_d \\
K &= \frac{1}{h} (M_d + C_d + K_d) \quad (11)
\end{align*}$$

**Identification of Structural Matrices**

The identification of the structural matrices $(M, C, K)$ can be performed by estimating $(M_d, C_d, K_d)$ of the discrete structural dynamics (9) via the relation (11).

From the physical point of view the direct measurement on the displacement $z(t)$ is rather hard to obtain, especially for the building structures. It will be rather the more practical to use accelerometers for detecting movement of the relevant structure. As well-known, the recent measurement technology develops cheap high-performance accelerometers. The principle of the accelerometer is based on the recovery of the displacement information which can be obtained by integrating the required acceleration data twice with the initial condition $z(0) = \dot{z}(0) = 0$. So, in this subsection, assuming the use of accelerometers, we investigate the more general case such that [4, Chap.8]

$$y(t_k) = H \ddot{z}(t_k) + H_v \dot{z}(t_k) + H_d z(t_k) + v(t_k), \quad (12)$$
where \( y(t_k) \in \mathbb{R}^m; \ v(t_k) \) is the measurement noise; and \( H_a, H_v, H_d \) are all \( m \times n \)-matrices.

Applying the approximations in (8) to (12), we get

\[
\hat{y}_k = \hat{H}_a z_{k+1} + \hat{H}_v z_k + \hat{H}_d z_{k-1} + \bar{v}_k,  \tag{13}
\]

where \( \bar{v}_k = h(y(t_k)) = h(y(kh)), \ v_k = hv(kh), \) and

\[
\begin{aligned}
\hat{H}_a &= \frac{H_a}{h},  \\
\hat{H}_v &= -\frac{2H_a}{h} + hH_d,  \\
\hat{H}_d &= \frac{H_a}{h} - \frac{H_v}{h} - 2.
\end{aligned}
\]

Defining the matrices \( \hat{H} = [\hat{H}_a, \hat{H}_v, \hat{H}_d] \in \mathbb{R}^{m \times 3n} \) and \( Z_k = [z_{k+1}^T, z_k^T, z_{k-1}^T]^T \in \mathbb{R}^{3n} \), we obtain the expression

\[
\hat{y}_k = \hat{H}Z_k + \bar{v}_k.  \tag{14}
\]

On the other hand, from (9) we get the expression

\[
A_0 \bar{Z}_k = Gw_k  \tag{15}
\]

where \( A_0 = [M_d, C_d, K_d] \). Eliminating \( Z_k \) from (14) and (15) using the pseudoinverse matrix \( \hat{H}^+ = \hat{H}^T(\hat{H}\hat{H}^T)^{-1} \), there follows that

\[
A_0 \hat{y}_k = Gw_k + e_k,  \tag{16}
\]

where \( \hat{y}_k = \hat{H}^+ \hat{y}_k \) and \( e_k \) is the vector related to the observation noise, \( e_k = A_0 \hat{H}^+ \bar{v}_k \).

Here, define matrices

\[
\begin{aligned}
\hat{Y}_k &= [\hat{y}_k \ \hat{y}_{k+1} \ \cdots \ \hat{y}_{k+N-1}] \in \mathbb{R}^{3m \times N}  \\
W_k &= [w_k \ w_{k+1} \ \cdots \ w_{k+N-1}] \in \mathbb{R}^{m \times N}  \\
E_k &= [e_k \ e_{k+1} \ \cdots \ e_{k+N-1}] \in \mathbb{R}^{n \times N},
\end{aligned}
\]

where \( N \) is a sufficiently large integer. Then, from (16) we have

\[
A_0 \hat{Y}_k = GW_k + E_k
\]

or

\[
E_k = A_0^{-1} - G \begin{bmatrix} \hat{Y}_k \\ W_k \end{bmatrix}.  \tag{17}
\]

Noting that the matrix \( E_k \) is related to the observation noise, the unknown matrix \( A_0 = [M_d, C_d, K_d] \) is determined as a least-squares estimate by minimizing the Frobenius norm \( \|E_k\|_F = \{|\text{tr}(E_k E_k^T)|\}^2 \). To this end, \( LQ \)-factorize the last matrix in the right-hand side of (17) as

\[
\begin{bmatrix} \hat{Y}_k \\ W_k \end{bmatrix} = \begin{bmatrix} L_{11} & O \\ L_{21} & L_{22} \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix}.  \tag{18}
\]

