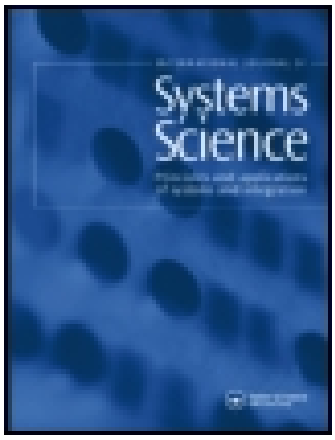


International Journal of Systems Science
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tsys20

Optimum system modelling using recent gradient methods

J. W. BANDLER [a], N. D. MARKETTOS [a] & N. K. SINHA [a]
[a] Department of Electrical Engineering and Communications Research Laboratory, McMaster University, Hamilton, Ontario, Canada
Published online: 08 Mar 2007.

To cite this article: J. W. BANDLER, N. D. MARKETTOS & N. K. SINHA (1973) Optimum system modelling using recent gradient methods, International Journal of Systems Science, 4:1, 33-44, DOI: 10.1080/00207727308919993

To link to this article: http://dx.doi.org/10.1080/00207727308919993

INT. J. SYSTEMS SCI., 1973, VOL. 4, NO. 1, 33–44

Optimum system modelling using recent gradient methods

J. W. BANDLER, N. D. MARKETTOS and N. K. SINHA
Department of Electrical Engineering and Communications Research Laboratory, McMaster University, Hamilton, Ontario, Canada

[Received 4 January 1972]

Three gradient optimization methods have been applied to the problem of determining an optimum low-order model for a high-order system. A comparison has been made between the three methods (the highly regarded Fletcher–Powell method, the new Fletcher method and the Jacobson–Oksman method) by obtaining second- and third-order models of a seventh-order system so that the step-response of the model approximates that of the system in the least-squares sense.

1. Introduction

In a recent paper (Sinha and Bereznai 1971), the derivation of optimum low-order models for a high-order system was discussed, with respect to any specified criterion. In particular, the pattern search algorithm (Hooke and Jeeves 1961) was applied to obtain second-, third- and fourth-order models of a seventh-order system, using a number of different criteria for optimization. Because of the poor convergence properties of the pattern search algorithm, it is worth while investigating the application of efficient gradient methods to this problem. One of the disadvantages of gradient methods is the necessity of determining gradients. However, in this case, it is not a serious problem since the dominant part of the expression consists of the partial derivatives of the response of the low-order model with respect to its parameters, and this is the same for any high-order system and any error criterion.

Among the available gradient methods, that due to Fletcher and Powell (1963) is highly regarded. Two other methods have been proposed recently (Fletcher 1970, Jacobson and Oksman 1970). Therefore, it was considered very desirable to make a comparative study of the application of these three methods to the problem of determining optimum low-order models for a given high-order system.

To date, as far as the authors are aware, no comparative study of these three gradient methods has appeared in the literature. Since two of the methods are very recent, and at least one of them is not very widely known, a brief review of these methods will be presented in this paper. This will be followed by the application of the methods to the determination of second- and third-order models of a seventh-order system minimizing the sum of squares of the error between the two responses obtained at a specified number of sampling instants over a given interval. Minimization of the maximum error between the two responses will be considered in a subsequent paper.

2. The new Fletcher method

To give a basic description of the new Fletcher method it is necessary to briefly review the Fletcher–Powell method first. Suppose that it is desired to

minimize a function $F(\mathbf{x})$ of an n-dimensional vector $\mathbf{x}$, given by

$$\mathbf{x} = [x_1\, x_2 \ldots x_n]^{\mathrm{T}}, \tag{1}$$

where the superscript T stands for transpose.

Let $\mathbf{g}(\mathbf{x})$ be the gradient of $F(\mathbf{x})$ with respect to $\mathbf{x}$ and let $\mathbf{G}$ denote the corresponding Hessian matrix. The inverse Hessian $\mathbf{G}^{-1}$ will be approximated by the matrix $\mathbf{H}$. Define

$$\boldsymbol{\delta} = \Delta\mathbf{x} \tag{2}$$

as the increment in $\mathbf{x}$, which is the correction made in the parameter vector $\mathbf{x}$ in order to decrease $F(\mathbf{x})$.

