Goodwin, G. C., and K. S. Sin, 1984. *Adaptive Filtering, Prediction and Control*, Prentice Hall, Englewood Cliffs, NJ.

Gunckel, T. L., and G. F. Franklin, 1968. A general solution for linear sampled-data control, *Trans. ASME: Journal of Basic Engineering*, Ser. D, *85*: 197–203.

Joseph, P. D., and J. Tou, 1961. On linear control theory, *Trans. AIEE*, *80*: 193–196.

Lewis, F. L., 1986. *Optimal Estimation with an Introduction to Stochastic Control Theory*, Wiley, New York.

Stengel, R. F., 1986. *Stochastic Optimal Control: Theory and Applications*, Wiley, New York.

Wonham, W. M., 1968. On the separation theorem of stochastic control, *SIAM Journal of Control*, 6(2): 312–326.

13
Introduction to System Identification

13.1 INTRODUCTION

In the fields of physical science and engineering we are often interested in the development of a mathematical model of some physical phenomenon in order to make analytical predictions about the behavior of the system. In control applications we are often interested in modeling a physical plant, which we wish to control, in order to predict the effect of control efforts and disturbances on that plant.

Often, the system model can be obtained by application of physical principles to the region of space designated as the system and obtaining the governing dynamic equations. Often these equations result from force, mass, energy, or momentum balances or the governing principles of electromechanical systems (i.e., Newton's, Kirchhoff's, Lenz's, and Faraday's laws). Sometimes the model cannot be obtained from physical arguments because of unknown chemical reactions, unknown boundary conditions on the physical processes, or the extreme complexity of the process. In these cases we must resort to the experimental method to develop a system model. This method is applied to the system of Fig. 13.1, wherein measurements of system stimuli and responses are made and the dynamic nature of the system is deduced from the relations between responses and stimuli. This process, known as *system identification* or *characterization*, is the epitome of the scientific method. There are many techniques for accomplishing this process and only two are given here; however, it is interesting to note that the more popular techniques are (1) random or pseudorandom

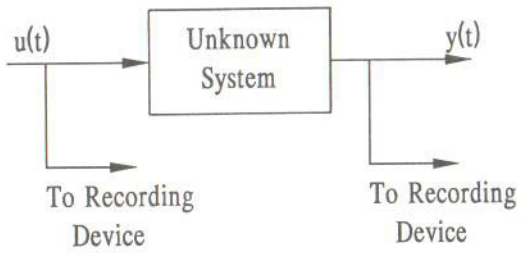

Figure 13.1. System to be identified.

excitation; (2) impulse excitation; (3) step excitation; and (4) sinusoidal excitation with a sweep through the frequencies.

A technique is given here that uses an arbitrary deterministic form of input and a system model of fixed structure, leaving only the parameters in that structure to be identified. Before we continue with our discussion of the identification algorithms, we uncover a few of the ideas of the least-squares method of data fitting.

13.2 LEAST-SQUARES TECHNIQUE

We will be interested in the identification of systems with constant parameters which will form the parameter vector $\boldsymbol{\theta}$, which will be related to the measurements by the linear vector relation

$$y_i = \mathbf{d}_i^T \boldsymbol{\theta} \qquad i = 1, 2, \ldots, n \tag{13.2.1}$$

where $\mathbf{d}_i$ is a p-vector of data corresponding to measurement y_i and $\boldsymbol{\theta}$ is a p-vector of parameters which we wish to find based on the data y_i and $\mathbf{d}_i^T$. Let us define our estimate of the parameter vector as $\hat{\boldsymbol{\theta}}$ which may differ from the actual parameter values of $\boldsymbol{\theta}$. The set of equations then are

$$y_i = \mathbf{d}_i^T \hat{\boldsymbol{\theta}} + e_i \qquad i = 1, 2, \ldots, n \tag{13.2.2}$$

where e_i is the error induced because we have employed the estimate of $\boldsymbol{\theta}$. The set of n equations of (13.2.2) can be written as the vector equation

$$\mathbf{y} = \mathbf{D}\hat{\boldsymbol{\theta}} + \mathbf{e} \tag{13.2.3}$$

where $\mathbf{D}$ is the $n \times p$ data matrix with an ith row of $\mathbf{d}_i^T$ and $\mathbf{y}$ is commonly called the $n \times 1$ measurement vector and $\mathbf{e}$ is an $n \times 1$ vector of errors.

There are many practical problems that we can put into the form of

(13.2.3), one of which is the curve-fitting problem, which is discussed in Examples 13.1 and 13.2.