Then, the matrix \( E_k E_k^T \) is evaluated as

\[
E_k E_k^T = (A_0 L_{11} - GL_{21})(A_0 L_{11} - GL_{21})^T + GL_{22} L_{22}^T G^T \geq GL_{22} L_{22}^T G^T,  \tag{19}
\]

where the equality holds if and only if

\[
A_0 L_{11} - GL_{21} = 0
\]

holds. From this, \( [M_d, C_d, K_d] = A_0 \) is determined as

\[
[M_d \ C_d \ K_d] = GL_{21} L_{11}^{-1}.  \tag{20}
\]

Consequently, the structural matrices \( (M, C, K) \) can be obtained from (11) using the relation (20).

For this approach, it should be mentioned that the data on the random wind load \( \{w(t)\} \) is necessarily required as input data. Fortunately, hardware have drastically increased and computational algorithms are equipped recently, so that the acquisition and estimation of data on the wind load will be possible (e.g., [18]).

Part II. Nonlinear Stochastic System Identification

Nonlinear System Identification

It is broadly believed that the identification theory for nonlinear systems is almost as old as that for linear systems. Up to the present time a great deal of studies has been made on the identification or modeling for linear and/or nonlinear gray- or black-box systems. Especially, the subspace-based identification methods have been attracting us in the framework of linear system modeling from the input and output data. However, several authors have extended subspace methods developed for linear time-invariant systems to the identification of Wiener and/or Hammerstein systems (e.g., [19]-[22]). Wiener or Hammerstein model is a linear dynamic model with static nonlinearity at the output or at the input, and the Wiener-Hammerstein model has both nonlinearities.

Recently, I have been investigating the identification problem under the situation that the relevant black-box dynamic system involves nonlinearity in the system state and, possibly, the input. In the system/control community, it is generally recognized that the system identification is the problem to construct a valid mathematical model from a couple of input and output data acquired from the real dynamical system. In this sense, the realization problem falls into the same category because it is viewed as “guessing the equations of motion of a dynamical system from its input/output behavior,” or “setting up a physical model which explains the experimental data” [sic] [23], and the realization theory has been investigated for the linear and time-invariant systems. In 1956, by identification Zadeh [24] meant the problem of identifying a black-box for the determination of its input-output relationship by experimental
means, and further, listing principal problems of system theory, he defined the system identification such that the system identification is the determination, on the basis of input and output data, of a system within a specified class of systems to which the system under test is equivalent [25], [26].

To make my purpose clear, building on his definition, I give the following definition for unknown systems including stochastic systems:

**Definition.** System identification is defined as the procedure of constructing a mathematical model of the input-output relation through state variables from the real input and output observation data. The system model thus constructed should reproduce the equivalent input-output relationship by the real (random) input and observation data in the least-squares sense.

**Posutlated System Model**
Assume that we are given a couple of sampled input and output data, \( \{u(k), y(k)\}_{k=0,1,2,\ldots,N} \) \( u(k) \in \mathbb{R}^d \), \( y(k) \in \mathbb{R}^m \), which has been acquired from the real black-box system. Although the relevant system is a black-box, however assume a priori that the output data comes out as realizations from the following MIMO state-space stochastic systems containing nonlinear term in the state and, possibly, input,

\[
x(k+1) = Ax(k) + Bu(k) + f(x(k), u(k)) + w(k) \quad (21)
\]
\[
y(k) = Cx(k) + Du(k) + v(k),
\]

where \( x(k) \in \mathbb{R}^n \) is the system state; \( w(k) \in \mathbb{R}^d \) and \( v(k) \in \mathbb{R}^m \) are mutually independent white Gaussian noise sequences with zero-means and covariance matrices \( Q \) and \( R \), i.e., \( \mathcal{E}\{w(k)w^T(j)\} = Q \delta_{kj} \), \( \mathcal{E}\{v(k)v^T(j)\} = R \delta_{kj} \) (Kronecker delta); and \( f(\cdot, \cdot) \) is the \( n \)-vector-valued nonlinear function. Assume that the rank \( C = m \), so the system order \( n \) is unknown.