The main feature of the Fletcher–Powell method is that the increment at the ith iteration

$$\boldsymbol{\delta} = \alpha\mathbf{s} \tag{3}$$

is taken along the direction

$$\mathbf{s} = -\mathbf{H}\mathbf{g}, \tag{4}$$

where α in (3) is that value of λ which minimizes $F(\mathbf{x} + \lambda\mathbf{s})$ along the direction of $\mathbf{s}$. In practice, α is determined by linear search. The matrix $\mathbf{H}$ is updated at each iteration using the formula

$$\mathbf{H}^{i+1} = \mathbf{H}^i + \frac{\boldsymbol{\delta}\boldsymbol{\delta}^{\mathrm{T}}}{\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{\gamma}} - \frac{\mathbf{H}^i\boldsymbol{\gamma}\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{H}^i}{\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{H}^i\boldsymbol{\gamma}}, \tag{5}$$

where

$$\boldsymbol{\gamma} = \mathbf{g}^{i+1} - \mathbf{g}^i \tag{6}$$

and the superscript i denotes the value at the ith iteration. Initially $\mathbf{H}$ is taken as the unit matrix, so that $\mathbf{H}^i$ is positive definite for all i.

This method has the property of quadratic termination, that is, for a quadratic function the minimum can be located in at most n iterations. However, it depends on accurate location of the minimum along each direction of search. This is done by cubic interpolation which, although it is the most efficient method of search, requires several function and gradient evaluations.

The new Fletcher method dispenses with the linear search. The property of quadratic termination, which depends on linear search, is replaced by a property which requires, for quadratic functions, that the eigenvalues of $\mathbf{H}$ tend monotonically towards those of $\mathbf{G}^{-1}$. Also ΔF, the decrease in F, must be sufficiently large to guarantee ultimate convergence. This is taken care of by the following test. The change ΔF in F on an iteration would be expected by Taylor's series expansion to be approximately $\mathbf{g}^{\mathrm{T}}\boldsymbol{\delta}$ for a small $\boldsymbol{\delta}$, but much less than $\mathbf{g}^{\mathrm{T}}\boldsymbol{\delta}$ in absolute value when the position of the minimum along a line is overestimated. The change in F relative to $\mathbf{g}^{\mathrm{T}}\boldsymbol{\delta}$ cannot become arbitrarily small if

$$\frac{\Delta F}{\mathbf{g}^{\mathrm{T}}\boldsymbol{\delta}} \geqslant \mu, \tag{7}$$

where $0 < \mu \ll 1$, a pre-assigned small quantity set at 0·0001. If corrections are determined by

$$\boldsymbol{\delta} = -\lambda\mathbf{H}\mathbf{g}, \tag{8}$$

then trying values of $\lambda = 1, w, w^2, w^3, \ldots$ for $w = 0{\cdot}1$ will eventually produce a $\boldsymbol{\delta}$ that satisfies eqn. (7).

To avoid ill-conditioning in $\mathbf{H}$, a new updating formula is introduced

$$\mathbf{H}^{i+1} = \mathbf{H}^i - \frac{\boldsymbol{\delta}\boldsymbol{\gamma}^T\mathbf{H}^i}{\boldsymbol{\delta}^T\boldsymbol{\gamma}} - \frac{\mathbf{H}^i\boldsymbol{\gamma}\boldsymbol{\delta}^T}{\boldsymbol{\delta}^T\boldsymbol{\gamma}} + \left(1 + \frac{\boldsymbol{\gamma}^T\mathbf{H}^i\boldsymbol{\gamma}}{\boldsymbol{\delta}^T\boldsymbol{\gamma}}\right)\frac{\boldsymbol{\delta}\boldsymbol{\delta}^T}{\boldsymbol{\delta}^T\boldsymbol{\gamma}}; \tag{9}$$

denote $\mathbf{H}^{i+1}$ in (9) as $\mathbf{H}_1{}^{i+1}$ and $\mathbf{H}^{i+1}$ in (5) as $\mathbf{H}_2{}^{i+1}$. The formula in (9) has the property that the eigenvalues of $\mathbf{H}$ tend monotonically to those of $\mathbf{G}^{-1}$. Note, however, that $\mathbf{H}_1{}^{i+1}$ does not necessarily replace $\mathbf{H}_2{}^{i+1}$, as one of the two is chosen for updating on the basis of the following test. If

$$\boldsymbol{\delta}^T\boldsymbol{\gamma} \geqslant \boldsymbol{\gamma}^T\mathbf{H}^i\boldsymbol{\gamma}, \tag{10}$$

then $\mathbf{H}_1{}^{i+1}$ is used ; otherwise $\mathbf{H}_2{}^{i+1}$ is used.

The algorithm is terminated when each component of $\boldsymbol{\delta}$ is less than ϵ which was set at $1{\cdot}0 \times 10^{-6}$.