Example 13.1. An example of least-squares curve fitting involves the set of experimental data shown in Fig. 13.2, which we wish to fit with a straight line, or

$$y_i \cong mx_i + b \qquad i = 1, 2, \ldots, n \tag{13.2.4}$$

or

$$y_i = mx_i + b + e_i \qquad i = 1, 2, \ldots, n \tag{13.2.5}$$

In this case the parameter vector is two-dimensional and contains the slope and y intercept of the straight line, or

$$\boldsymbol{\theta} = \begin{bmatrix} m \\ b \end{bmatrix} \tag{13.2.6}$$

The n relations of (13.2.5) can be written in matrix form as

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} \hat{m} \\ \hat{b} \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \tag{13.2.7}$$

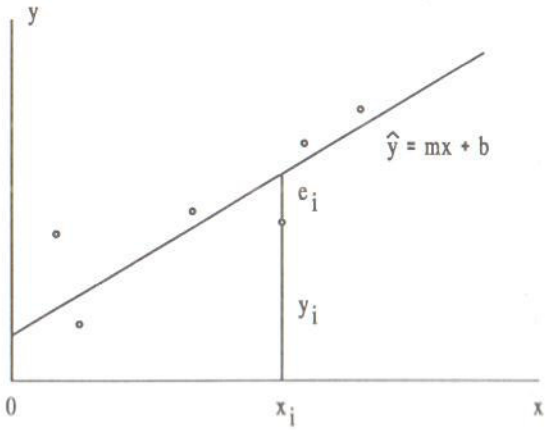

Figure 13.2. Data points to be fit with a straight line.

In this example the number of measurements (n) must be greater than the number of parameters (p), which in this case is two.

The technique of least squares is not restricted to models that are linear in the independent variable x, but the models *must* be linear in the parameters, as is this one in m and b. Polynomial curve fits are common, but note that these fits are linear in the coefficients which are the parameters. Similarly, Fourier series representations are linear in the parameters, which in that case are the unknown Fourier coefficients.

Let us return now to our general discussion of least squares. The model of the estimator is

$$\mathbf{y} = \mathbf{D}\boldsymbol{\theta} \tag{13.2.8}$$

while in reality

$$\mathbf{y} = \mathbf{D}\hat{\boldsymbol{\theta}} + \mathbf{e} \tag{13.2.9}$$

where the $\mathbf{e}$ vector represents the errors in the estimates of the points y_i. We want to minimize the sum of the squares of the errors or minimize the scalar function

$$J = \mathbf{e}^T\mathbf{e} = \sum_{i=1}^{n} e_i^2 \tag{13.2.10}$$

We see that the error vector is the difference of $\mathbf{y}$ and the predicted value of $\mathbf{y}$ (namely $\mathbf{D}\hat{\boldsymbol{\theta}}$), or

$$J = (\mathbf{y} - \mathbf{D}\hat{\boldsymbol{\theta}})^T(\mathbf{y} - \mathbf{D}\hat{\boldsymbol{\theta}}) \tag{13.2.11}$$

Now to minimize J with respect to $\hat{\boldsymbol{\theta}}$ we differentiate the function J with respect to the vector $\hat{\boldsymbol{\theta}}$ and we get

$$\frac{\partial J}{\partial \hat{\boldsymbol{\theta}}} = (\mathbf{y} - \mathbf{D}\hat{\boldsymbol{\theta}})^T(-\mathbf{D}) + (-\mathbf{D}^T)(\mathbf{y} - \mathbf{D}\hat{\boldsymbol{\theta}}) = 0 \tag{13.2.12}$$

These two terms are the same, and hence it is sufficient that one of them be zero and hence

$$\mathbf{D}^T\mathbf{D}\hat{\boldsymbol{\theta}} = \mathbf{D}^T\mathbf{y} \tag{13.2.13}$$

This set of equations is often referred to as the *normal equations*, which must be solved for the parameter estimate vector $\hat{\boldsymbol{\theta}}$. One way would be to invert the matrix $\mathbf{D}^T\mathbf{D}$ to give

$$\hat{\boldsymbol{\theta}} = (\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T\mathbf{y} \tag{13.2.14}$$

The matrix $\mathbf{D}^T\mathbf{D}$ is $p \times p$ and must be inverted. If $p > n$, then the rank of $\mathbf{D}^T\mathbf{D}$ is less than p and it will be singular and hence not invertible.

Clearly, the number of data points (n) taken should be larger than the number of parameters to be estimated (p). When dealing with a large number of data points it is possible that numerical problems will arise in the inversion required in (13.2.14). It might be easier to solve (13.12.13) by an iterative technique such as relaxation.