The legitimacy of the postulated system model (21) lies on the following RQ-factorization of the vector consisting of the input \( u(k) \), state \( x(k) \) and the subsequent state \( x(k+1) \),

\[
\begin{bmatrix}
  u(k) \\
  x(k) \\
  x(k+1)
\end{bmatrix} =
\begin{bmatrix}
  R_{11} & 0 & 0 \\
  R_{21} & R_{22} & 0 \\
  R_{31} & R_{32} & R_{33}
\end{bmatrix}
\begin{bmatrix}
  q_1 \\
  q_2 \\
  q_3
\end{bmatrix}.
\]

This yields

\[
x(k+1) = R_{32} R_{22}^{-1} x(k)
\]
\[
+ (R_{31} - R_{32} R_{22}^{-1} R_{21}) R_{11}^{-1} u(k) + R_{33} q_3
\]

in which the last term in the right-hand side will not be independent of the input and state. Thus, the postulated model (21) can be reasonably acceptable.

Then, the problem is to identify the system order \( n \), the system quadruplet \((A, B, C, D)\), the covariances \((Q, R)\), and to model the nonlinear function \( f(x(k), u(k)) \) (within a similarity transformation) from the acquired input and output data \( \{u(k), y(k)\}_{k=0,1,2,\ldots,N} \).

**Pre-processing**
In order to determine the system order \( n \) and the quadruplet \((A, B, C, D)\), recall that if the nonlinearity is so weak (i.e., \( f(x(k), u(k)) \equiv 0 \)), we can treat the system as linear one. So, to do this, regard temporarily the nonlinear term \( f(x(k), u(k)) \) in (21) as a random disturbance \( f(k) \) with zero-mean and independent of \( w(k) \) to get the temporary model,

\[
\begin{cases}
  x(k+1) = Ax(k) + Bu(k) + f(k) + w(k) \\
  y(k) = Cx(k) + Du(k) + v(k).
\end{cases}
\]

Using the standard 4SID method, we get the (similarity-transformed) estimates \((\hat{A}, \hat{B}, \hat{C}, \hat{D})\), \((\hat{Q}_{f+w}, \hat{R})\) and the system order \( \hat{n} \), where \( \hat{Q}_{f+w} \) is the covariance matrix of the combined noise \( f(k) + w(k) \). Although these estimates are rough ones, we use the estimate of \( Q \) to be \( \hat{Q} = \rho \hat{Q}_{f+w} \) where \( \rho \) is a design parameter such that \( 0 < \rho < 1 \).

**Proposed Method of Modeling**
The proposed method for identifying the nonlinear black-box consists of the three steps: (i) Extraction of nonlinear effect from the observation data \( \{y(k)\} \), (ii) Estimation of nonlinearity using the Kalman filter, and (iii) Modeling of nonlinearity using auxiliary basis functions.

(i) **Extraction of Nonlinear Effect from Observation Data:** Since we have postulated that the system model is given in the form (21), it will be natural to consider that the output \( y(k) \) consists of “linear” and “nonlinear” portions. So, in order to extract the nonlinear portion hidden in observation data conspicuous, we diminish the “linear” portion in \( \{y(k)\} \) as thoroughly as possible. To realize this, consider the auxiliary linear system,

\[
\begin{cases}
  x_L(k+1) = \hat{A} x_L(k) + \hat{B} u(k) \\
  y_L(k) = \hat{C} x_L(k) + \hat{D} u(k)
\end{cases}
\]

with a proper initial value \( x_L(0) \), where \( x_L(k) \in \mathbb{R}^d \), \( y_L(k) \in \mathbb{R}^m \), \( u(k) \) is the same input as that in (21) and (22), and \((\hat{A}, \hat{B}, \hat{C}, \hat{D})\) are rough and ready estimates as mentioned above. Then, the data defined by

\[
y_F(k) = y(k) - y_L(k)
\]

is considered to be mainly due to the “nonlinear” portion and the random disturbances \( w(k) \) and \( v(k) \).

