The design engineer often finds himself in the situation where he is sure about the mathematical description of a system but is uncertain about the numerical values of the parameters. It is necessary in this case that he determine the numerical values of the parameters from observed data, a process known as "parameter estimation." Particularly as it relates to adaptive control systems, however, the process is not a simple one; the estimation has to be continually updated in the light of new observations. This requirement has led to the development of recursive methods. This article, the first of two, shows how recursive parameter estimation equations are obtained, and how the equations may be modified to allow for the estimation or "tracking" of possible parameter variations. The concluding article will discuss the application of recursive least-squares-estimation algorithms. It will explain why these algorithms are in an ideal form for the on-line analysis of data, and will demonstrate how they are used in the identification of dynamic processes from normal operating data.

Applying Parameter Estimation to Dynamic Systems—Part I

P. C. YOUNG, Naval Weapons Center

The basic problem in parameter estimation is deciding the best way to use observed data to form estimates of the unknown parameters. What is required immediately is a definition of what is "best," in terms of the mathematical language in which the estimation problem has been posed. This is by no means an easy thing to do since an answer depends on the particular application under consideration. However, in general, that estimate of the unknown parameter vector will be selected which finds the maximum or minimum of a chosen criterion function. Perhaps the best-known technique for finding this parameter vector is the minimization of the sum of the squares of the residuals or, as it is usually called, the "least squares" technique.

Estimating a single parameter

For a body moving in a straight line with constant velocity, the distance s, at time t, is related to the velocity, v, by the equation

$$s = tv \tag{1}$$

In practice, distance cannot be observed exactly; there will normally be some unavoidable error. Accordingly, what is actually measured is s^*, where

$$s^* = s + \epsilon_s \tag{2}$$

and ϵ_s are the random errors associated with the measurement. Table I shows a set of observations which, when plotted as shown in Figure 1, yields an approximate straight line whose slope may be expected to be v. No straight line, however, will pass exactly through all the data points because of the measurement errors. The question then is: Which of the many straight lines that could be drawn through the points will be the best representation of the data? This is not an easy question, since an answer requires a definition of what is "best" in the particular context.

Table I. Observed Values of Distance and Time

distance, s^* (ft)	5.71	9	15	19	20	45	55	78
time, t (sec)	0	1	2	3	4	10	12	18

(*Note datum distance s_0 is 5.71 ft*)

Probably the most straightforward method for arriving at this definition is the least squares method (Ref. 1). The basic procedure consists of taking an estimate of the slope and minimizing the sum of the squares of the difference between the ordinates of the actual and estimated values. By differentiating the

The major portion of this article was written while the author was with the Control Group, Engineering Department, University of Cambridge, England. The author is at present in the Guidance and Control Systems Division (Code 4041), Naval Weapons Center, China Lake, California.

sum of squares function, the estimated slope can be found analytically. Suppose there are k pairs of values for two variables, x_i and y_i. If the variables are related by a linear expression of the form:

$$y = xa \tag{3}$$

where x is exactly known, but the measure of y is only approximate (y^*), then:

$$y^* = y + \epsilon_y = xa + \epsilon_y \tag{4}$$

The method of least squares requires that the estimate $\hat{a}$ be chosen which minimizes the least squares criterion function J_2:

$$J_2 = \sum_{i=1}^{k} (x_i\hat{a} - y_i^*)^2 \tag{5}$$

Since J_2 represents a unimodal, or single minimum, function in the criterion-function parameter space, the solution is found by taking the gradient of J_2 with respect to $\hat{a}$ and equating to zero:

$$\nabla_{\hat{a}}(J_2) = \frac{\delta J_2}{\delta \hat{a}} = 2\sum_{i=1}^{k} x_i(x_i a - y_i^*) = 0 \tag{6}$$

As a result, the estimate $\hat{a}_k$ that minimizes J_2 after k samples is given by

$$\hat{a}_k = \left(\sum_{i=1}^{k} x_i^2\right)^{-1} \sum_{i=1}^{k} x_i y_i^* = p_k b_k \tag{7}$$

where

$$p_k = \left(\sum_{i=1}^{k} x_i^2\right)^{-1}$$

and

$$b_k = \sum_{i=1}^{k} x_i y_i^*$$

The attractive feature of the least squares formulation is its basic simplicity; it is particularly straightforward in theory and easy to handle in practice. Other criterion functions, such as the least magnitude

$$J_1 = \sum_{i=1}^{k} |x_i a - y_i^*|,$$

could also be used, but they are less convenient.