Example 13.2. Consider the least-squares curve fit of a straight line of the form

$$y = mx + b$$

to the following data and shown in Fig. 13.3.

x	0	2	4
y	1	2	4

The parameter vector is

$$\boldsymbol{\theta} = \begin{bmatrix} m \\ b \end{bmatrix}$$

According to relation (13.2.7), the data matrix and measurement vector are

$$\mathbf{D} = \begin{bmatrix} 0 & 1 \\ 2 & 1 \\ 4 & 1 \end{bmatrix} \qquad \mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$$

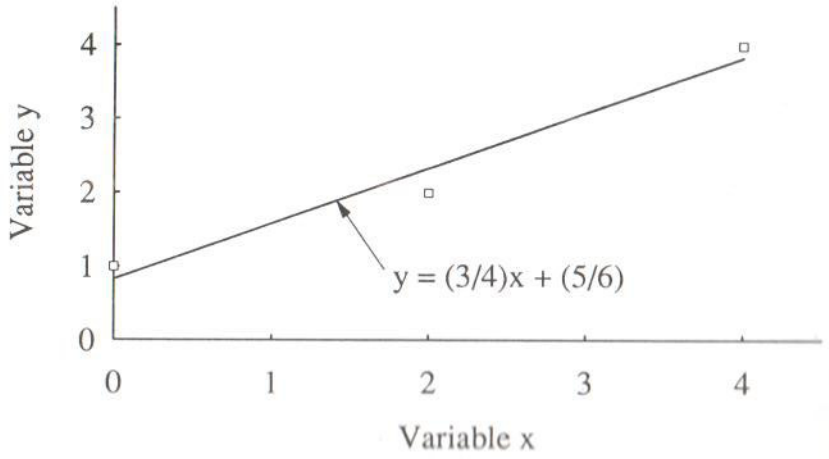

Figure 13.3. Data points to be fit by least-squares straight line.

We shall use relation (13.2.14) to solve the problem

$$\mathbf{D}^T\mathbf{D} = \begin{bmatrix} 20 & 6 \\ 6 & 3 \end{bmatrix} \qquad \mathbf{D}^T\mathbf{y} = \begin{bmatrix} 20 \\ 7 \end{bmatrix}$$

Using relation (13.2.14), we get the parameter vector estimate

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} \frac{3}{4} \\ \frac{5}{6} \end{bmatrix}$$

So the best-fit least-square straight line is

$$y = \frac{3}{4}x + \frac{5}{6}$$

and this is plotted as the straight line in Fig. 13.3.

13.3 TRANSFER FUNCTION ESTIMATION USING LEAST SQUARES

We shall assume that we have a discrete-data system and that we have measured stimulus and response sequences $u(k)$ and $y(k)$, respectively. We must assume a priori a form of the z-domain transfer function. It is sometimes not an easy task, so one must assume several forms for the transfer function and see which one gives the minimum aggregate error. An example will best illustrate this technique, which then can easily be applied to other transfer functions.

Assume that the process involved is governed by the transfer function

$$G(z) = \frac{Y(z)}{U(z)} = \frac{b_{n-1}z^{n-1} + \cdots + b_1 z + b_0}{z^n - a_{n-1}z^{n-1} - \cdots - a_1 z - a_0} \tag{13.3.1}$$

The equivalent difference equation is

$$\begin{aligned} y(k) = {} & a_{n-1}y(k-1) + \cdots + a_0 y(k-n) \\ & + b_{n-1}u(k-1) + \cdots + b_0 u(k-n) \end{aligned} \tag{13.3.2}$$

We shall assume that $y(k)$ and $u(k)$ are zero for negative indices k.

Let us define the kth data vector to be

$$\begin{aligned} \mathbf{d}(k) = {} & [y(k-1)\ y(k-2) \cdots \\ & y(k-n)\ u(k-1) \cdots u(k-n)]^T \end{aligned} \tag{13.3.3}$$

and the parameter vector to be

$$\boldsymbol{\theta} = [a_{n-1} \cdots a_0 \quad b_{n-1} \cdots b_0]^T \tag{13.3.4}$$

Then the output at time k can be written as the inner product of the parameter vector and the data vector

$$y(k) = \mathbf{d}^T(k)\boldsymbol{\theta} \tag{13.3.5}$$

If an estimate of $\boldsymbol{\theta}$, say $\hat{\boldsymbol{\theta}}$, is used, there will be an error in the prediction of $y(k)$, or

$$y(k) = \mathbf{d}^T(k)\hat{\boldsymbol{\theta}} + e(k) \tag{13.3.6}$$

If we start k at n and end it at N, the resulting set of equations is

$$\begin{aligned} y(n) &= \mathbf{d}^T(n)\hat{\boldsymbol{\theta}} + e(n) \\ &\vdots \qquad\qquad \vdots \\ y(N) &= \mathbf{d}^T(N)\hat{\boldsymbol{\theta}} + e(N) \end{aligned} \tag{13.3.7}$$

Define the vector of output data $\mathbf{y}(N)$ as

$$\mathbf{y}(N) = [y(n) \cdots y(N)]^T \tag{13.3.8}$$

Define the data matrix as

$$\mathbf{D}(N) = [\mathbf{d}(n) \cdots \mathbf{d}(N)]^T = \begin{bmatrix} \mathbf{d}^T(n) \\ \vdots \\ \mathbf{d}^T(N) \end{bmatrix} \tag{13.3.9}$$

