# Identification of Stochastic Linear Dynamic Systems Using Kalman Filter Representation

RAMAN K. MEHRA*
*Systems Control, Inc., Palo Alto, Calif.*

**Using a Kalman filter representation, a new set of equations has been derived for the maximum likelihood identification of linear dynamic systems. These equations are shown to be directly related to the filtering and smoothing equations for linear dynamic systems. Numerical results are obtained for a fourth-order single-input single-output system using Davidon's Conjugate Gradient Method.**

## I. Introduction

THE identification of linear dynamic systems has been the subject of several investigations in engineering, physical sciences and econometrics.[1–5] Two important aspects of the identification problem are, 1) the choice of a suitable representation for the system and 2) a criterion for identification.

In econometrics, the most commonly chosen form is an autoregressive scheme with moving average residuals.[5] In communications, a frequency-domain representation and in control applications a time-domain or a state space representation is more useful.[6] Irrespective of whichever form is used, it is important to choose a parsimonious representation, i.e., one which involves minimum number of unknowns.

In this paper, we consider a special form of state space representation for the system. It is based on the Kalman filter[7] for a linear dynamic system and has been called Levy's "proper canonical representation."[8] In addition to the obvious advantage of this form in a number of applications, it turns out to be very suitable for identification purposes.

The choice of a criterion for identification varies with the application. In this paper, we use maximum likelihood criterion which under certain restrictions leads to estimates which are asymptotically unbiased, consistent and efficient.[2,3] For finite samples, however, there may be more than one maximum likelihood estimate. To overcome this problem partially, we shall obtain starting values from a correlation scheme which in addition to giving consistent (but inefficient) estimates also determines the order of the system.[10] Two other criterions commonly used for parameter estimation viz. the Equation Error Criterion and the Output Error Criterion[11,20] will also be considered. It will be shown that they are special cases of the Maximum Likelihood Criterion.

The earlier work in this area to which the present work is related is that of Åström[2] and Kashyap.[3] Although the basic ideas are the same, the interpretation of the models and the details of the algorithm are different. It is believed that the viewpoint taken here is extended more readily to situations where the physical model of the system is known up to certain unknown parameters. This is the situation in most of the aerospace applications.

A block diagram of the identification procedure is shown in Fig. 1.

## II. A Canonical Form for the System

A state variable representation of a discrete, single-input, single-output linear dynamic system† is

$$x_{i+1} = \Phi x_i + g u_i \tag{1}$$

$$z_i = h^T x_i + v_i \tag{2}$$

where $x_i$ is $n \times 1$ state vector, $z_i$ is the output or the observed sequence, $\Phi$ is $n \times n$ transition matrix, $g$ and $h$ are $n \times 1$ constant vectors and $u_i$ and $v_i$ are Gaussian white noise sequences with zero mean and unknown variances

$$E(u_i) = 0 \qquad E(u_i u_j) = q\delta_{ij}$$

$$E(v_i) = 0 \qquad E(v_i v_j) = r\delta_{ij}$$

$$E(u_i v_j) = 0 \text{ for all } i \text{ and } j$$

Both $x_i$ and $z_i$ are zero mean Gaussian random sequences for all $i \geq 0$.

By suitable linear transformations $\Phi$ and $h$ can be reduced to the following canonical form for a completely observable system.[12]‡

$$\Phi = \begin{bmatrix} 0 & 1 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & 1 & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & 1 \\ -\varphi_1 & \cdot & \cdot & \cdot & \cdot & -\varphi_n \end{bmatrix} \qquad h = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

An alternate representation for this system is provided by a Kalman filter[7] for this system viz.

$$\hat{x}_{i+1/i} = \Phi \hat{x}_{i/i-1} + \Phi k \nu_i \tag{3}$$

$$z_i = h^T \hat{x}_{i/i-1} + \nu_i \tag{4}$$

with initial condition $\hat{x}_{0/-1} \equiv \hat{x}_0$.