For parameter estimation a recursive form of Equation 7 is needed in which the estimate at the kth instant is a linear sum of the estimate at the previous $(k-1)$th instant plus a corrective term based on the data at the kth instant. The relationships given in Equation 7 can be written in the alternative form,

$$p_k^{-1} = \sum_{i=1}^{k} x_i^2 = p_{k-1}^{-1} + x_k^2 \tag{8}$$

and

$$b_k = \sum_{i=1}^{k} x_i y_i^* = b_{k-1} + x_k y_k^* \tag{9}$$

Table II. Weighting Functions

p_k	p_1	p_2	p_3	p_4	p_5	p_6	p_7
Stage-wise	1	0.2	0.0714	0.0330	0.00769	0.00364	0.00167
$p_0 =$ 100.0	0.99	0.199	0.0714	0.0332	0.00769	0.00364	0.00167
$p_0 =$ 1.0	0.50	0.1667	0.0667	0.0323	0.00763	0.00364	0.00617

Rearranging Equation 8:

$$p_{k-1} = p_k + p_k x_k^2 p_{k-1}$$

and

$$p_{k-1}x_k = p_k x_k + p_k x_k^2 p_{k-1} = p_k x_k(1 + p_{k-1}x_k^2) \tag{10}$$

so that:

$$p_{k-1}x_k(1 + p_{k-1}x_k^2)^{-1} = p_k x_k$$

Now, multiplying by $p_{k-1}x_k$ and using Equation 10 gives:

$$p_{k-1}^2 x_k^2(1 + p_{k-1}x_k^2)^{-1} = p_k x_k^2 p_{k-1} = p_{k-1} - p_k$$

Consequently,

$$p_k = p_{k-1} - p_{k-1}^2 x_k^2(1 + p_{k-1}x_k^2)^{-1} \tag{Ia}$$

Then, substituting this result in Equation 7 and using Equation 9:

$$\hat{a}_k = [p_{k-1} - p_{k-1}^2 x_k^2(1 + p_{k-1}x_k^2)^{-1}](b_{k-1} + x_k y_k^*)$$

With

$$\hat{a}_{k-1} = p_{k-1}b_{k-1}$$

this expression can be expanded to yield:

$$\hat{a}_k = \hat{a}_{k-1} - k_k(x_k\hat{a}_{k-1} - y_k^*) \tag{Ib}$$

where

$$k_k = p_{k-1}x_k(1 + p_{k-1}x_k^2)^{-1} \tag{Ic}$$

Equation Ic can be written:

$$k_k = (p_k p_k^{-1})p_{k-1}x_k(1 + p_{k-1}x_k^2)^{-1}$$

so that by substituting for p_k^{-1} from Equation 8, the following alternative definition of k_k is obtained:

$$\begin{aligned} k_k &= p_k(p_{k-1}^{-1} + x_k^2)\, p_{k-1}x_k(1 + p_{k-1}x_k^2)^{-1} \\ &= p_k(x_k + x_k^2 p_{k-1}x_k)(1 + p_{k-1}x_k^2)^{-1} \\ &= p_k x_k \end{aligned} \tag{Id}$$

The estimation algorithm given by Equations Ia through Ic or Id is the required recursive version of Equation 7. Since it is in this form, starting values for $\hat{a}$ and p must be chosen. One approach is to compute $\hat{a}_1$ and p_1 from Equation 7. Another is to allow $\hat{a}_0$ to have some arbitrary initial value, say zero, and set p_0 to some large positive number. Although less obvious, this second scheme yields asymptotically equivalent results, provided p_0 is chosen large enough.

Equation Ia or its equivalent, Equation 8, shows that p_k is a strictly decreasing function of the number of samples; so that as the estimation proceeds, the weight attached to the gradient measure is reduced. This means that large corrections are possible at the

start of the estimation procedure, when the estimate may be in gross error. However, as the estimation progresses, the estimates will converge. Less and less correction is warranted, since it becomes more likely that the observed gradient is the result of measurement noise rather than estimation error.

To see how the simple recursive least-squares algorithm I works in practice, consider the moving body problem mentioned earlier, and let $y = s^* - s_0$; $x = t$; $a = v$. Figure 2 shows the estimation results obtained using the algorithm from initial conditions $\hat{a}_0 = 0$; $p_0 = 100$ and $\hat{a}_0 = 0$; $p_0 = 1.0$. Also shown are the results obtained from the stagewise solution of Equation 7. While the results for $p_0 = 100$ are virtually identical with the stagewise results, those for $p_0 = 1.0$ show some discrepancy. Thus the need to choose large values for p_0 in order to ensure equivalence. Table II gives the weighting functions, p_k, for this same example; note their strictly decreasing nature.

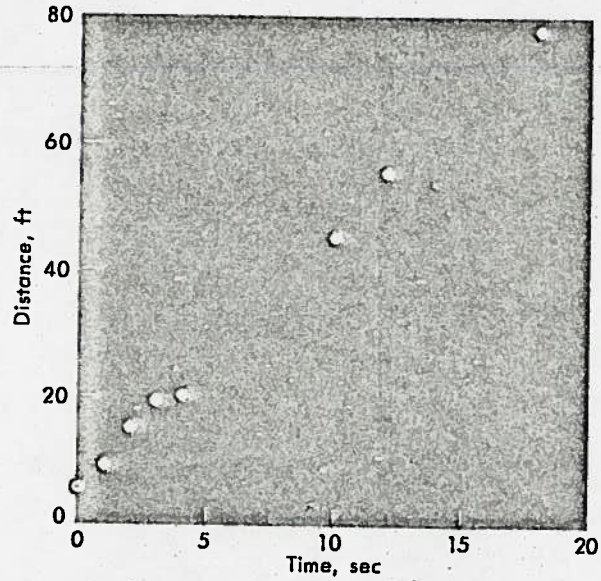

FIG. 1. In the case of a body moving with constant velocity, time and distance should vary linearly. However, observed values will not lie exactly on a straight line. An estimate of the best fit can be found by using a least squares technique.