3. The Jacobson–Oksman method

This method is based on homogeneous rather than quadratic functions. A consequence of this is that convergence is obtained in $n+2$ steps for a homogeneous function of the form

$$F(\mathbf{x}) = \frac{1}{\theta}(\mathbf{x} - \hat{\mathbf{x}})^T\mathbf{g}(\mathbf{x}) + F(\hat{\mathbf{x}}), \tag{11}$$

where $\mathbf{x}$ is an n-dimensional parameter vector as before, $\mathbf{g}(\mathbf{x})$ is the gradient of $F(\mathbf{x})$ as defined previously, θ is the degree of homogeneity and $\hat{\mathbf{x}}$ is the location of the minimum of $F(\mathbf{x})$. On the other hand, a quadratic objective function may be expressed as

$$F(\mathbf{x}) = \tfrac{1}{2}(\mathbf{x} - \hat{\mathbf{x}})^T\mathbf{Q}(\mathbf{x} - \hat{\mathbf{x}}) + F(\hat{\mathbf{x}}), \tag{12}$$

where $\mathbf{Q}$ is a constant positive definite matrix. Thus, it will be seen that eqn. (12) is a special case of the homogeneous eqn. (11) with $\theta = 2$.

The basis of the method will now be discussed. Equation (11) may be arranged as

$$\hat{\mathbf{x}}^T\mathbf{g}(\mathbf{x}) + \theta F(\mathbf{x}) - w = \mathbf{x}^T\mathbf{g}(\mathbf{x}), \tag{13}$$

where

$$w = \theta F(\hat{\mathbf{x}}). \tag{14}$$

Let

$$\left.\begin{aligned} v &\triangleq \mathbf{x}^T\mathbf{g}(\mathbf{x}), \\ \mathbf{y} &\triangleq [\mathbf{g}^T(\mathbf{x})\ F(\mathbf{x})\ -1]^T, \\ \boldsymbol{\alpha} &\triangleq [\hat{\mathbf{x}}^T\quad \theta\quad w]^T. \end{aligned}\right\} \tag{15}$$

For some point $\mathbf{x}_{i+1}$, eqn. (13) becomes

$$v_{i+1} = \boldsymbol{\alpha}^T\mathbf{y}_{i+1} = \mathbf{y}_{i+1}{}^T\boldsymbol{\alpha}, \tag{16}$$

where $\boldsymbol{\alpha}$ and $\mathbf{y}$ are $(n+2)$-vectors. The vector $\boldsymbol{\alpha}$ contains the unknowns $\hat{\mathbf{x}}$ and w and must be determined. If we evaluate v and $\mathbf{y}$ at $n+2$ distinct points

$\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n+2}$, so that the resultant $\mathbf{y}_i$'s are linearly independent, we have

$$\begin{bmatrix} \mathbf{y}_1^{\mathrm{T}} \\ \mathbf{y}_2^{\mathrm{T}} \\ \vdots \\ \mathbf{y}_{n+2}^{\mathrm{T}} \end{bmatrix} \boldsymbol{\alpha} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_{n+2} \end{bmatrix}, \tag{17}$$

or, in the matrix form

$$\mathbf{Y}\boldsymbol{\alpha} = \mathbf{v}. \tag{18}$$

Since the $\mathbf{y}_i$'s are linearly independent, the matrix $\mathbf{Y}$ is non-singular, giving

$$\boldsymbol{\alpha} = \mathbf{Y}^{-1}\mathbf{v}. \tag{19}$$

Matrix inversion can be avoided by using a recursive formula as new $\mathbf{y}_i$ and v_i are evaluated. Starting with $\mathbf{P}_0 = \mathbf{I}$, an $(n+2)\times(n+2)$ identity matrix and $\mathbf{v}_0 = \boldsymbol{\alpha}_0$, an arbitrary initial guess, successive estimates of the vector $\boldsymbol{\alpha}$ are given by

$$\boldsymbol{\alpha}_{i+1} = \boldsymbol{\alpha}_i + \frac{\mathbf{P}_i\mathbf{e}_{i+1}(v_{i+1} - \mathbf{y}_{i+1}^{\mathrm{T}}\boldsymbol{\alpha}_i)}{\mathbf{y}_{i+1}^{\mathrm{T}}\mathbf{P}_i\mathbf{e}_{i+1}}, \tag{20}$$

where $\mathbf{e}_{i+1}$ is a unit $(n+2)$-vector having unity as the $(i+1)$th element and zero elsewhere, and where the $\mathbf{P}_i$ are obtained successively from the formula

$$\mathbf{P}_{i+1} = \mathbf{P}_i - \frac{\mathbf{P}_i\mathbf{e}_{i+1}(\mathbf{y}_{i+1}^{\mathrm{T}}\mathbf{P}_i - \mathbf{e}_{i+1}^{\mathrm{T}})}{\mathbf{y}_{i+1}^{\mathrm{T}}\mathbf{P}_i\mathbf{e}_{i+1}}. \tag{21}$$

It can be shown that the algorithm finds the minimum, the degree of homogeneity and the value of the minimum after $n+2$ iterations.

