R & D NOTES

Comparing Sensitivity Methods for Estimating Parameters in Nonlinear Ordinary Differential Equations

ALBERT J. BERGER

Rensselaer Polytechnic Institute
Troy, New York 12181

The problem of parameter estimation is of general importance to the chemical engineer. Typically parameter estimation problems arise when one attempts to fit various models to experimental observations. A very complex situation occurs when the model of the system takes the form of a vector set of differential equations. These equations can be linear or nonlinear ordinary or partial differential equations. In a general sense the techniques described in this paper, as well as other techniques, can be applied to any of these classes of problems (Seinfeld, 1969 and 1970).

The specific purpose of this note is to present the results of a computational study. In this study the modified Gauss-Newton (MGN) and Fletcher-Powell-Davidon (FPD) methods were compared for solving the nonlinear parameter estimation problem as applied to systems whose states can be modeled by a vector set of nonlinear ordinary differential equations. It is hoped that these comparisons may be helpful in the selection and designing of algorithms to solve parameter estimation problems.

PREVIOUS STUDIES

Bard (1970) presented an extensive comparison of many gradient methods, or alternatively referred to as sensitivity methods, for the estimation of parameters arising from systems modeled by nonlinear ordinary differential equations. Essentially the parameter estimation problem is cast into the form of a minimization problem in function space. Bard's criterion for comparison was based upon the minimization of the number of equivalent function evaluations. He found that the MGN method was superior to the FPD method. This conclusion was not at all obvious, since the FPD method had proved to be very useful and rapid for function minimization (Fletcher and Powell, 1963). The current study extends Bard's work by comparing the MGN and FPD methods on the bases of the size of the initial region in the parameter space in which convergence can be achieved and the effect of noisy data upon the ability of these algorithms to converge to the best estimate of the parameters. In addition, this study also includes the effect of limited state observability. These additional comparisons are important because the effects of a poor initial choice of a parameter vector, noisy experimental data, or limited state observability are present in many situations and can lead to severe computational difficulties.

PARAMETER ESTIMATION PROBLEM AND ESTIMATION METHODS

The reader is referred to the works of Bard (1970) and of Bekey (1970) which present in a general form the parameter estimation problem, and the MGN and FPD algorithms. Among the more pertinent assumptions incorporated in this study are:

1. The form of the model's ordinary differential equations is known explicitly.
2. Initial conditions for all the states are known exactly.
3. Observed states may have experimental errors associated with their measurement.
4. Calculations were based upon an unweighted sum of square objective function although other forms of the objective function such as weighted sum of squares, maximum likelihood, or Bayesian type could be used.

Both the MGN and FPD algorithms are iterative and at the $k + 1$ iteration step a new value for the parameter vector q is calculated:

$$\Delta q^k = q^{k+1} - q^k = -\alpha^k \underline{\underline{A}}^k g^k \qquad (1)$$

The new estimate of q should result in a lower value of the objective function. The two algorithms are similar in that the scalar α^k in Equation (1) is determined by a one-dimensional search in the $-\underline{\underline{A}}^k g^k$ direction. Also, both require the calculation of a gradient vector of the objective function with respect to the parameter vector g^k. Analytical gradients can be obtained by the simultaneous integration of the sensitivity and state differential equations. The two algorithms differ in the elements of the $\underline{\underline{A}}$ matrix. The MGN method $\underline{\underline{A}}$ is an approximate inverse Hessian. It is an approximate representation because certain terms which occur in the exact Hessian involve residuals and are assumed to be small, and therefore neglected in calculating $\underline{\underline{A}}$ of the MGN method. The FPD method $\underline{\underline{A}}$ is determined iteratively, and at the minimum becomes equal to the true inverse Hessian.

In application of the MGN and FPD methods some

A. J. Berger is with the Department of Physiology, University of California, San Francisco, California 94122.

degrees of freedom are open. Among these degrees of freedom are choice of the stopping or convergence criterion, the form of one dimensional search, and how bounds are placed on parameter values. In the comparative study reported here these three features took the same form for each algorithm. For example, the stopping criterion was taken as the sum of the absolute values of relative change in the parameters being less than some preassigned level. The one-dimensional search fit the objective function at three points to a quadratic expression. A fit using only the value of the objective function and not its gradient increases computational speed because state equations only need be integrated and not the full state-sensitivity set. Finally, bounds were placed upon the parameter values by setting the parameter to the bound value if during the one dimensional search a parameter violated its bound.

EXAMPLE I. COMPETITIVE CONSECUTIVE SECOND ORDER REACTION

The reaction scheme for the irreversible competitive consecutive second-order reaction has the form

$$A + B \xrightarrow{k_1} C + E \tag{2}$$

$$A + C \xrightarrow{k_2} D + E \tag{3}$$

The following special set of initial conditions were assumed:

$$A(0) = 1 \quad B(0) = 1/2$$

$$C(0) = \quad D(0) = \quad E(0) = 0 \tag{4}$$

The resulting nonlinear ordinary differential equations represent a transient kinetic description for this system:

$$\frac{dA}{dt} = \{-k_1 + 2k_2\} AB - k_2A^2 \tag{5}$$

$$\frac{dB}{dt} = -k_1AB \tag{6}$$

Thus the system is a two-state variable two-parameter estimation problem. It was assumed that the true values of the parameters, k_1 and k_2, were both equal to one. Discrete experimental data were generated at 10 unequally spaced points in the domain $0 < t \leq 4$ by integrating Equations (5) and (6). Figure 1 shows the objective function contours for this system as a function of parameter values. This figure shows that the objective function has a characteristic banana shape which can lead to computational difficulties in locating the minimum at (1, 1). The approximate maximum region of convergence with the two algorithms is presented in Figure 2. This figure shows that the MGN method has a very much larger region of convergence than the FPD method.