(ii) **Estimation of Nonlinearity:** In order to estimate the nonlinearity, provide the following system whose output is \( y_F(k) \):

\[
\begin{cases}
  x_F(k+1) = \hat{A} x_F(k) + \hat{f}(k) + \hat{w}(k) \\
  y_F(k) = \hat{C} x_F(k) + \hat{v}(k),
\end{cases}
\]

where \( x_F(k) \in \mathbb{R}^n \); \( \hat{w}(k) \in \mathbb{R}^d \) and \( \hat{v}(k) \in \mathbb{R}^m \) are zero-mean system and observation noises having estimated
covariances \( \tilde{Q} \) and \( \tilde{R} \), respectively; the term \( \tilde{f}(k) \) is the similarity-transformed \( f(k) \) and this should be determined in such a way that the output of the system (26) is exactly the data \( \{y_F(k)\} \). Obviously, this implies that the problem to determine \( \{f(k)\} \) is just the estimation problem, that is, the inverse problem, based on the nonlinearity-extracted observation data \( \{y_F(k)\} \).

Assume that the unknown \( \tilde{f}(k) \) obeys the dynamics,

\[
\tilde{f}(k+1) = \tilde{f}(k) + \eta(k),
\]

where \( \eta(k) \in R^n \) is a zero-mean white Gaussian sequence with known covariance \( Q_\eta(\geq 0) \). Augmenting (27) with \( x_F(k) \)-process in (26) and constructing the Kalman filter for the augmented state vector, we get the estimate process for \( \tilde{f}(k) \) as

\[
\begin{aligned}
\tilde{f}(k+1|k) = & \tilde{f}(k|k-1) + \hat{P}_{12} \hat{C}^T (\hat{C} \hat{P}_{11} \hat{C}^T + \hat{R})^{-1} \\
& \cdot \{y_F(k) - \hat{C} \tilde{x}_F(k|k-1)\}
\end{aligned}
\]

\[
\begin{aligned}
\tilde{x}_F(k+1|k) = & \hat{A} \tilde{x}_F(k|k-1) + \tilde{f}(k|k-1) \\
& + (\hat{A} \hat{P}_{11} + \hat{P}_{21}) \hat{C}^T (\hat{C} \hat{P}_{11} \hat{C}^T + \hat{R})^{-1} \\
& \cdot \{y_F(k) - \hat{C} \tilde{x}_F(k|k-1)\},
\end{aligned}
\]

where \( \tilde{f}(k|k-1) \) and \( \tilde{x}_F(k|k-1) \) are Kalman estimates of \( f(k) \) and \( x_F(k) \), respectively; and \( \hat{P}_{1j} (\hat{P}_{11} = \hat{P}_{11}^T) \) is the \( ij \)-component of the estimation error covariance matrix \( \hat{P} \in R^{2n \times 2n} \) which is the positive-definite solution to the algebraic Riccati equation:

\[
\hat{P} = A_0 [\hat{P} - \hat{P} C_0^T (C_0 \hat{P} C_0^T + \hat{R})^{-1} C_0 \hat{P}] A_0^T + Q_0,
\]

where

\[
\begin{bmatrix}
\hat{A} & I_n \\
0 & I_n
\end{bmatrix} \in R^{2n \times 2n}, \quad C_0 = [\hat{C} \ 0] \in R^{m \times 2n}
\]

\( Q_0 = \) block diag \( \{\tilde{Q}, Q_\eta\} \) \( R^{2n \times 2n} \).

In (28), the initial conditions are set properly as \( \tilde{f}(0|0) = \tilde{f}_0 \) and \( \tilde{x}_F(0|0) = \tilde{x}_F(0) \).

The obtained \( \tilde{f}(k|k-1) \) is used as the estimate of \( f(k) \), viz.,

\[
\tilde{f}(k) = \tilde{f}(k|k-1).
\]

(iii) **Modeling of Nonlinearity via Basis Functions:**

Thus, the nonlinearity \( \{\tilde{f}(k)\} \) has been estimated by processing the Kalman filter. However, it should be noted that the sequence \( \{\tilde{f}(k|k-1)\} \) is given only as a numerical data. This awkward situation is fatal rather than unfavorable to the system modeling.