Here $\hat{x}_{i/i-1}$ denotes the best estimate of $x_i$ based on ($z_j$, $j = 0, \ldots, i-1$) and $\nu_i$ is a Gaussian white noise sequence called the "innovation sequence"[15]

$$E(\nu_i) = 0 \qquad E(\nu_i \nu_j) = b\delta_{ij}$$

and $k$ is an $n \times 1$ constant vector.§ The steady-state values of $k$ and $b$ are obtained from $g$, $q$, $r$ by solving the following set of equations,[7,13]

$$k = Mh/b \tag{5}$$

$$b = h^T M h + r \tag{6}$$

$$M = \Phi(I - kh^T)M(I - kh^T)^T\Phi^T + r\Phi k k^T \Phi^T + qgg^T \tag{7}$$

Equations (3–7) have been called Levy's "proper canonical representation."[8] It will be useful to identify a system in this representation since the filtered estimates of the state will be obtained without any extra effort. Moreover, it will be

Received December 15, 1969; revision received May 28, 1970. This work was done while the author was with the Analytic Sciences Corporation, Reading, Mass.

* Senior Research Engineer. Member AIAA.

† The extension to continuous systems is given in the Appendix.

‡ It is not essential to use a canonical form as long as a suitable representation with a uniquely identifiable set of parameters is chosen for the system.

§ $k$ and $b$ are in general time-varying even for a constant system, but they reach steady state value for completely observable and controllable systems.[7] We neglect the initial transient portion in which they are time-varying.

easy to obtain predicted and smoothed estimates once the filtered estimates are known.[15] Another advantage of using Eqs. (3) and (4) for maximum likelihood estimation will become clear in the next section.

For obtaining $g$, $q$ and $r$ from $k$ and $b$, it is necessary to invert Eqs. (5–7). This gives

$$Mh = kb \quad (8)$$

$$r = (1 - h^T k)b \quad (9)$$

$$qgg^T = M - \Phi(I - kh^T)M(I - kh^T)^T\Phi^T - r\Phi kk^T\Phi^T \quad (10)$$

The solution for $r$ is straightforward, but the solution for $g$ and $q$ is somewhat tedious. Note that, in general, Eqs. (8) and (10) do not have a unique solution. Therefore it is necessary to impose further constraints on $g$ and $q$ to obtain a unique solution. For more on these questions and for different methods to solve them, see Refs. 13 and 16.

## III. Maximum Likelihood Estimation

The system of Eqs. (3) and (4) has $(3n + 1)$ unknowns viz. $\varphi, k, \hat{x}_0$ and $b$ where $\varphi$ is $n \times 1$ vector consisting of $\varphi_1, \ldots \varphi_n$.

Let $\theta$ denote the vector of $(3n + 1)$ unknowns

$$\theta = \begin{bmatrix} \varphi \\ k \\ \hat{x}_0 \\ b \end{bmatrix}$$

It is required to estimate $\theta$ from a set $Z$ of $N$ observations where $Z \equiv (z_i, i = 0, N - 1)$.

The maximum likelihood estimate is obtained by maximizing the conditional density function $p(Z/\theta)$ or its logarithm. In order to obtain $p(Z,\theta)$, we first "whiten" $Z$ through a causal invertible linear transformation. Such a transformation is provided by a Kalman filter in the sense that the innovation sequence $\nu_i = (z_i - h^T\hat{x}_{i/i-1})$ is a zero mean Gaussian white noise sequence.[15] Following Schweppe,[14] the likelihood function $J(\theta)$ can be written as

$$J(\theta) = -\frac{1}{2b}\sum_{i=0}^{N-1}\nu_i^2 - \frac{N}{2}\ln b \quad (11)$$

The estimation problem may now be stated as follows: Maximize $J(\theta)$ with respect to $\theta$ subject to constraints of Eqs. (3) and (4).

We now adjoin the constraints of Eq. (3) to $J(\theta)$ using multipliers $\lambda_i$. Then considering variations, and simplifying the results, we obtain the following set of necessary conditions of optimality:

$$\nu_i = z_i - h^T\hat{x}_{i/i-1}$$

$$\hat{x}_{i+1/i} = \Phi\hat{x}_{i/i-1} + \Phi k\nu_i$$

$$\lambda_i = (I - kh^T)^T\Phi^T\lambda_{i+1} + hb^{-1}\nu_i, \ \lambda_N = 0 \quad (12)$$

$$\frac{\partial J}{\partial b} = \frac{1}{2b^2}\left(-Nb + \sum_{i=0}^{N-1}\nu_i^2\right) = 0 \quad (13)$$

$$\frac{\partial J}{\partial \varphi} = \sum_{i=0}^{N-1}[\lambda_{i+1}]_n(\hat{x}_{i/i-1} + k\nu_i) = 0 \quad (14)$$

$$\frac{\partial J}{\partial k} = \sum_{i=0}^{N-1}\nu_i\Phi^T\lambda_{i+1} = 0 \quad (15)$$

$$\frac{\partial J}{\partial \hat{x}_0} = \lambda_0 = 0 \quad (16)$$

where $[\lambda_{i+1}]_n$ denotes the $n$th element of the vector $\lambda_{i+1}$.