Also define the error vector

$$\mathbf{e}(N) = [e(n) \cdots e(N)]^T \tag{13.3.10}$$

If we intend to estimate the parameter vector, the system of equations (13.3.7) can be written as

$$\mathbf{y}(N) = \mathbf{D}(N)\hat{\boldsymbol{\theta}} + \mathbf{e}(N) \tag{13.3.11}$$

If we want to estimate the parameter vector at time NT, we want to minimize the sum of the squares of the errors, or

$$J = \sum_{k=n}^{N} e^2(k) = \mathbf{e}^T(N)\mathbf{e}(N) \tag{13.3.12}$$

If we solve relation (13.3.11) for $\mathbf{e}(N)$ and substitute into the relation for J,

$$J = (\mathbf{y}(N) - \mathbf{D}(N)\hat{\boldsymbol{\theta}})^T (\mathbf{y}(N) - \mathbf{D}(N)\hat{\boldsymbol{\theta}}) \tag{13.3.13}$$

If we differentiate this product with respect to the vector $\hat{\boldsymbol{\theta}}$, we get

$$\frac{\partial J}{\partial \hat{\boldsymbol{\theta}}} = -\mathbf{y}(N)\mathbf{D}(N) + \hat{\boldsymbol{\theta}}^T\mathbf{D}^T(N)\mathbf{D}(N) = 0 \qquad (13.3.14)$$

or

$$\mathbf{D}^T(N)\mathbf{D}(N)\hat{\boldsymbol{\theta}} = \mathbf{D}^T(N)\mathbf{y}(N) \qquad (13.3.15)$$

This set of equations is called the *normal equations* from least-squares theory, and much attention has been given to the solution of them. We may write the solution symbolically as

$$\hat{\boldsymbol{\theta}} = (\mathbf{D}^T(N)\mathbf{D}(N)^{-1}\mathbf{D}^T(N)\mathbf{y}(N) \qquad (13.3.16)$$

although it is not very likely that we would actually invert the matrix to get the parameter estimate.

Example 13.3. We shall consider the identification of the thermal system of Example 3.3. Since the idea would be to control this system with the digital computer in the loop, it stands to reason that we could use the same digital computer to identify a discrete-time model for the plant, the analog-to-digital (A/D) converter, and the digital-to-analog (D/A) converter. The forcing sequence, in this case a unit step, is generated by the computer, and the A/D converter is used to sample the system output. The forcing and response sequences are used in expression (13.3.16) to perform the parameter identification. The physical situation is shown in Fig. 13.4. The driving sequence applied is a unit-step function applied at $k = 0$, so the response sequence $y(k)$ recorded by the A/D converter and the driving sequence are given in Table 13.1. For this problem the data matrix is

$$\mathbf{D} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0.01953 & 0 & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1.21561 & 1.16435 & 1 & 1 \end{bmatrix}$$

and the measurement matrix is

$$\mathbf{y} = \begin{bmatrix} 0.01953 \\ 0.07567 \\ \vdots \\ 1.26199 \end{bmatrix}$$

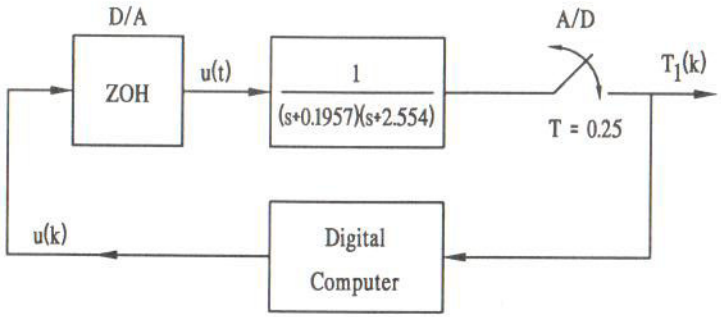

Figure 13.4. Digital computer system identification.

Using relation (13.3.16) the estimates of the parameters are

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} 1.47023 \\ -0.49018 \\ 0.01953 \\ 0.02776 \end{bmatrix}$$

The actual discrete-time parameters found in Example 3.3 are

$$\boldsymbol{\theta} = \begin{bmatrix} 1.48 \\ -0.5026 \\ 0.025 \\ 0.0204 \end{bmatrix}$$

Table 13.1. Discrete Data for Thermal System Identification with $T = 0.25$ s

k	$y(k)$	$u(k)$	k	$y(k)$	$u(k)$
0	0	1	11	0.76159	1
1	0.01953	1	12	0.82750	1
2	0.07567	1	13	0.89096	1
3	0.14890	1	14	0.95199	1
4	0.22945	1	15	1.00812	1
5	0.31001	1	16	1.06427	1
6	0.39300	1	17	1.11553	1
7	0.47111	1	18	1.16435	1
8	0.54922	1	19	1.21561	1
9	0.62245	1	20	1.26199	1
10	0.69324	1			