Estimating multiple parameters

A more general problem is the estimation of the set of n unknown parameters, a_j, that appear in a linear relationship of the form

$$y = a_1x_1 + a_2x_2 + \ldots + a_nx_n$$

where, once again, the observation y^* of y is approximate, or contaminated by noise, ϵ_y, while the x_j are exactly known quantities. In this case, the minimization of the least-squares criterion function,

$$J_2 \triangleq \sum_{i=1}^{k}\left(\sum_{j=1}^{n} x_{ji}\hat{a}_j - y_i^*\right)^2 \qquad (11)$$

requires that all the partial derivatives of J_2 with respect to each of the parameter estimates $\hat{a}_j$ should be set simultaneously to zero. Such a procedure yields a set of n linear simultaneous algebraic equations that are sometimes termed the normal equations, and that can be solved for the parameter estimates $\hat{a}_{jk}$ at the kth instant. In this case the least squares results can be obtained by using a vector-matrix formulation. Thus, by writing Equation 11 in the alternative vector form

$$y = \mathbf{x}^T\mathbf{a} \qquad (12)$$

where

$$\mathbf{x}^T = (x_1x_2 \ldots\ldots x_n);\ \mathbf{a} = (a_1a_2 \ldots\ldots a_n),$$

where T denotes the transpose of a matrix, J_2 can be defined as

$$J_2 \triangleq \sum_{i=1}^{k}(\mathbf{x}_i^T\hat{\mathbf{a}} - y_i^*)^2 \qquad (13)$$

where

$$y_i^* = \mathbf{x}_i^T\mathbf{a} + \epsilon_y \qquad (14)$$

Now the normal equations become

$$\nabla_{\hat{a}}(J_2) = \left(\sum_{i=1}^{k}\mathbf{x}_i\mathbf{x}_i^T\right)\hat{\mathbf{a}} - \sum_{i=1}^{k}\mathbf{x}_iy_i^* = 0 \qquad (15)$$

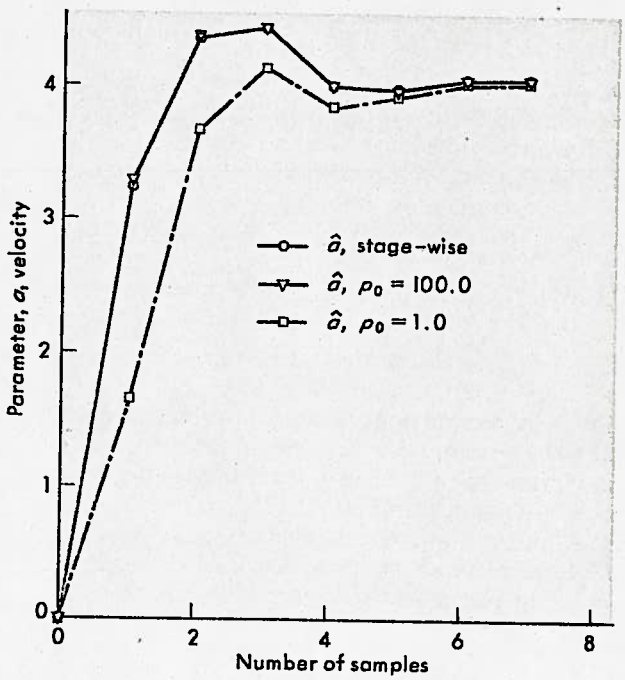

FIG. 2. The accuracy of estimation can be improved with the number of samples taken. The circles show points obtained by a step-by-step solution of the basic least squares formula; the other two sets of points are obtained by using the recursive least squares equation. Where p_0 is large, the results agree with the original formula.

where $\nabla_{\hat{a}}(J_2)$ denotes the gradient of J_2 with respect to all the elements of $\hat{a}$. The solution to the equations takes the form,

$$\hat{a}_k = P_k B_k \tag{16}$$

where P_k is an $n \times n$ matrix and B_k is an $n \times 1$ vector, and are defined accordingly.

The scalar relationships of Equations 8 and 9 of the single-parameter example are now replaced by the following vector matrix expressions:

$$P_k^{-1} = P_{k-1}^{-1} + \mathbf{x}_k \mathbf{x}_k^T \tag{17}$$

and

$$B_k = B_{k-1} + \mathbf{x}_k y_k^* \tag{18}$$

In order to develop a recursive version of Equation 16, premultiply Equation 17 by P_k and then post-multiply by P_{k-1}, to give

$$P_{k-1} = P_k + P_k \mathbf{x}_k \mathbf{x}_k^T P_{k-1} \tag{19}$$

Post-multiplying by $\mathbf{x}_k$:

$$\begin{aligned} P_{k-1}\mathbf{x}_k &= P_k \mathbf{x}_k + P_k \mathbf{x}_k \mathbf{x}_k^T P_{k-1} \mathbf{x}_k \\ &= P_k \mathbf{x}_k (1 + \mathbf{x}_k^T P_{k-1} \mathbf{x}_k) \end{aligned}$$