### Comments

1) It is interesting to note that the equation for adjoint multipliers $\lambda_i$ is the same as that in the formulation of smoothing equations by Bryson[12] and Kailath.[15]

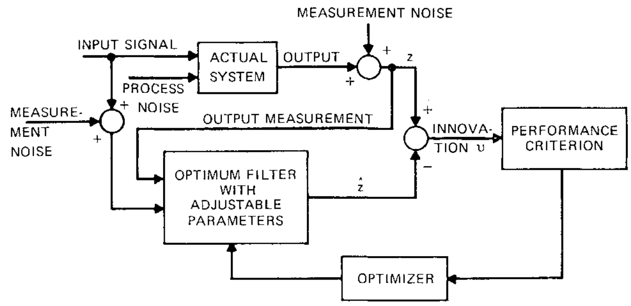


**Fig. 1 Identification scheme.**

2) Asymptotically as $N \to \infty$ the time averages in Eqs. (13–15) can be replaced by their expected values. Then by expressing $\hat{x}_{i/i-1}$ and $\lambda_i$ in terms of $\nu_i$'s from Eqs. (3) and (12), it can be shown that a necessary and suffidient condition for $\partial J/\partial\varphi$ and $\partial J/\partial k$ to vanish is

$$E(\nu_i) = 0, E(\nu_i\nu_j) = 0, i \neq j$$

This is exactly the innovation property of an optimal filter when all the parameters are known.[15]

3) Equations (3–16) for maximum likelihood estimation are different from the equations reported in the literature.[2,3] The difference results from using the Kalman filter representation and the multipliers $\lambda_i$.

4) The case in which there is a known deterministic input to the system can be treated in a similar manner, e.g., let $y_i$ denote a scalar input to the system. The state equations are

$$x_{i+1} = \Phi x_i + \gamma y_i + gu_i \quad (17)$$

where $\gamma$ is an $n \times 1$ vector of unknowns. The Kalman filter for this system is

$$\hat{x}_{i+1/i} = \Phi\hat{x}_{i/i-1} + \gamma y_i + \Phi k\nu_i \quad (18)$$

Now we have $n$ extra unknowns $\gamma$, but Eqs. (12–16) remain unchanged. The extra stationarity conditions are

$$\frac{\partial J}{\partial \gamma} = \sum_{i=0}^{N-1}\lambda_{i+1}y_i = 0 \quad (19)$$

It has been shown by Åström and Wensmark[2] that certain special conditions must be imposed on $y_i$ in order to obtain consistent and unbiased estimates of $\theta$. In case, the process noise $y_i$ is zero, these conditions require that the input $y_i$ be sufficiently varied and persistent so that all the modes of system are excited.

## IV. Numerical Solution

In general, Eqs. (3–16) have to be solved numerically. A large number of methods are available, out of which gradient methods are particularly attractive. Here we use Davidon's Conjugate Gradient Method as given in Fletcher and Powell.[17] This method also gives us the inverse of the Hessian matrix of second partial derivatives viz.

$$\left[\frac{\partial^2 J}{\partial\theta_i\partial\theta_j}\right]^{-1}$$

which is directly related to the variance of the maximum likelihood estimates.[2]

Example: A fourth-order system representing the short period dynamics and the first bending mode of a missile was simulated on a digital computer using the following values[22]