The sum of the squares of the errors can be evaluated using relation (13.5.13) to give

$$J = \mathbf{y}^T(N)(\mathbf{y}(N) - \mathbf{D}(N)\hat{\boldsymbol{\theta}}) = 4.63 \times 10^{-5}$$

Using these parameters, a digital simulation was run, and in plots of response data no difference could be detected between the data of Table 13.1 and those given in the digital simulation

The experiment was run several times, and each time data similar to those in Table 13.1, but not exactly the same, were obtained and each time yielded a somewhat different parameter vector estimate. The technique developed in this section is sometimes referred to as a batch mode identification technique because the estimation process used a block of previously collected data.

13.4 WEIGHTED LEAST SQUARES

Often it is important to weight the data nearest our current time more heavily than those in the past, so it is sometimes wise to incorporate a weighting matrix into the problem, or

$$J = \mathbf{e}^T(N)\mathbf{W}(N)\mathbf{e}^T(N) \tag{13.4.1}$$

A popular choice for the elements of the weighting matrix is to make it diagonal, such that

$$\mathbf{W}(N) = \begin{bmatrix} a\gamma^{N-n} & & \bigcirc \\ & \ddots & \\ \bigcirc & & a\gamma \\ & & & a \end{bmatrix} \tag{13.4.2}$$

so the ith diagonal element is $W_i(N) = a\gamma^{N-n+1-i}$.

Minimizing J yields the following normal equations which involve the weighting matrix:

$$\hat{\boldsymbol{\theta}}(N) = (\mathbf{D}^T(N)\mathbf{W}(N)\mathbf{D}(N)^{-1}\mathbf{D}^T(N)\mathbf{W}(N)\mathbf{y}(N) \tag{13.4.3}$$

If $\mathbf{W}(N) = \mathbf{I}$, we have the ordinary unweighted least squares of relation (13.3.16).

13.5 RECURSIVE LEAST SQUARES

In Section 13.3 we explored a batch mode identification technique and in this section we explore a technique with which we may reevaluate the parameter estimate with each newly obtained response sample.

Let us look back at the weighted least-squares technique, which we shall examine for the $(N + 1)$st estimate of the parameter vector

$$\begin{aligned}\hat{\boldsymbol{\theta}}(N + 1) = {} & [\mathbf{D}^T(N + 1)\mathbf{W}(N + 1) \\ & \times \mathbf{D}(N + 1)]^{-1}\mathbf{D}^T(N + 1)\mathbf{W}(N + 1) \\ & \times \mathbf{y}(N + 1)\end{aligned} \tag{13.5.1}$$

Let us now examine the matrix product in the square brackets, which may be written out as

$$\mathbf{D}^T(N + 1)\mathbf{W}(N + 1)\mathbf{D}(N + 1) = \sum_{k=n}^{N+1} \mathbf{d}^T(k)\mathbf{W}_k(N + 1)\mathbf{d}^T(k)$$

and if we choose exponential weighting,

$$= \sum_{k=n}^{N+1} \mathbf{d}(k)a\gamma^{N+1-k}\mathbf{d}^T(k) \tag{13.5.2}$$

This can be partitioned into two terms:

$$\begin{aligned}&\mathbf{D}^T(N + 1)\mathbf{W}(N + 1)\mathbf{D}(N + 1) \\ &\quad = \sum_{k=n}^{N} \mathbf{d}(k)a\gamma^{N-k}\mathbf{d}^T(k) + \mathbf{d}(N + 1)a\mathbf{d}^T(N + 1) \\ &\quad = \gamma\mathbf{D}^T(N)\mathbf{W}(N)\mathbf{D}(N) + \mathbf{d}(N + 1)a\mathbf{d}^T(N + 1)\end{aligned} \tag{13.5.3}$$

Define a $2n \times 2n$ matrix $\mathbf{P}(N + 1)$ as

$$\mathbf{P}(N + 1) = [\mathbf{D}^T(N + 1)\mathbf{W}(N + 1)\mathbf{D}(N + 1)]^{-1} \tag{13.5.4}$$

From relation (13.5.3) we see that

$$\mathbf{P}(N + 1) = [\gamma\mathbf{P}^{-1}(N) + \mathbf{d}(N + 1)a\mathbf{d}^T(N + 1)]^{-1} \tag{13.5.5}$$