4. Statement of the modelling problem

The problem of determining optimum low-order models for a given high-order system may be stated as follows: It is required to find the transfer function of a model of a given order the response of which is the best approximation to the response of the actual system to a particular input in a specified sense. For this work, the input was taken as a unit step.

Often it is desirable to impose certain constraints on the model. For instance, it may be required that the initial slope of the step-response of the model be the same as that of the high-order system, and that the steady-state values of the two responses be equal. In general, the transfer function of the model may be written as

$$H(s) = \frac{b_m s^m + b_{m-1}s^{m-1} + \ldots + b_1 s + b_0}{s^n + a_{n-1}s^{n-1} + \ldots + a_1 s + a_0}, \tag{22}$$

where m is less than or equal to n for physical systems. The problem is, therefore, the determination of the parameter vector $\boldsymbol{\phi}$, defined as

$$\boldsymbol{\phi} = [b_m\, b_{m-1} \ldots b_1\, b_0\, a_{n-1}\, a_{n-2} \ldots a_1\, a_0]^{\mathrm{T}}, \tag{23}$$

such that the step-response of the model approximates the step-response of the system in an optimal manner.

The choice of the criterion for optimization depends on the purpose for which the model is to be used, and the portion of the response on which emphasis should be placed. Thus, a least squares criterion may be used if it is desired to distribute the error. On the other hand, a weighted least-pth with a large p, or a minimax criterion may be used if most emphasis is placed on the maximum deviation (Bandler 1969). Also, if it is desired to make the approximating response look close to the actual response, the perpendicular error criterion (Bereznai and Sinha 1970) may be used, where the maximum perpendicular distance between the responses is minimized.

The optimum model is obtained by starting with an initial choice of the parameter vector and then using one of the three gradient methods to successively obtain the optimum choice. The use of these techniques makes it necessary to obtain the first partial derivatives of the objective function with respect to the various parameters. Analytical expressions for the gradient with respect to the parameters of the step-response of second- and third-order models are shown in the appendix.

5. Example

In order to compare the performance of the three minimization techniques on the modelling problem, it was decided to consider the seventh-order system used in a previous paper (Sinha and Bereznai 1971), namely, the control system for the pitch rate of a supersonic transport aircraft. The transfer function of the system is given by

$$G(s)=\frac{375000(s+0{\cdot}08333)}{s^7+83{\cdot}64s^6+4097s^5+70342s^4+853703s^3+2814271s^2+3310875s+281250} \tag{24}$$

with a steady-state value of 0·1111.

Second- and third-order models of this system were obtained, of the following forms

$$H_1(s)=\frac{Ea_0}{s^2+a_1s+a_0}, \tag{25}$$

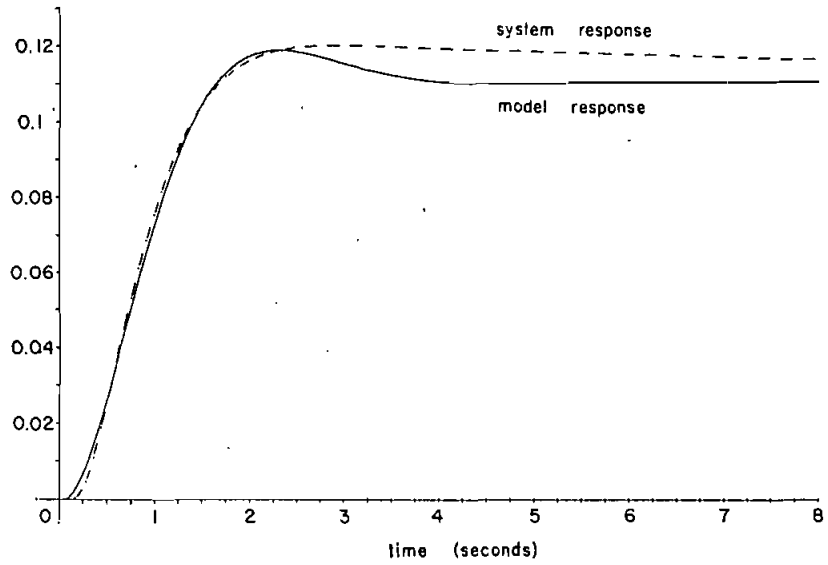
$$H_2(s)=\frac{b_1s+Ea_0}{s^2+a_1s+a_0}, \tag{26}$$

$$H_3(s)=\frac{b_2s^2+b_1s+Ea_0}{s^3+a_2s^2+a_1s+a_0}=\frac{x_5s^2+x_4s+Ex_1x_3}{(s+x_3)(s^2+x_2s+x_1)}. \tag{27}$$