Then, post-multiplying by

$$(1 + \mathbf{x}_k^T P_{k-1} \mathbf{x}_k)^{-1} \mathbf{x}_k^T P_{k-1}:$$

$$P_{k-1}\mathbf{x}_k [1 + \mathbf{x}_k^T P_{k-1} \mathbf{x}_k]^{-1} \mathbf{x}_k^T P_{k-1} = P_k \mathbf{x}_k \mathbf{x}_k^T P_{k-1}$$

Finally, substituting from Equation 19:

$$P_k = P_{k-1} - P_{k-1}\mathbf{x}_k [1 + \mathbf{x}_k^T P_{k-1} \mathbf{x}_k]^{-1} \mathbf{x}_k^T P_{k-1} \tag{IIa}$$

As in the single-parameter case (care being taken to obey the rules of matrix algebra (Ref. 2)), the following equivalent recursive equation is obtained by substituting in Equation 16:

$$\hat{\mathbf{a}}_k = \hat{\mathbf{a}}_{k-1} - P_{k-1}\mathbf{x}_k (1 + \mathbf{x}_k^T P_{k-1} \mathbf{x}_k)^{-1} (\mathbf{x}_k^T \hat{\mathbf{a}}_{k-1} - y_k^*) \tag{IIb}$$

or

$$\hat{\mathbf{a}}_k = \hat{\mathbf{a}}_{k-1} - P_k (\mathbf{x}_k \mathbf{x}_k^T \hat{\mathbf{a}}_{k-1} - \mathbf{x}_k y_k^*) \tag{IIc}$$

Equation set II constitutes the recursive form of the least squares solution, Equation 16. Once again it is necessary to specify starting values. The terms $\hat{\mathbf{a}}_0$ and P_0, having large diagonal elements, will yield performances commensurate with the stagewise solution of the same problem (Refs. 2 and 3).

Another similarity with the scalar case is that algorithm II can be interpreted as a special form of gradient procedure. In this multiparameter case, the instantaneous gradient, $\mathbf{x}_k \mathbf{x}_k^T \hat{\mathbf{a}}_{k-1} - \mathbf{x}_k y_k^*$, is a vector. Thus, the scalar weighting factor is replaced by the time variable weighting matrix P_k. However, the strictly decreasing nature and consequent smoothing effect of this matrix is directly analogous to that of p_k in the single input case.

In contrast to the scalar situation, algorithm II does provide some considerable advantage over the stagewise solution of Equation 16. In addition to the now convenient recursive form, which provides for a minimum of computer storage, the term $(1 + \mathbf{x}_k^T P_{k-1} \mathbf{x}_k)$ is simply a scalar quantity. As a result, there is no requirement for direct matrix inversion even though the equivalent classical solution of Equation 16 entails inverting an $n \times n$ matrix.

To see how to use the recursive least-squares algorithm II, consider a slightly expanded form of the moving body problem discussed in the previous section. Suppose that the datum distance, s_0, is unknown, so that Equation 1 has to be written,

$$s = s_0 + vt$$

This equation can be written in the vector form of Equation 12, where

$$\mathbf{x}^T = [1 \quad t]; \quad \mathbf{a}^T = [s_0 \quad v]$$

Consequently, the simultaneous estimation of s_0 and v by reference to the measurements s^* represents a two-parameter estimation problem that can be solved in a least squares manner. Figure 3 shows the estimation results obtained from algorithm II with $\mathbf{a}_0 = [0]$ and

$$P_0 = \begin{bmatrix} 10^4 & 0 \\ 0 & 10^4 \end{bmatrix}.$$

Figure 4 provides the time history of the P matrix elements.

Statistical regression analysis

The recursive least squares algorithm is a deterministic estimation procedure in the sense that it makes no assumptions about the statistical nature of the signals or the noise and does not provide any statistical information on the nature of the estimates. To indicate how even the minimum of statistical information can help to improve the algorithm, assume that the measurement noise ϵ_y can be considered as a zero mean independent, or uncorrelated, random sequence with variance σ^2. Then the estimates are asymptotically unbiased and the covariance matrix of the estimation error at the kth instant, P_k^*, is directly related to the P_k matrix by the equation:

$$P_k^* = \mathrm{E}(\tilde{\mathbf{a}}\tilde{\mathbf{a}}^T) = \sigma^2 P_k$$

where $\tilde{\mathbf{a}} = \hat{\mathbf{a}} - \mathbf{a}$ and $\mathrm{E}(\cdot)$ is the expectation operator. Substituting this second result in algorithm II:

$$\hat{\mathbf{a}}_k = \hat{\mathbf{a}}_{k-1} - P_{k-1}^* \mathbf{x}_k (\sigma^2 + \mathbf{x}_k^T P_{k-1}^* \mathbf{x}_k)^{-1} (\mathbf{x}_k^T \hat{\mathbf{a}}_{k-1} - y_k^*) \tag{IIIa}$$

or,

$$\hat{\mathbf{a}}_k = \hat{\mathbf{a}}_{k-1} - P_k^* (\mathbf{x}_k \mathbf{x}_k^T \hat{\mathbf{a}}_{k-1} - \mathbf{x}_k y_k^*) \tag{IIIb}$$

while

$$P_k^* = P_{k-1}^* - P_{k-1}^* \mathbf{x}_k (\sigma^2 + \mathbf{x}_k^T P_{k-1}^* \mathbf{x}_k)^{-1} \mathbf{x}_k^T P_{k-1}^* \tag{IIIc}$$