$$\varphi_1 = -0.656, \varphi_2 = 0.784, \varphi_3 = -0.18, \varphi_4 = 1.0$$

$$\hat{x}_0 = 0, g^T / [0,1,0,1], q = 1.0, r = 0.25$$

**Table 1 Parameter estimates for a fourth-order example**

| Iteration | $J \times 10^{-3}$ | $b$ | $\varphi_1$ | $\varphi_2$ | $\varphi_3$ | $\varphi_4$ | $k_1$ | $k_2$ | $k_3$ | $k_4$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | — | — | — | — | — | — | — | — | — | — |
| Results from correlation technique, Ref. 10 | | | | | | | | | | |
| 1 | −1.0706 | 2.3800 | −0.5965 | 0.8029 | −0.1360 | 0.8696 | 0.6830 | 0.2837 | 0.4191 | 0.8207 |
| M. L. estimates using 1000 points | | | | | | | | | | |
| 2 | −1.0660 | 2.3811 | −0.5938 | 0.8029 | −0.1338 | 0.8759 | 0.6803 | 0.2840 | 0.4200 | 0.8312 |
| 3 | −1.0085 | 2.4026 | −0.6054 | 0.7452 | −0.1494 | 0.9380 | 0.6304 | 0.2888 | 0.4392 | 1.0311 |
| 4 | −0.9798 | 2.4409 | −0.6036 | 0.8161 | −0.1405 | 0.8540 | 0.6801 | 0.3210 | 0.6108 | 1.1831 |
| 5 | −0.9785 | 2.4412 | −0.5999 | 0.8196 | −0.1370 | 0.8580 | 0.6803 | 0.3214 | 0.6107 | 1.1835 |
| 6 | −0.9771 | 2.4637 | −0.6014 | 0.8086 | −0.1503 | 0.8841 | 0.7068 | 0.3479 | 0.6059 | 1.2200 |
| 7 | −0.9769 | 2.4603 | −0.6023 | 0.8130 | −0.1470 | 0.8773 | 0.7045 | 0.3429 | 0.6106 | 1.2104 |
| 8 | −0.9744 | 2.5240 | −0.6313 | 0.8105 | −0.1631 | 0.9279 | 0.7990 | 0.3756 | 0.6484 | 1.2589 |
| 9 | −0.9743 | 2.5241 | −0.6306 | 0.8108 | −0.1622 | 0.9296 | 0.7989 | 0.3749 | 0.6480 | 1.2588 |
| 10 | −0.9734 | 2.5270 | −0.6374 | 0.7961 | −0.1630 | 0.9505 | 0.7974 | 0.3568 | 0.6378 | 1.2577 |
| 11 | −0.9728 | 2.5313 | −0.6482 | 0.7987 | −0.1620 | 0.9577 | 0.8103 | 0.3443 | 0.6403 | 1.2351 |
| 12 | −0.9720 | 2.5444 | −0.6602 | 0.7995 | −0.1783 | 0.9866 | 0.8487 | 0.3303 | 0.6083 | 1.2053 |
| 13 | −0.9714 | 2.5600 | −0.6634 | 0.7919 | −0.2036 | 1.0280 | 0.8924 | 0.3143 | 0.6014 | 1.2054 |
| 14 | −0.9711 | 2.5657 | −0.6624 | 0.7808 | −0.2148 | 1.0491 | 0.9073 | 0.3251 | 0.6122 | 1.2200 |
| M. L. estimates using 100 points | | | | | | | | | | |
| 30 | −0.9659 | 2.620 | −0.6094 | 0.7663 | −0.1987 | 1.0156 | 1.24 | 0.136 | 0.454 | 1.103 |
| Actual values | | | | | | | | | | |
| | −0.94 | 2.557 | −0.6560 | 0.7840 | −0.1800 | 1.0000 | 0.8937 | 0.2957 | 0.6239 | 1.2510 |
| Estimates of standard deviation using 1000 points | | | | | | | | | | |
| | | 0.0317 | 0.0277 | 0.0247 | 0.0275 | 0.0261 | 0.0302 | 0.0323 | 0.0302 | 0.029 |
| Estimates of standard deviation using 100 points | | | | | | | | | | |
| | | 0.149 | 0.104 | 0.131 | 0.084 | 0.184 | 0.303 | 0.092 | 0.082 | 0.09 |

Using covariance Eqs. (5–7) optimum values of $k$ and $b$ were calculated and are shown in Table 1. It was assumed that $\hat{x}_0$ was known exactly. The starting values for the maximum likelihood scheme were obtained using a correlation technique given in Ref. 10. The results of successive iterations are shown in Table 1. The variances of the estimates obtained from the matrix of second partial derivatives are also given. For comparison purposes, results obtained by using 1000 data points and 100 data points are given.

## V. Comments and Generalizations

1) The convergence rate of the gradient procedure can be improved by using a second order gradient or Newton-Raphron method. This would require evaluation of the Hessian Matrix

$$\left[\frac{\partial^2 J}{\partial\theta_i \partial\theta_j}\right]$$

during each iteration. The computations can be simplified by using results of Wilkie and Perkins[18] for evaluating the sensitivity functions viz.

$$\partial\hat{x}_{i/i-1}/\partial\theta \text{ and } \partial\lambda_i/\partial\theta$$

which are needed in calculating the Hessian matrix. (Only 2n equations need to be solved.)