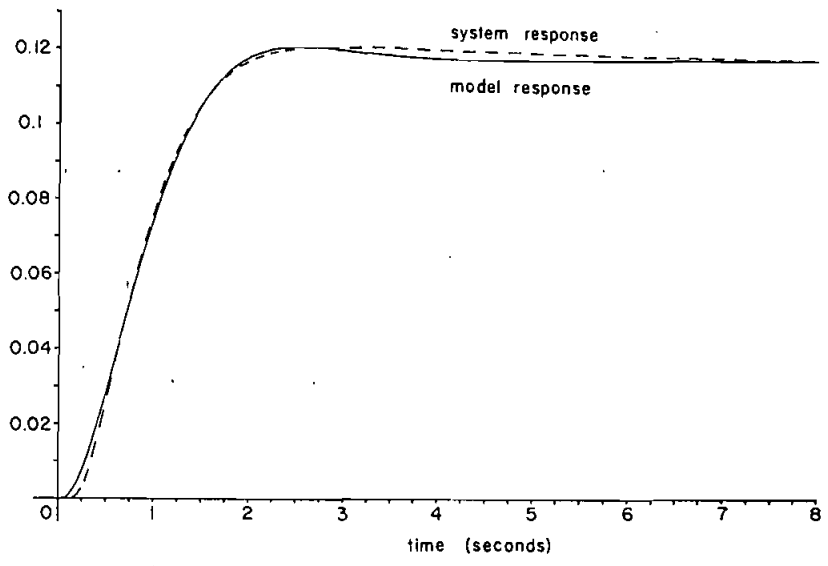
For each of the models, the parameter b_0 was made equal to Ea_0 in order that the steady-state response of the model could be equal to E, which was either the steady-state response of the system to a unit step or the value of that response at 8 sec. It was found necessary to introduce this constraint, as otherwise the problem did not have a unique solution. The response was approximated over an interval of 8 sec using 21 uniformly spaced samples.

First the two-parameter problem resulting from eqn. (25) was tried. The computer used was a CDC 6400 and typical C.P. times for this problem were about 4 sec. E was 0·1111. Three different starting points were considered

Fig. 1



(*a*)



(*b*)

(*a*) Response of the two-parameter model with steady-state value 0·1111. (*b*) Response of the two-parameter model with steady-state value 0·11706.

for each of the three methods and in every case the algorithms ultimately converged to the same optimum parameters, $a_0 = 3{\cdot}195912$ and $a_1 = 2{\cdot}281056$, with the optimum value of the objective function $7{\cdot}50758 \times 10^{-4}$ and the components of the gradient less than $1{\cdot}0 \times 10^{-9}$. Figure 1 (*a*) shows the corresponding response. Table 1 compares the number of function evaluations required for each method for the objective function to reach the value of $7{\cdot}50759 \times 10^{-4}$, this

value being $1{\cdot}0 \times 10^{-9}$ higher than the optimum ultimately obtained. From table 1 it can be seen that both the new Fletcher method and the Jacobson–Oksman method show an improvement over the Fletcher–Powell method, with the Jacobson–Oksman method being slightly better than the new Fletcher method.

Table 1. Number of function evaluations required to reach the objective function value $7{\cdot}50759 \times 10^{-4}$ for the two-parameter problem