Algorithm III, which is a recursive version of the well-known least-squares regression equations (Refs. 1 and 4), not only supplies the parameter estimates at each sampling instant but provides an indication of the accuracy of these estimates through the error covariance matrix P_k^*. P_k^* behaves in a similar manner to P_k and is a strictly decreasing function of the sample size. This property is a physical indication that the estimates are statistically consistent; that is, that their accuracy increases as more data is utilized.

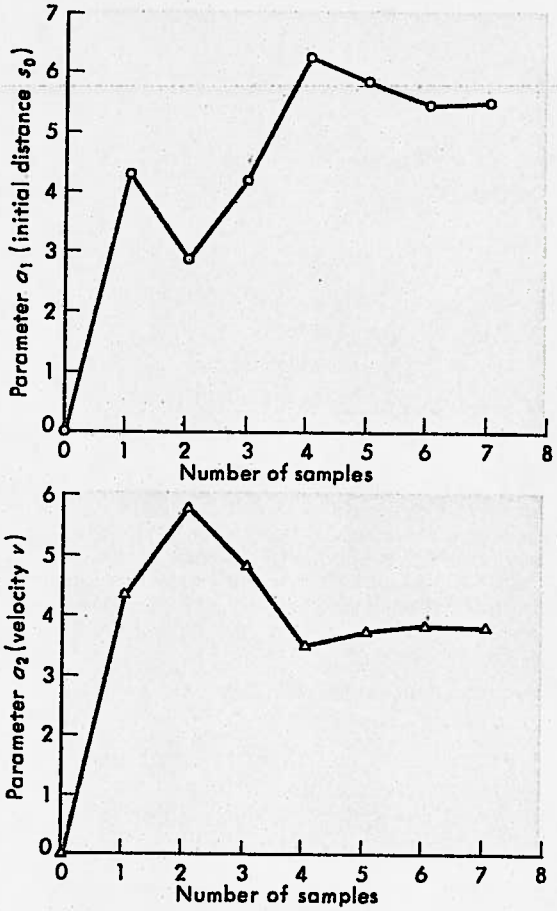

FIG. 3. If both velocity and initial distance are unknown, the moving body example becomes a two-parameter estimation problem. Shown here are the results of successive estimation by the recursive multiparameter least squares method.

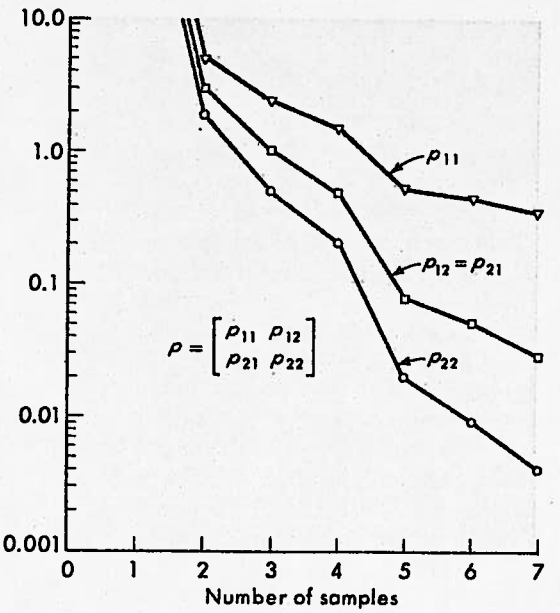

FIG. 4. A measure of the accuracy of the estimation is provided by the size of the elements of the P matrix. For the two-parameter moving body example, the elements become less than 1.0 within five samples.

Time variable parameter estimation

If a common factor of $1/k$ is introduced into Equation 15, then the terms

$$\frac{1}{k}\sum_{i=1}^{k} \mathbf{x}_i\mathbf{x}_i^T \text{ and } \frac{1}{k}\sum_{i=1}^{k} \mathbf{x}_i y_i^*$$

represent finite-time averaging operations (Ref. 5). This means that all data is weighted equally over the observation interval of k samples and there is an implicit assumption that the parameters remain constant during this period. In order to allow for possible parameter variation, therefore, it is necessary to modify the estimation procedure in some way.