Now employ the matrix inversion lemma, which is proven in Appendix C, to yield

$$\begin{aligned}\mathbf{P}(N + 1) = {} & \frac{\mathbf{P}(N)}{\gamma} - \frac{\mathbf{P}(N)}{\gamma}\mathbf{d}(N + 1) \\ & \left(\frac{1}{a} + \mathbf{d}^T(N + 1)\frac{\mathbf{P}(N)}{\gamma}\mathbf{d}(N + 1)\right)^{-1} \\ & \mathbf{d}^T(N + 1)\frac{\mathbf{P}(N)}{\gamma}\end{aligned} \tag{13.5.6}$$

We also need to examine the quantity $\mathbf{D}^T(N + 1)\mathbf{W}(N + 1)\mathbf{y}(N + 1)$,

which may also be partitioned as

$$\begin{aligned}\mathbf{D}^T(N+1)\mathbf{W}(N+1)\mathbf{y}(N+1)&\\ &= \sum_{k=n}^{N} \mathbf{d}(k)a\gamma^{N+1-k}\mathbf{y}(N) + \mathbf{d}(N+1)ay(N+1)\\ &= \gamma\mathbf{D}^T(N)\mathbf{W}(N)\mathbf{y}(N) + \mathbf{d}(N+1)ay(N+1) \qquad (13.5.7)\end{aligned}$$

From the weighted least-squares result (13.4.3), we get

$$\begin{aligned}\hat{\boldsymbol{\theta}}(N+1) = &\left[\frac{\mathbf{P}(N)}{\gamma} - \frac{\mathbf{P}(N)}{\gamma}\mathbf{d}(N+1)\right.\\ &\left(\frac{1}{a} + \mathbf{d}^T(N+1)\frac{\mathbf{P}(N)}{\gamma}\mathbf{d}(N+1)\right)^{-1} \qquad (13.5.8)\\ &\left.\mathbf{d}^T(N+1)\frac{\mathbf{P}(N)}{\gamma}\right]\\ &\times [\gamma\mathbf{D}(N)\mathbf{W}(N)\mathbf{y}(N) + \mathbf{d}(N+1)ay(N+1)]\end{aligned}$$

When we multiply this out and note that

$$\mathbf{P}(N)\mathbf{D}^T(N)\mathbf{W}(N)\mathbf{y}(N) = \hat{\boldsymbol{\theta}}(N) \qquad (13.5.9)$$

we get

$$\begin{aligned}&\hat{\boldsymbol{\theta}}(N+1)\\ &= \hat{\boldsymbol{\theta}}(N) + \frac{\mathbf{P}(N)}{\gamma}\mathbf{d}(N+1)ay(N+1) - \frac{\mathbf{P}(N)}{\gamma}\\ &\left(\frac{1}{a} + \mathbf{d}(N+1)\frac{\mathbf{P}(N)}{\gamma}\mathbf{d}^T(N+1)\right)^{-1}\mathbf{d}^T(N+1)\hat{\boldsymbol{\theta}}(N)\\ &- \frac{\mathbf{P}(N)}{\gamma}\mathbf{d}(N+1)\left(\frac{1}{a} + \mathbf{d}^T(N+1)\frac{\mathbf{P}(N)}{\gamma}\mathbf{d}(N+1)\right)^{-1}\\ &\times \mathbf{d}^T(N+1)\frac{\mathbf{P}(N)}{\gamma}\mathbf{d}(N+1)ay(N+1) \qquad (13.5.10)\end{aligned}$$

Now insert the following indentity between the $\mathbf{d}(N+1)$ and the a in the second term:

$$\left(\frac{1}{a} + \mathbf{d}^T(N+1)\frac{\mathbf{P}(N)}{\gamma}\mathbf{d}(N+1)\right)^{-1}\left(\frac{1}{a} + \mathbf{d}^T(N+1)\frac{\mathbf{P}(N)}{\gamma}\mathbf{d}(N+1)\right) \qquad (13.5.11)$$

After expanding we get the following estimate of the parameter vector at time $(N+1)T$:

$$\hat{\boldsymbol{\theta}}(N+1) = \hat{\boldsymbol{\theta}}(N) + \mathbf{k}(N+1)\big(y(N+1) - \mathbf{d}^T(N+1)\hat{\boldsymbol{\theta}}(N)\big) \qquad (13.5.12)$$

where the "gain" vector $\mathbf{k}(N+1)$ is defined as

$$\mathbf{k}(N+1) = \frac{\mathbf{P}(N)}{\gamma}\mathbf{d}(N+1)\left(\frac{1}{a} + \mathbf{d}^T(N+1)\frac{\mathbf{P}(N)}{\gamma}\mathbf{d}(N+1)\right)^{-1} \qquad (13.5.13)$$

One can note the similarity between this and the structure of the Kalman filter from Chapter 11. Note that the parameter estimator given by (13.5.12) is recursive and that it is driven by the error in the predicted system output at time $(N+1)T$.