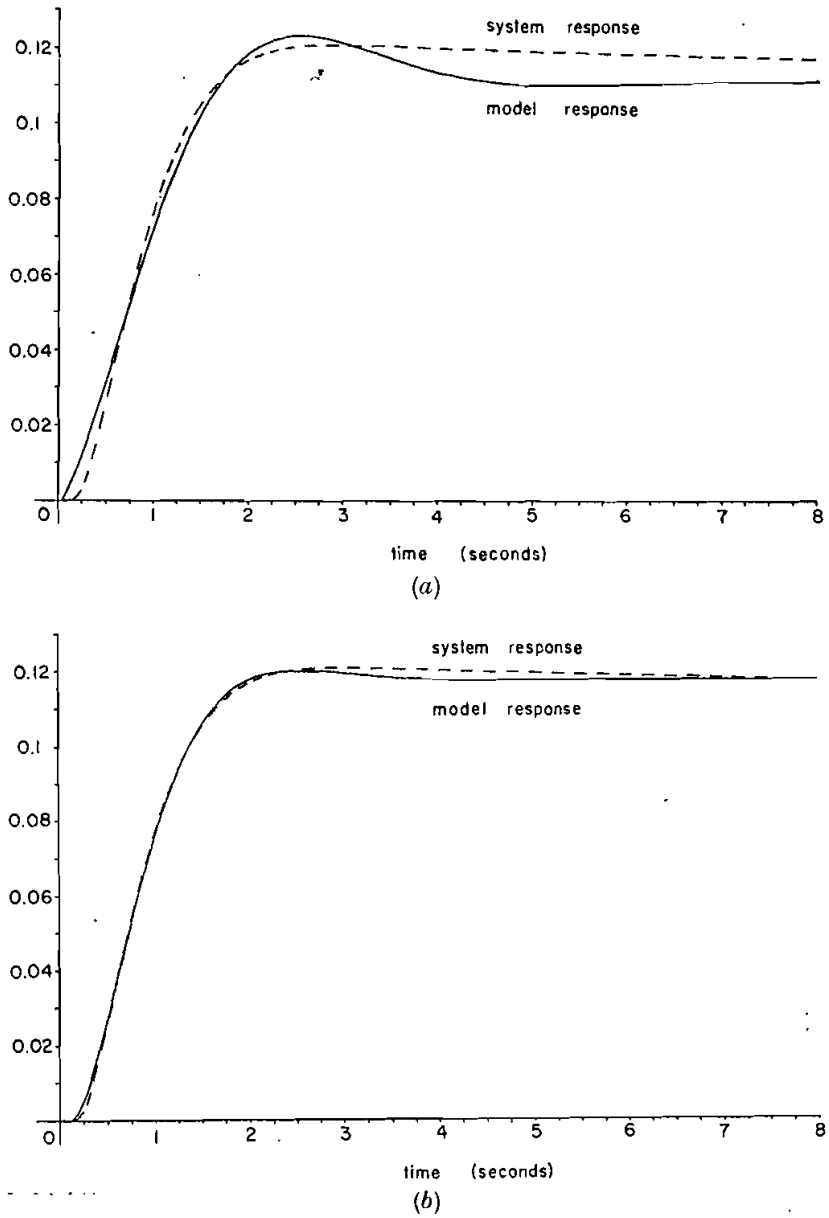
Starting point **x**	$a_0 = 3{\cdot}0$, $a_1 = 2{\cdot}0$	$a_0 = 0{\cdot}5$, $a_1 = 0{\cdot}5$	$a_0 = 1{\cdot}0$, $a_1 = 1{\cdot}0$
Jacobson–Oksman	21	19	14
New Fletcher	21	22	19
Fletcher–Powell	29	49	32

For the three-parameter problem represented by eqn. (26) with $E = 0{\cdot}1111$, typical computer C.P. times were about 6 sec on the CDC 6400. Again three different starting points were tried with each method and in each case the algorithms converged to the optimum $a_0 = 1{\cdot}997397$, $a_1 = 1{\cdot}660663$, $b_1 = 4{\cdot}370715 \times 10^{-2}$ with the minimum value of the objective function $1{\cdot}582215 \times 10^{-4}$ and the components of the gradient less than $1{\cdot}0 \times 10^{-9}$. Figure 2 (*a*) shows the corresponding response. Table 2 compares the number of function evaluations required to reach the objective function value of $1{\cdot}588225 \times 10^{-4}$ in each case. Here again it is seen that the Jacobson–Oksman and the new Fletcher methods are superior to the Fletcher–Powell method and the Jacobson–Oksman method is slightly better than the new Fletcher method. However, there was one case, not shown, when the Jacobson–Oksman method failed. This was due to a very large step that made the objective function too

Table 2. Number of function evaluations required to reach the objective function value $1{\cdot}582225 \times 10^{-4}$ for the three-parameter problem

Starting point **x** \ Method	Jacobson–Oksman	New Fletcher	Fletcher–Powell
1·0 1·0 1·0	39	27	60
0·5 2·0 4·0	39	76	274
1·0 0·5 0·1	29	35	58

Fig. 2



(*a*) Response of the three-parameter model with steady-state value 0·1111. (*b*) Response of the three-parameter model with steady-state value 0·11706.

large for the computer to handle. In situations like this, one is tempted to limit the step size. However, it was decided not to interfere with the algorithm in any way, as this might slow down the method at the beginning.

Since the approximation is over 8 sec and in that time the system does not reach steady state, the steady-state value of the model, E, can be fixed to the system response value at 8 sec. Figures 1 (*b*) and 2 (*b*) show the corresponding

Table 3. Number of function evaluations required to reach the objective function value of $1{\cdot}027406 \times 10^{-6}$ and central processor time if it is less than 64 sec, or the value of the objective function reached in the time limit of 64 sec