Shaping the memory of the estimation scheme

One approach to the problem of detecting parameter variation is simply to curtail the memory of the estimation procedure in some manner so that as new observations are included, the effect of earlier observations is reduced. This can be done by replacing the finite time averaging operations by exponentially weighted past (ewp) averages (Ref. 5). This technique, which entails processing the data by means of a discrete, low-pass filter with an exponential weighting function, requires replacing Equations 17 and 18 by the following (Refs. 6 and 7):

$$P_k^{-1} = (1-\alpha)P_{k-1}^{-1} + \alpha[\mathbf{x}_k\mathbf{x}_k^T] \quad (21)$$

$$B_k = (1-\alpha)B_{k-1} + \alpha[\mathbf{x}_k y_k^*] \quad (22)$$

where $0 < \alpha \ll 1.0$ represents a scalar weighting factor that specifies the length of the filter memory. Using these equations, which simply represent discrete, first-order, low-pass filtering operations on the elements of $\mathbf{x}_k\mathbf{x}_k^T$ and $\mathbf{x}_k y_k^*$, respectively, is equivalent to minimizing a criterion function of the type

$$J_k^{\text{EWP}} = \sum_{i=1}^{k} (\mathbf{x}_i^T\hat{a}_i - y_i^*)^2(1-\alpha)^{k-i}\alpha \quad (23)$$

This is simply the normal least-squares criterion function modified by the exponential weighting term $(1-\alpha)^{k-i}\alpha$. Figure 5 gives examples of this weighting effect for $k = 10$ and $\alpha = 0.1$ and 0.3, respectively. Proceeding now as in the time-invariant situation, it is not difficult to obtain the following recursive algorithm (Refs. 8 and 9):

$$\hat{\mathbf{a}}_k = \hat{\mathbf{a}}_{k-1} - \frac{\alpha}{1-\alpha}P_{k-1}\mathbf{x}_k\left(1 + \frac{\alpha}{1-\alpha}\mathbf{x}_k^TP_{k-1}\mathbf{x}_k\right)^{-1} \times (\mathbf{x}_k^T\hat{\mathbf{a}}_{k-1} - y_k^*) \quad \text{(IVa)}$$

or

$$\hat{\mathbf{a}}_k = \hat{\mathbf{a}}_{k-1} - \alpha P_k(\mathbf{x}_k\mathbf{x}_k^T\hat{\mathbf{a}}_{k-1} - \mathbf{x}_k y_k^*) \quad \text{(IVb)}$$

while

$$P_k = \frac{1}{1-\alpha}P_{k-1} - \frac{\alpha}{(1-\alpha)^2}P_{k-1}\mathbf{x}_k\left(1 + \frac{\alpha}{1-\alpha}\mathbf{x}_k^TP_{k-1}\mathbf{x}_k\right)^{-1} \times \mathbf{x}_k^TP_{k-1} \quad \text{(IVc)}$$

The physical effect of the factor α is simply to prevent the P matrix elements from becoming too small, so that new data continues to have some effect on the estimates. In this way, any modification to the

measured gradient caused by parameter variation can be detected and used to update the parameter estimates. Of course, this approach has the disadvantage that the effects of noise will also be detected and used to modify the estimates. Thus the estimation scheme only functions satisfactorily if the parameter variations are larger than the residual fluctuations due to noise. In practice, α in algorithm IV is better replaced by a time variable scalar α_k that is a strictly decreasing function of sample size and tends to α as k tends to infinity:

$$\alpha_k = \alpha/(1 - (1 - \alpha)^k); \; k > 1$$

Here α will normally need to be <0.1.

Modeling the parameter variations

The statistical interpretation of the least squares equations discussed earlier suggests an alternative method of detecting parameter variations (Refs. 2 and 3). For instance, suppose that any parameter variations can be described by the discrete vector equation

$$\mathbf{a}_k = \Phi \mathbf{a}_{k-1} + \mathbf{q}_{k-1} \tag{24}$$

where $\Phi = \Phi(k, k-1)$ is an $n \times n$ transition matrix (Ref. 10) and $\mathbf{q}_{k-1}$ is an n vector of independent random variables with zero mean and covariance matrix $E(\mathbf{q}_i \mathbf{q}_j^T) = Q\delta_{ij}$ where δ_{ij}, the Kronecker delta, equals

$$\delta_{ij} = \begin{pmatrix} 0, & i = j \\ 1, & i = j \end{pmatrix}$$

which is uncorrelated with the measurement noise $(\epsilon_y)_{k-1}$. When $\Phi = I$, the identity or unit matrix, then this model merely states that the parameters may undergo random fluctuations $\mathbf{q}_{k-1}$ between samples, the expected variance of these fluctuations being defined by the diagonal elements of the covariance matrix Q. In other words, the presence of $\mathbf{q}_{k-1}$ introduces a statistical degree of freedom to the parameter variation in Equation 24.

If something is known about the parameter variation—for example, that it is sinusoidal with a given frequency—then a suitable Φ matrix can be selected and used in Equation 24. In these circumstances a simple statistical analysis of the modified estimation problem (Ref. 3) yields a new least-squares regression algorithm. The only difference between this new algorithm and the basic least squares regression (Equations III) is that the between-samples $\mathbf{a}_{k-1}$ and P_{k-1}^* are updated to $\hat{\mathbf{a}}_{k|k-1}$ and $P_{k|k-1}^*$, respectively. Here

$$\mathbf{a}_{k|k-1} = \Phi \hat{\mathbf{a}}_{k-1} \tag{25}$$

$$P_{k|k-1}^* = P_{k-1}^* + Q \tag{26}$$

where $\hat{\mathbf{a}}_{k|k-1}$ and $P_{k|k-1}^*$ are respectively the *a priori* updates of $\hat{\mathbf{a}}_{k-1}$ and P_{k-1}^* at the kth instant, based on the information obtained up to the $(k-1)$th instant and a knowledge of the parameter variation law, Equation 24. In the absence of any evidence to the contrary, Q is chosen purely diagonal with elements