This algorithm is useful for real-time identification and hence can be used to monitor systems for which the parameters drift slowly in time. An adaptive control strategy can thus be employed which not only updates the values of the estimates of the system parameters but also updates the control law either in the form of a conventional compensator or a state-space controller composed of a state estimator and estimated state feedback control law. The algorithm can be started in many ways, and hence we make the following suggestions. Let

$$\mathbf{P}(n) = \alpha\mathbf{I}$$

where α is large and positive. If we assume that the first n data points are noiseless, the exact solution for $\hat{\boldsymbol{\theta}}(n)$ would be

$$\hat{\boldsymbol{\theta}}(n) = \begin{bmatrix}\mathbf{d}^T(1)\\ \mathbf{d}^T(2)\\ \vdots\\ \mathbf{d}^T(n)\end{bmatrix}^{-1}\begin{bmatrix}y(1)\\ \vdots\\ y(n)\end{bmatrix} \qquad (13.5.14)$$

and this would be one way of starting the recursive process.

The algorithm for recursive identification is then summarized as

1. Choose $\mathbf{P}(k)$, γ, a, and $\hat{\boldsymbol{\theta}}(k)$ to start the algorithm.
2. Compute $\mathbf{k}(k+1)$ from

$$\mathbf{k}(k+1) = \frac{\mathbf{P}(k)}{\gamma}\mathbf{d}(k+1)\left(\frac{1}{a} + \mathbf{d}^T(k+1)\frac{\mathbf{P}(k)}{\gamma}\mathbf{d}(k+1)\right)^{-1}$$

3. Estimate the parameters at time $k+1$ or

$$\hat{\boldsymbol{\theta}}(k+1) = \hat{\boldsymbol{\theta}}(k) + \mathbf{k}(k)\big(y(k+1) - \mathbf{d}^T(k+1)\hat{\boldsymbol{\theta}}(k)\big)$$

4. Compute $\mathbf{P}(k+1)$ from

$$\mathbf{P}(k+1) = \frac{1}{\gamma}[\mathbf{I} - \mathbf{k}(k)\mathbf{d}^T(k+1)]\mathbf{P}(k)$$

5. Let $k = k + 1$.
6. Go to step 2.

Now we shall illustrate the technique with the data of Example 13.2.

Example 13.4. Let us now consider the same data used in the batch mode identification performed in Example 13.3. Let us identify the system recursively and choose an initial estimate of the parameter vector to be

$$\hat{\boldsymbol{\theta}}(0) = \begin{bmatrix} 1 \\ -0.5 \\ 0.02 \\ 0.02 \end{bmatrix}$$

and the initial value of $\mathbf{P}(0)$ as

$$\mathbf{P}(0) = 10^4\mathbf{I}$$

The weighting factors are chosen such the weighting of the past data is exponential, or

$$a = 0.2$$
$$\gamma = (1 - a) = 0.8$$

The algorithm as outlined was programmed and runs very well on a personal computer. In fact, the algorithm is so efficient that it can run in real time in parallel with a control law such that the processor can both control and identify the system at the same time. Plots of the identified parameter histories are given in Fig. 13.5.

13.6 EFFECTS OF NOISE

Often in an engineering environment the input and output measurements may be contaminated by random noise. For low levels of noise the methods just presented may produce excellent estimates of the system parameters,