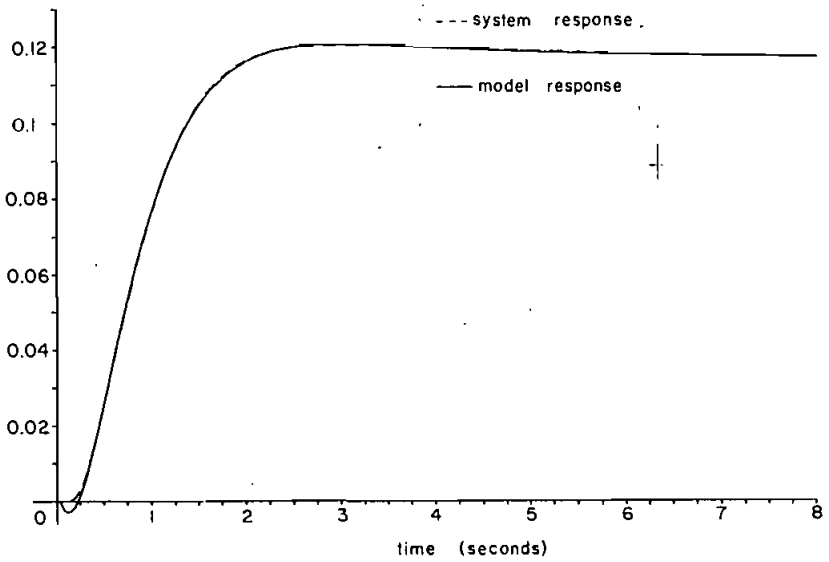
Starting point x \ Method	New Fletcher	Fletcher–Powell	Jacobson–Oksman
1·25 2·8 2·3 0·7 −0·1	$N=535$ $F=1{\cdot}027406\times10^{-6}$ C.P. Time = 48 sec	225 $1{\cdot}027406\times10^{-6}$ 22 sec	88 $F=1{\cdot}027952\times10^{-6}$
5·0 4·5 3·5 4·0 2·0	140 $1{\cdot}027406\times10^{-6}$ 16 sec	780 $1{\cdot}528774\times10^{-6}$ > 64 sec	
3·2 0·8 5·3 −2·6 2·1	465 $1{\cdot}027406\times10^{-6}$ 45 sec	745 $8{\cdot}449308\times10^{-6}$ > 64 sec	
1·0 1·5 3·0 4·0 5·0	298 $1{\cdot}027406\times10^{-6}$ 30 sec		
2·5 1·5 3·5 0·1 1·0	132 $1{\cdot}027406\times10^{-6}$ 15 sec	800 $1{\cdot}02883\times10^{-6}$ > 64 sec	

responses for the two- and three-parameter problems, respectively, with the value of E equal to 0·11706. For the two-variable case the optimum parameters were $a_0=3{\cdot}475712$, $a_1=2{\cdot}766808$ with $F=4{\cdot}750160\times10^{-5}$. For the three-variable case the optimum parameters were $a_0=3{\cdot}975313$, $a_1=3{\cdot}039002$, $b_1=-2{\cdot}087874\times10^{-2}$ with $F=2{\cdot}261484\times10^{-5}$.

For the five-parameter problem represented by eqn. (27), with $E=0{\cdot}11706$ a number of starting points were considered in an effort to get some convergence with the Jacobson–Oksman method, but with the exception of one case, it always failed. The failure was, computationally, due to the objective function becoming too large, as a result of large variations in the parameters. The new Fletcher method converged to the same optimum in every case, given by $x_1=1{\cdot}027405$, $x_2=2{\cdot}855360$, $x_3=2{\cdot}301252$, $x_4=0{\cdot}662057$ and $x_5=-0{\cdot}076045$. The Fletcher–Powell method was much slower, and in only

one case it converged within the central processor time limit of 64 sec ; this particular case being when the starting point was very close to the optimum. The results for five starting points are shown in table 3, where in addition to the number of function evaluations, the minimum value of the objective function obtained as well as the central processor time required are also given. The response for the optimum is shown in fig. 3.

Fig. 3



Response of the five-parameter model with steady-state value 0·11706.

6. Conclusion

From these examples it appears that both the new Fletcher method and the Jacobson–Oksman method converge much faster than the Fletcher–Powell method, as might have been expected. Although the Jacobson–Oksman method gave a slightly better performance than the new Fletcher method in a number of cases, it often failed, especially if the number of variable parameters was increased. This difficulty could perhaps be overcome by limiting the step-size but probably it may slow down the initial performance of the algorithm. As the new Fletcher method gave consistently good results, the authors feel that it is the most suitable algorithm for problems of this type.

Acknowledgments

This work was supported by grants A7239 and A3374 from the National Research Council of Canada. The authors are particularly grateful to Dr. D. H. Jacobson and Dr. R. Fletcher for making available their computer programmes.