Given the (numerically) estimated nonlinearity \( \{\tilde{f}(k)\} \), consider the system

\[
\Sigma_D: \begin{cases}
    x_D(k+1) = \hat{A} x_D(k) + \hat{B} u(k) + \tilde{f}(k) \\
    y_D(k) = \hat{C} x_D(k) + \hat{D} u(k),
\end{cases}
\]

and, as opposed to this, let

\[
\Sigma_M: \begin{cases}
    x_M(k+1) = \hat{A} x_M(k) + \hat{B} u(k) + \mu(k, \varphi; \theta) \\
    y_M(k) = \hat{C} x_M(k) + \hat{D} u(k),
\end{cases}
\]

where \( x_D(k), x_M(k) \in R^n; \ y_D(k), y_M(k) \in R^m; \ u(k) \in R^p \) is the same input as that in (21) and (22); \( \mu(k, \varphi; \theta) \) is an \( n \)-vector-valued function to model \( f(k) \) in which \( \varphi = \varphi(x_M, u) \) is the user-defined basis function such as (possibly anticipatory) polynomials, power series, etc., and \( \theta = \theta(k) \) is the parameter (vector) to be determined.

There are many possibilities to set the vector-valued function \( \mu(k, \varphi; \theta) \). For instance, let

\[
\mu(k, \varphi; \theta) = \sum_{i=1}^p \theta_i(k) \varphi_i[x_M(k), u(k)],
\]

where \( \{\varphi_i(\cdot, \cdot)\} \) are \( n \)-vector-valued basis functions and \( \{\theta_i(k)\} \) are scalar parameters to be determined; so that \( \mu(k, \varphi; \theta) \) is expressed as

\[
\mu(k, \varphi; \theta) = \Phi[x_M(k), u(k)] \theta(k)
\]

in which \( \Phi(\cdot, \cdot) = [\varphi_1(\cdot, \cdot), \cdots, \varphi_p(\cdot, \cdot)] \in R^{n \times p} \) and \( \theta(k) = [\theta_1(k), \cdots, \theta_p(k)]^T \in R^p \).

Then, the parameter vector \( \theta(k) \) is determined sequentially in such a way that the squared norm of the discrepancy between two outputs of \( \Sigma_D \) and \( \Sigma_M \) becomes minimal. Define

\[
J_k(\theta) = \| y_D(k) - y_M(k) \|^2_W \quad (k = 1, 2, \cdots),
\]

where the notation \( \|u\|^2_W = (y^T W y)^{1/2} \) is the Euclidean norm, and \( W \) is a symmetric and positive-definite weight matrix. Under the condition \( x_D(0) \equiv x_M(0) \), the minimization of \( J_k(\theta) \) with respect to \( \theta(k) \) gives

\[
\theta(k) = \left\{ \Phi^T[x_M(k), u(k)] \hat{C}^T W \hat{C} \Phi[x_M(k), u(k)] \right\}^{-1} \cdot \Phi^T[x_M(k), u(k)] \hat{C}^T W \hat{C} \lambda_k,
\]

where \( \lambda_k \) is the sequentially computable \( n \)-vector (given below). For the derivation of (36), see [5], [6].

Consequently, (by dropping the subscript \( M \) on \( x_M(k) \) and \( y_M(k) \) to write), the identified mathematical model is given as follows:

\[
\begin{cases}
    x(k+1) = \hat{A} x(k) + \hat{B} u(k) \\
    \quad + \Phi[x(k), u(k)] \theta(k) + \hat{w}(k) \\
    y(k) = \hat{C} x(k) + \hat{D} u(k) + \hat{v}(k)
\end{cases}
\]

with
Selection of Basis Functions

Although several basis functions in modeling nonlinear systems have been introduced in the literature [27], [28] and [29, Chap. 5], the problem how to select the basis functions \{\varphi_i(x, u)\} is quite ad hoc, and there seems no general way how to select them. In the identified model (37) with (38), it should be noted that the computation of the matrix inverse is necessary for obtaining \(\theta(k)\). A crucial point is how we can guarantee the existence of the inverse of \(\Phi^T(x(k), u(k)) (\hat{C}^TW\hat{C}) \Phi(x(k), u(k))\)

Here, I propose a recommendable way. Let \(p = \tilde{n}\) and all components of the \(\tilde{n}\)-vector-valued function \(\varphi_i(\cdot, \cdot)\) except its ith one be zero, i.e.,

\[
\varphi_i(\cdot, \cdot) = [0 \cdots 0 \varphi_{i\tilde{n}}(\cdot, \cdot) 0 \cdots 0]^T
\]