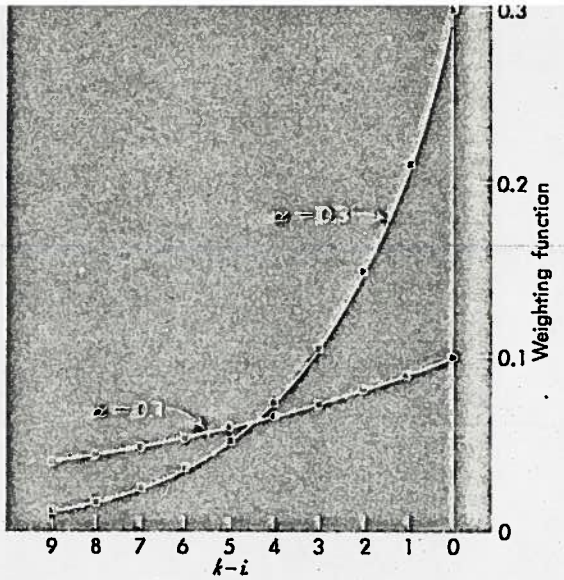

FIG. 5. For time-varying parameters, less weight should be given to the earlier observations so that the parameter variation can be detected. This can be done by weighting the least-squares criterion function with a factor that acts in a manner similar to a low-pass filter. Shown here are the exponential weighting functions for a sample length $k = 10$ and two values of the weighting factor, α.

reflecting the expected rate of variation of the parameters between samples. If there is no possibility of random variation, then $\mathbf{q}_{k-1}$ is zero and Q is removed from Equation 26.

By using a purely heuristic argument (Ref. 2), it is now possible to consider modifying the simple deterministic least-squares algorithm II in a manner similar to that shown above. In the simplest and most generally useful case, where $\Phi = I$, we obtain the following estimation equations:

$$\hat{\mathbf{a}}_k = \hat{\mathbf{a}}_{k-1} - P_{k|k-1}\mathbf{x}_k(1 + \mathbf{x}_k^T P_{k|k-1}\mathbf{x}_k)^{-1}(\mathbf{x}_k^T \hat{\mathbf{a}}_{k-1} - y_k^*) \tag{Va}$$

$$P_{k|k-1} = P_{k-1} + D \tag{Vb}$$

$$P_k = P_{k|k-1} - P_{k|k-1}\mathbf{x}_k(1 + \mathbf{x}_k^T P_{k|k-1}\mathbf{x}_k)^{-1}\mathbf{x}_k^T P_{k|k-1} \tag{Vc}$$

In these equations, D is a positive, definite and usually diagonal matrix which is analogous in its effect to the covariance matrix Q in the regression case. It can be chosen initially by reference to the expected rate of variation of the parameters and modified if necessary in the light of experiment. In the case where a parameter is constant, the corresponding diagonal element of D should be set at zero.

In computational terms, the introduction of the D matrix is similar in effect to the introduction of the exponential data-weighting function; the addition of D to P_{k-1} at each instant introduces a lower bound on the magnitude of the P matrix elements, preventing them from getting too small. In general, however, algorithm V is more attractive than the equivalent exponential memory scheme since it is rather simpler to implement and basically more flexible. For instance, the choice of a random walk model (i.e., $\Phi = I$ in Equation 24) is rather arbitrary; in certain circumstances it may be more realistic to specify other statistical models (Ref. 3). In order

to emphasize this point, consider for the moment the following very general model:

$$\hat{\mathbf{a}}_k = \Phi\hat{\mathbf{a}}_{k-1} + \Gamma\mathbf{q}_{k-1}$$

where Φ is a known $n \times n$ transition matrix, Γ is a known $n \times m$ matrix, and $\mathbf{q}_{k-1}$ is an m vector of independent random variables with zero mean and covariance Q. In this special case, update equation (26) in the regression case becomes:

$$p_{k|k-1}^* = \Phi p^*_{k-1}\Phi^T + \Gamma Q\Gamma^T \qquad (27)$$

A heuristic argument can once more be used to modify the deterministic least squares algorithm into the following "dynamic" form:

$$\hat{a}_k = \hat{a}_{k|k-1} - P_{k|k-1}x_k[1 + x_k^T P_{k|k-1}x_k]^{-1} \{x_k^T\hat{a}_{k|k-1} - y_k^*] \quad \text{(VIa)}$$

$$\hat{a}_{k|k-1} = \Phi\hat{a}_{k-1} \quad \text{(VIb)}$$

$$p_{k|k-1} = \Phi P_{k-1}\Phi^T + \Gamma D\Gamma^T \quad \text{(VIc)}$$

$$P_k = p_{k|k-1} - P_{k|k-1}x_k[1 + x_k^T p_{k|k-1}x_k]^{-1} x_k^T P_{k|k-1} \quad \text{(VId)}$$

where D is an $m \times m$ matrix selected by experiment to allow for any random parameter variations not accounted for in Equation VIb.