2) The results of Secs. 2 and 3 can be extended to multi-input multi-output systems by using appropriate canonical forms.[20] The numerical problems due to multiple relative maxima are however more serious. An extension to non-linear systems is also possible.[21]

3) If $\Phi$ and $h$ are prespecified with some unknowns, $\beta_i$, then we must compute $\partial J/\partial\beta_i$. This can be done directly or as follows:

$$\frac{\partial J}{\partial \beta_i} = \sum_{j=1}^{n} \frac{\partial J}{\partial \varphi_j} \cdot \frac{\partial \varphi_j}{\partial \beta_i}$$

Notice that $\partial\varphi_j/\partial\beta_i$ can be obtained from Leverrier's algorithm.[18]

4) In the derivation of maximum likelihood estimates, we assumed that no a priori knowledge about $\theta$ was available. We now assume that $\hat{\theta}_0$ is an a priori estimate of $\theta$ and that the a priori distribution of $\theta$ is known. For example, if $\hat{\theta}_0$ is the maximum likelihood estimate based on a large previous sample, $\theta$ is approximately Gaussian with mean $\hat{\theta}_0$ and a known covariance matrix $P_0$.[2,9]

The a posteriori estimate of $\theta$ can be obtained using Bayes' formula

$$p(\theta/Z) = p(Z/\theta)p(\theta)/p(Z)$$

Now we maximize $p(Z/\theta)p(\theta)$ with respect to $\theta$ to obtain a posteriori estimates. The case in which $p(\theta)$ is Gaussian, the gradient $\partial J/\partial\theta$ gets modified by the addition of the term

$$P_0^{-1}(\hat{\theta}_0 - \theta)$$

The other equations viz. Eqs. (3) and (12) remain unchanged.

5) It can be shown that the Equation Error Method and the Output Error Method[11,20] are special cases of the Maximum Likelihood Method. Consider the case in which the process noise $u_i$ in Eq. (1) is zero. The steady-state Kalman filter gain $k$ for this case tends to zero and the likelihood function $J(\theta)$ reduces to the output error criterion. Similarly when the measurement noise $v_i$ is zero and $g$ is of the special form

$$g = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

the system of Eq. (1) and (2) reduces to an autoregressive series for which the Equation Error Method gives the maximum likelihood estimates.[5,10]

## VI. Conclusions

By using a Kalman filter representation, a new set of equations is derived for the maximum likelihood identification of stochastic linear dynamic systems. The adjoint set of equations is found to be the same as the smoothing equations for such systems. A fourth-order example is solved using Davidon's conjugate gradient method. Various extensions and special cases of the method are considered.

## Appendix

In analogy with the discrete case, the following set of equations is derived for the continuous case

$$\dot{\hat{x}} = F\hat{x} + k\nu \tag{A1}$$

$$\dot{\lambda} = -(F - kh^T)^T\lambda + h\nu, \ \lambda(T) = 0 \tag{A2}$$

$$\nu = z - h^T\hat{x} \tag{A3}$$

$$J = \int_0^T \left[ z(t)\hat{x}_1(t) - \frac{1}{2}\hat{x}_1^2(t) \right] dt \tag{A4}$$

$$\frac{\partial J}{\partial \varphi} = \int_0^T \lambda_n(t)\hat{x}(t)dt = 0 \tag{A5}$$

$$\frac{\partial J}{\partial k} = \int_0^T \nu(t)\lambda(t)dt = 0 \tag{A6}$$

$$\partial J/\partial x_0 = \lambda(0) = 0 \tag{A7}$$

where $\hat{x}_1$ denotes the first element of vector $\hat{x}$ and $\lambda_n$ denotes the $n$th element of vector $\lambda$.