Appendix 1. The two-parameter problem

$$H_1(s)=\frac{C(s)}{R(s)}=\frac{Ea_0}{s^2+a_1s+a_0}. \tag{A 1}$$

For step input,

$$R(s)=\frac{1}{s}. \tag{A 2}$$

Hence,

$$C(s)=\frac{Ea_0}{s(s^2+a_1s+a_0)}=\frac{E}{s}-\frac{E(s+a_1)}{s^2+a_1s+a_0}, \tag{A 3}$$

$$\frac{\partial C(s)}{\partial a_0}=\frac{E(s+a_1)}{(s^2+a_1s+a_0)^2}, \tag{A 4}$$

$$\frac{\partial C(s)}{\partial a_1}=-\frac{Ea_0}{(s^2+a_1s+a_0)^2}. \tag{A 5}$$

The sensitivity function $\partial c(t)/\partial a_0$ and $\partial c(t)/\partial a_1$ can now be evaluated by taking the inverse Laplace transforms of eqns. (A 4) and (A 5). These can be easily obtained from standard tables. For example, if the poles of the transform function are complex, i.e.

$$s^2+a_1s+a_0=(s+\alpha)^2+\beta^2, \tag{A 6}$$

then from eqns. (A 4) and (A 5) we have

$$\frac{\partial c(t)}{\partial a_0}=\frac{E\exp(-\alpha t)}{2\beta^3}[(\alpha+\beta^2t)\sin\beta t-\alpha\beta t\cos\beta t], \tag{A 7}$$

$$\frac{\partial c(t)}{\partial a_1}=-\frac{Ea_0\exp(-\alpha t)}{2\beta^3}(\sin\beta t-\beta t\cos\beta t). \tag{A 8}$$

Appendix 2. The three-parameter problem

$$H_2(s)=\frac{b_1s+Ea_0}{s^2+a_1s+a_0}, \tag{A 9}$$

$$C(s)=\frac{b_1s+Ea_0}{s(s^2+a_1s+a_0)}=\frac{E}{s}-\frac{E(s+a_1)-b_1}{s^2+a_1s+a_0}, \tag{A 10}$$

$$\frac{\partial C(s)}{\partial a_0}=\frac{E(s+a_1)-b_1}{(s^2+a_1s+a_0)^2}, \tag{A 11}$$

$$\frac{\partial C(s)}{\partial a_1}=-\frac{b_1s+Ea_0}{(s^2+a_1s+a_0)^2}, \tag{A 12}$$

$$\frac{\partial C(s)}{\partial b_1}=\frac{1}{s^2+a_1s+a_0}. \tag{A 13}$$

The sensitivities in the time domain are again obtained by taking inverse Laplace transforms.

Appendix 3. The five-parameter problem

$$H_3(s)=\frac{x_5s^2+x_4s+Ex_1x_3}{(s+x_3)(s^2+x_2s+x_1)}, \tag{A 14}$$

$$C(s)=\frac{x_5s^2+x_4s+Ex_1x_3}{s(s+x_3)(s^2+x_2s+x_1)}$$

$$=\frac{E}{s}-\frac{E[s^2+(x_2+x_3)s+x_1+x_2x_3)]-x_5s-x_4}{(s+x_3)(s^2+x_2s+x_1)}, \tag{A 15}$$

$$\frac{\partial C(s)}{\partial x_1}=-\frac{x_5s+x_4-Ex_3(s+x_2)}{(s+x_3)(s^2+x_2s+x_1)^2}, \tag{A 16}$$

$$\frac{\partial C(s)}{\partial x_2}=-\frac{x_5s^2+x_4s+Ex_1x_3}{(s+x_3)(s^2+x_2s+x_1)^2}, \tag{A 17}$$

$$\frac{\partial C(s)}{\partial x_3}=-\frac{x_5s+x_4-Ex_1}{(s+x_3)^2(s^2+x_2s+x_1)}, \tag{A 18}$$

$$\frac{\partial C(s)}{\partial x_4}=\frac{1}{(s+x_3)(s^2+x_2s+x_1)}, \tag{A 19}$$

$$\frac{\partial C(s)}{\partial x_5}=\frac{s}{(s+x_3)(s^2+x_2s+x_1)}. \tag{A 20}$$

References

Bandler, J. W., 1969, *I.E.E.E. Trans. microw. Theory Tech.*, **17**, 533.
Bereznai, G. T., and Sinha, N. K., 1970, *Electron. Lett.*, **6**, 847.
Fletcher, R., 1970, *Computer J.*, **13**, 317.
Fletcher, R., and Powell, M. J. D., 1963, *Computer J.*, **6**, 163.
Hooke, R., and Jeeves, T. A., 1961, *J. Ass. Comput. Mach.*, **8**, 212.
Jacobson, D. H., and Oksman, W., 1970, Harvard Univ. Tech. Rep. No. 618.
Sinha, N. K., and Bereznai, G. T., 1971, *Int. J. Control*, **14**, 551.