Note that this dynamic least squares regression algorithm is a special form of the famous Kalman filter-estimator equations (Ref. 11 and 13). In basic terms there is no difference between the parameters and the state of a process (Ref. 12). Segregation of the two is often based on purely physical reasoning, to the effect that the parameters are normally either time invariant or only slowly variable while, by comparison, the state is rapidly varying. This unified concept of parameter and state estimation is emphasized in the simple moving body example discussed earlier. Here the unknown "parameters" are actually the states of a dynamic system.

Structural models

Up to this point only a particular form of linear estimation model has been considered; namely, the regression model, in which the variables associated with the unknown parameters are *exactly* known quantities. In practice a number of different estimation models are encountered (Ref. 14), and the techniques described here may need to be modified in some way. This whole topic is discussed fully in a book by Graybill (Ref. 15). Although Graybill's classification of the various models is open to question, he does treat all the major possibilities in considerable detail. In the present article, it will suffice to mention one of the most important models met in practical situations, the structural model (Ref. 1). In the structural model, the basic relationship between the parameters is still in the form shown in Equation 12. However, the elements of x are no longer exactly known quantities and can only be observed in error. In other words, the observed value of x is x*, where

$$\mathbf{x}^* = \mathbf{x} + \epsilon_x$$

and ϵ_x is a n vector of measurement noise associated with the observation of x. In this situation, it can be shown (Refs. 16 and 17) that the estimates obtained via a least squares analysis are asymptotically biased to a degree dependent upon the noise/signal ratio on the observation x*.

There are a number of ways of solving the structural model problem (Ref. 18). If the noise statistics are known *a priori*, then there is no real difficulty, since it is possible to compensate directly for the bias (Ref. 19). Another approach to the problem, attractive because it does not require detailed information on the noise statistics, is the "instrumental variable" method (Refs. 2 and 17).

REFERENCES

1. *The Advanced Theory of Statistics*, Vol. 2, M. G. Kendall and A. Stuart, Griffin, London, 1961.
2. *An Instrumental Variable Method for Real-Time Identification of a Noisy Process*, P. C. Young, IFAC Congress, Warsaw, Poland, 1969.
3. *Optimal Estimation Identification and Control*, R. C. K. Lee, Research Monograph 28, MIT Press, Cambridge, Mass., 1964.
4. "Process Parameter Estimation," P. C. Young, *Control*, Vol. 12, No. 125, November 1968, pp. 931–936.
5. *Random Process Simulation and Measurements*, G. A. Korn, McGraw-Hill, New York, 1966.
6. "Design of a Self-Optimizing Control System," R. E. Kalman, *Trans. ASME, Basic Engineering*, Vol. 80D, No. 2, February 1958.
7. "Digital Adaptive Flight Control System for Aerospace Vehicles," R. E. Andeen and P. O. Shipley, *AIAA Journal*, Vol. 1, No. 5, May 1963.
8. *On the Relationship Between Various Least Square Non-stationary Parameter Estimation Techniques*, P. C. Young, Engineering Laboratory, Cambridge University, England, internal report, December 1966 (available from author).
9. "On-Line Model Making for a Chemical Plant," J. W. Bray et al., *Trans. Soc. Inst. Tech.*, Vol. 17, No. 8, 1965.
10. *State Variables for Engineers*, P. M. Derusso, R. J. Roy, and C. M. Close, Wiley, New York, 1965.
11. "Kalman Filtering Techniques," H. W. Sorenson, in *Advances in Control Systems*, Vol. 3, C. T. Leondes, editor; Academic Press, New York, 1966.
12. "Linear Regression Applied to System Identification for Adaptive Control Systems," R. E. Kopp and R. J. Orford, *AIAA Journal*, Vol. 1, No. 10, October 1963.
13. "Application of State Space Methods to Navigation Problems," S. F. Schmidt, *Advances in Control Systems*, Vol. 3, C. T. Leondes, editor; Academic Press, New York, 1966.
14. "Statistical, Structural and Regression Models in Linear Process Identification," R. B. Kerr, *Proceedings Joint Automatic Control Conference*, 1956, pp. 458–466.
15. *An Introduction to Linear Statistical Models—Vol. 1*, F. A. Graybill, McGraw-Hill, New York, 1961.
16. "Process Parameter Estimation and Self-Adaptive Control," P. C. Young, *IFAC Conference on Theory of Self-Adaptive Control Systems*, P. H. Hammond, editor, pp. 118–139, Teddington, England, 1965.
17. "Errors in Variables," J. Durbin, *Review of Int. Statist. Inst.*, Vol. 22, No. 23, 1954.
18. "Regression Analysis and Process Parameter Estimation: a Cautionary Message," P. C. Young, *Simulation*, Vol. 10, No. 3, pp. 125–128, March 1968.
19. "Estimation of System Pulse Transfer Function in the Presence of Noise," M. J. Levin, *Proceedings, Joint Automatic Control Conference*, 1963, pp. 452–458.