Then, \(\theta(k) \in \mathbb{R}^\tilde{n}\), and \(\Phi(x, u) \in \mathbb{R}^{\tilde{n} \times \tilde{n}}\) becomes diagonal,

\[
\Phi(x, u) = \begin{bmatrix}
\varphi_{1\tilde{n}}(x, u) & \cdots & 0 \\
0 & \ddots & 0 \\
0 & \cdots & \varphi_{\tilde{n}\tilde{n}}(x, u)
\end{bmatrix}
\]

Since \(\det[\Phi^T(x, u) (\hat{C}^TW\hat{C}) \Phi(x, u)] = [\det \Phi(x, u)]^2 \cdot \det(\hat{C}^TW\hat{C})\), we know that the inverse of \(\Phi^T(x, u) (\hat{C}^TW\hat{C}) \Phi(x, u)\) exists as far as \(\det(\hat{C}^TW\hat{C}) \neq 0\) and \(\det \Phi(x, u) \neq 0\) hold for any \((x(k), u(k))\). For example, we may select \(\varphi_{i\tilde{n}}(x, u)\) as a Fourier series-like function,

\[
\varphi_{i\tilde{n}}(x(k), u(k)) = \sum_{\nu=1}^{q} \sin \nu x_i(k) + \cos \nu x_i(k)
\]

for all \(i\), where \(q\) is a properly set integer.

The identified system model thus established has been tested to show whether it works well for several simulation data generated by the dynamics of a quarter-car, and the simulation experiments have been reported in [6]. The algorithm tracks the simulated data satisfactorily well.

On-line Use of Identified System Model

Thus, the mathematical model of the relevant nonlinear system has been successfully established. It should be noted that a set of large amount of input and output data, \(\{u(k), y(k)\}_{k=0,1,2,\ldots,N}\), where \(N\) is arbitrary but sufficiently large, is required in advance. So, in this sense, the modeling is done by identifying nonlinear system off-line.

Once we get the identified model (37) with (38), then how can we work it on-line? In order to use the identified model for the system design or control-oriented purpose, the output sequence \(\{y(k)\}\) should be generated on-line for any input sequence. To make the problem clear, suppose that we have obtained the output \(\{y(j), j = 0, 1, 2, \ldots, k\}\) up to the present time step \(k\) for the inputs \(\{u(j), j = 0, 1, 2, \ldots, k\}\). Then, how can the future output \(y(k+1)\) be produced?

The algorithm is stated as follows:

**Step 0:** [Prerequisite] As the major premise, let us assume that the estimates \(\hat{\theta}\), \((\hat{A}, \hat{B}, \hat{C}, \hat{D})\) and \((\hat{Q}, \hat{R})\) have been already obtained using a standard 4SID method (as mentioned in Subsection ‘Pre-processing’).

**Step 1:** [Preassignment of initial conditions] The initial condition for the system state \(x(k)\) should be estimated beforehand, otherwise set it arbitrarily as \(x(0) = x_0\). For the auxiliary systems, preassign their initial values: \(x_i(0) = x_{L0}, \hat{f}(0) = \hat{f}_0, \hat{x}_F(0) = \hat{x}_{F0}\), \(\lambda_0 = \hat{f}(0) (\equiv \hat{f}_0)\); and obtain the block matrices \(\hat{P}_{ij}\) by solving the algebraic Riccati equation (29).

**Step 2:** [Computation of \(f(k)\) and \(\lambda_k\)] At the present time step \(j = k\), using the one-step behind observation \(y(k-1)\), compute the nonlinearity-extracted data \(y_r(k-1) = y(k-1) - y_L(k-1)\), where \(y_L(k-1)\) is computed by (24):

\[
\begin{align*}
\begin{bmatrix}
y_{L}(k-1) \\
x_{L}(k-1)
\end{bmatrix} &= \begin{bmatrix}
\hat{C}x_{L}(k-1) + \hat{D}u_{k-1} \\
\hat{A}x_{L}(k-2) + \hat{B}u_{k-2}
\end{bmatrix}.
\end{align*}
\]

**Step 3:** [Computation of \(f(k) = \hat{f}(k-1)\)] Using \(y_r(k-1)\) obtained in Step 2 in the Kalman filter (28), the (similarity-transformed) nonlinearity \(\hat{f}(k)\) is computed as its output \(\hat{f}(k|k-1)\), and then compute

\[
\lambda_k = \hat{A}\{\lambda_{k-1} - \Phi(x(k-1), u(k-1)) \theta(k-1)\} + \hat{f}(k).
\]

**Step 4:** [Computation of \(\theta(k)\)] Thus, the adjust-parameters vector \(\theta(k)\) is computed by the first equation in (38) using \(u(k)\) and \(\lambda_k\).