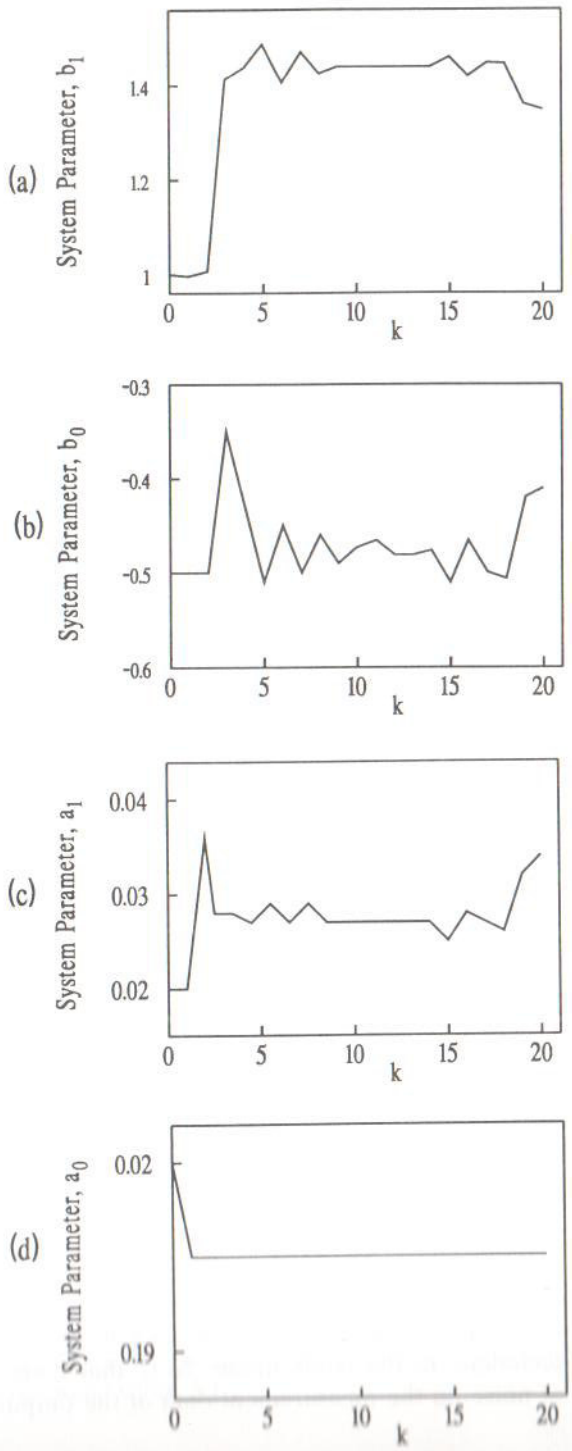

Figure 13.5. (a) Parameter estimate history for $\theta_1 = b_1$; (b) parameter estimate history for $\theta_2 = b_0$; (c) parameter estimate history for $\theta_3 = a_1$; (d) parameter estimate history for $\theta_4 = a_0$.

however, with larger levels of noise may require modification of the weighting factors γ and a. For still larger levels of noise there may be no hope for estimation of the parameters.

Another source of random noise in the measured data is that the system to be identified is also driven by random disturbances at some point. In this case the noise at the output is not purely random, but rather, statistically related (correlated) from sample to sample. There exists a method that will find the least-squares estimate of the parameters, which accounts for the random excitation of the system and thus improved parameter estimates result. This technique is called the *stochastic least-squares method* (Franklin et al., 1990).

One way to handle the random noise in the measurements is to perform the experiment by which the data are gathered a number of times under identical conditions. The data from each of these independent experiments are then averaged at each given time step; thus the random fluctuations in the responses tend to average to zero and the signal-to-noise ratio is greatly increased. In fact, it can be shown that the standard deviation of the noise on the averaged response decreases by a factor of the reciprocal of the square root of the number of records taken.

13.7. SUMMARY

In this chapter we have addressed the subject of transfer-function parameter identification. The least-squares batch method was derived and an example given. The recursive least-squares method was also derived. The recursive technique is useful because it can be employed for simultaneous real-time identification and control.

PROBLEMS

13.1. Given the input–output data given in the table below and a system model of the form

$$y(k + 1) = ay(k) + bu(k)$$

identify the parameters a and b using the data given using the least-squares technique in the batch mode. Note that there is a small amount of noise on the measurement data of the output.

k	$u(k)$	$y(k)$	k	$u(k)$	$y(k)$
0	1	0	6	1	1.843
1	1	0.490	7	1	1.976
2	1	0.905	8	1	2.091
3	1	1.218	9	1	2.165
4	1	1.476	10	1	2.240
5	1	1.681			

13.2. Given the data of Problem 13.1, identify the system employing the recursive least-squares technique. Choose $a = 0.2$ and $\gamma = 0.8$ and $\mathbf{P}(0) = 10^4\mathbf{I}$.

13.3. Consider the sprung-mass positioning problem posed in Problem 3.9. A unit-step force input is employed and the responses are tabulated below.

k	$y(k)$	k	$y(k)$	k	$y(k)$
0	0	7	0.0296	14	0.0472
1	0.0215	8	0.0510	15	0.0584
2	0.0639	9	0.0686	16	0.0615
3	0.0877	10	0.0685	17	0.0555
4	0.0770	11	0.0539	18	0.0469
5	0.0465	12	0.0387	19	0.0435
6	0.0254	13	0.0375	20	0.0470

Identify a second-order model for this system using the least-squares technique.

13.4. For the sprung-mass problem of Problem 13.3., identify the system recursively using $a = 0.2$, $\gamma = 0.8$, and $\mathbf{P}(0) = 10^4\mathbf{I}$.

REFERENCES

Anderson, B. D. O., and J. B. Moore, 1979. *Optimal Filtering*, Prentice Hall, Englewood Cliffs, NJ.

Franklin, G. F., J. D. Powell, and M. L. Workman, 1990. *Digital Control of Dynamic Systems*, 2nd Ed., Addison-Wesley, Reading, MA.

Goodwin, G. C., and K. S. Sin, 1984. *Adaptive Filtering, Prediction and Control*, Prentice Hall, Englewood Cliffs, NJ.

Johnson, C. R., Jr., 1988. *Lectures on Adaptive Parameter Identification*, Prentice Hall, Englewood Cliffs, NJ.