Notice that in order to avoid certain mathematical problems, $J$ is taken as the likelihood ratio for the hypotheses[14]

$$H_1: \ z = h^Tx + v \tag{A8}$$

and

$$H_0: \ z = v \tag{A9}$$

The integrals with respect to the white noise process $\nu(t)$ are interpreted in the Ito sense.[19]

$r$ is obtained as

$$r = \frac{1}{N\Delta}\sum_{i=1}^{N} dw_i^2 \tag{A10}$$

where

$$dw_i = \int_{t_{i-\Delta}}^{t_i} \nu(t)dt \tag{A11}$$

and the time interval $[O,T]$ has been divided into $N$ equal parts at $t_0, t_1, \ldots, t_N$

$$t_i - t_{i-1} = \Delta$$

## References


[1] Bellman, R., Kagiwada, H., and Kalaba, R., "Quasilinearization, System Identification and Prediction," Memorandum RM-3812-PR, 1963, Rand Corp., Santa Monica, Calif.

[2] Åström, K. J. and Wensmark, S., "Numerical Identification of Stationary Time Series," Sixth International Instruments and Measurements Congress, Sept. 1964, Section on Automatic Control, Stockholm, Sweden.

[3] Kashyap, R. L., "Maximum Likelihood Identification of Stochastic Linear Dynamic Systems," *IEEE Transactions on Automatic Control*, Feb. 1970, pp. 25–34.

[4] Bartlett, M. S., *Stochastic Processes*, Cambridge University Press, England, 1961.

[5] Box, G. E. P. and Jenkins, G. M., *Statistical Models for Forecasting and Control*, Holden Day, San Francisco, Calif., June 1970.

[6] Zadeh, L. A. and Desoer, A. A., *Linear System Theory—The State Space Approach*, McGraw-Hill, New York, 1963.

[7] Kalman, R. E., "A New Approach to Linear Filtering and Prediction Problems," *Transactions of the ASME, Journal of Basic Engineering*, Vol. 82, 1960, pp. 35–45.

[8] Geesey, R. and Kailath, T., "Comments on the Relationship of Alternate State-Space Representations in Linear Filtering Problems," *IEEE Transactions on Automatic Control*, Feb. 1969, pp. 113–114.

[9] Kendall, M. G. and Stuart, A., *The Advanced Theory of Statistics*, Vol. 2, Hafner Publishing Co., New York, 1961.

[10] Mehra, R. K., "On-Line Identification of Linear Dynamic Systems with Applications to Kalman Filtering," *Joint Automatic Control Conference*, June 24–26, 1970, Atlanta, Ga.

[11] Taylor, L. W. and Kenneth, W. I., "A Modified Newton-Raphson Method for Determining Stability Derivatives From Flight Data," *2nd International Conference on Computing Methods in Optimization Problems*, Sanremo, Italy, Sept. 9–13, 1968.

[12] Bryson, A. E. and Ho, Y. C., *Applied Optimal Control*, Blaisdell, Waltham, Mass., 1969.

[13] Mehra, R. K., "On the Identification of Variances and Adaptive Kalman Filtering," *Joint Automatic Control Conference*, Boulder, Colo., Aug. 1969, also *IEEE Transactions on Automatic Control*, April, 1970.

[14] Schweppe, F. C., "Evaluation of Likelihood Functions for Gaussian Signals," *IEEE Transactions of the Information Theory*, Jan. 1965.

[15] Kailath, T., "An Innovation Approach to Least Squares Estimation—Parts I and II," *IEEE Transactions on Automatic Control*, Dec. 1968, pp. 646–661.

[16] Mehra, R. K., "An Algorithm to Solve Matrix Equations $PH^T = G$ and $P = \Phi P \Phi^T + \Gamma\Gamma^T$," *IEEE Transactions on Automatic Control*, Oct. 1970.

[17] Fletcher, R. and Powel, M. J. D., "A Rapid Descent Method for Minimization," *Computer Journal*, Vol. 6, No. 2, 1963.

[18] Wilkie, D. F. and Perkins, W. R., "Essential Parameters in Sensitivity Analysis," *Automatica*, Vol. 5, No. 2, March 1969.

[19] Ito, K., "On Stochastic Differential Equations," *Mem. American Mathematical Society*, No. 4, 1951, pp. 1–51.

[20] Denery, D. G., "An Identification Algorithm Which Is Insensitive to Initial Parameter Estimates," AIAA Paper 70-34, New York, 1970.

[21] Mehra, R. K., "Maximum Likelihood Identification of Aircraft Parameters," *Joint Automatic Control Conference*, June 24–26, 1970, Atlanta, Ga.

[22] Ho, Y. C. and Lee, R. C. K., "Identification of Linear Dynamic Systems," *Joint Information and Control*, Vol. 8, Feb. 1965, pp. 93–110.