**Step 5:** [Generation of future output \(y(k+1)\)] Therefore, the future output \(y(k+1)\) is obtained with new input \(u(k+1)\) as

\[
\begin{align*}
y(k+1) &= \hat{C}x(k+1) + \hat{D}u(k+1) + \hat{e}(k+1) \\
x(k+1) &= \hat{A}x(k) + \hat{B}u(k) + \Phi(x(k), u(k)) \theta(k) + \hat{w}(k), \\
& \quad \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ x(0) = x_0.
\end{align*}
\]

Repeating Steps 2-5, the output sequence \(\{y(k)\}_{k=1,2,\ldots}\) can be computed on-line.

If we use the recursive identification algorithm developed by my colleagues and myself [30], [31] in Step 0, then all steps including Step 0 are treated on-line, and can mutatis mutandis cope with even though the case where the black-box system is time-varying. The recursive algorithm is based on the recursion of \(LQ\)-factorization of data matrix consisting of the inputs and
Concluding Remarks

In this lecture, I have attempted to give my personal outlook on the recently developing system identification of MIMO linear and nonlinear stochastic systems.

In Part I, for the linear system, different from the viewpoint of the subspace-based identification method, an approach to identify the structural matrices of building structure has been proposed using the second-order discrete-time model which is a counterpart of the familiar Langevin-type continuous-time second-order dynamics subjected to white noise as disturbance.

In Part II, my recent work on the identification of black-box system with some nonlinearity has been reviewed, extending the identified model to be applicable for on-line use.

I believe that the modeling/identification of real, linear or nonlinear, system is truly one of fundamental disciplines in modern engineering/science, and therefore, undoubtedly it will become more and more important in the future. Specifically, the art of nonlinear system modeling is vast and quite ad hoc, so that it still remains open to all people in the system/control community. I want to close the talk by quoting from [32] which cites [33]: ‘identification of nonlinear system’ is like a statement about ‘non-elephant zoology.’

References


Akira Ohsumi was born in Kyoto, Japan, on 1943. Received the B.Sc. and M.Sc. degrees in Engineering from Kyoto Institute of Technology in 1967 and 1969, respectively, and the Ph.D. degree from Kyoto University in 1976. Since 1969 he was with Kyoto Institute of Technology as Research Assistant, Associate Professor, and then as Professor of Department of Mechanical & System Engineering since 1989, retired in 2007. After retired, he was a Professor at the Department of Computer Science & System Engineering, University of Miyazaki, Miyazaki, Japan, during 2007-2009. In 1983 he was a Visiting Scholar at the Division of Applied Sciences, Harvard University, Cambridge, Massachusetts, U.S.A. His research areas are stochastic processes, estimation, system identification, stochastic control, distributed parameter systems, and signal detection in random noise.

He has published over 160 archival and conference papers and the author of *Introduction to Stochastic Systems* (Asakura Publ., 2002), *Linear Systems Control Theory* (Morikita Publ. 2003), and *Control Theory of Structural Systems* (Morikita Pul., 2013), and a co-author of several text books. He was the President of the Institute of System, Control and Information Engineers (ISCIIE) during 2005-2006. He is a recipient of the Paper Award in 1982, 2007, and the Best Book Award in 2005 all from the Society of Instruments and Control Engineers (SICE), the 1987 Shimadzu Fellowship for Promotion of Science and Technology, the Second Yoshihito Kurose Award from the Japan Microwave Tunneling Association in 1991, and the Paper Award and, at the same time, The Yoshihumi Sunahara Award in 2001 and 2009 from ISCIE.

He is now Professor Emeritus of Kyoto Institute of Technology, and he is a SICE Fellow, ISCIE Honorary Member, and IEEE Life Member.