Next, the effect upon the algorithms of noise in the observed data was assessed. Noisy data were produced using a random number generator. The noise was normally distributed with a mean of zero and variable percent standard deviation σ. Both algorithms were started from the same point, $k_1 = k_2 = 2$, within the domain of convergence shown in Figure 2. The computational results indicated that for $0\% \leq \sigma \leq 20\%$ the MGN algorithm, took approximately 6 iterations versus 12 iterations for the FPD algorithm, and about one half the computation time. In almost all cases studied, both algorithms converged to the same point with the same value of the objective function. These results are comparable with

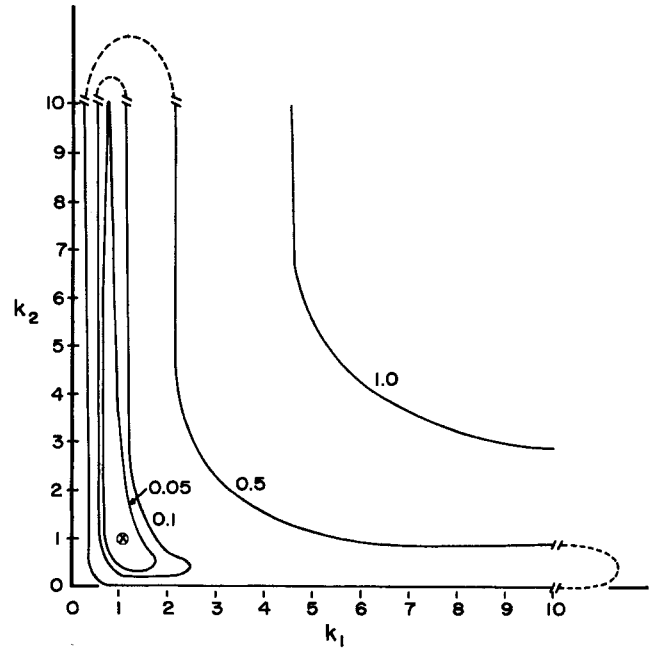

Fig. 1. Unweighted sum of squares objective function for Example I. Noise free observed states A and B. True parameter values $k_1 = k_2 = 1$. Ten discrete data points in the domain $0 < t \leq 4$.

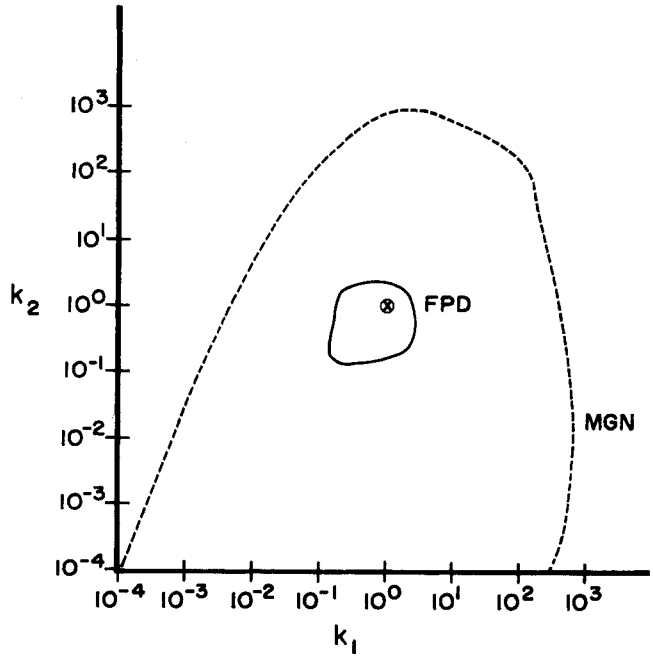

Fig. 2. Approximate maximum region of convergence for Example I. Noise free observed states A and B. True parameter values $k_1 = k_2 = 1$. Parameter bounds $10^{-4} \leq k_1, k_2 \leq 10^3$.

those of Bard (1970) for noise free data.

The final comparative test of these algorithms related to their ability to converge in the face of limited state observability. In this case noise free experimental data were used, and the initial point was again, $k_1 = k_2 = 2$. Two situations were analyzed. First where only A is observed, and secondly where only B is observed. In both cases the MGN method rapidly converged to approximately $k_1 = k_2 = 1$, yielding a value of the sum of square objective function of about 0.5×10^{-8}. On the other hand, the FPD method failed to converge to this point, and was only able to reduce the objective function to about 0.2×10^{-3} in both cases.

TABLE 1. CONVERGENCE AS A FUNCTION OF INITIAL PARAMETER VALUES—EXAMPLE II

Initial k_1	MGN No. of iterations	MGN Converged objective function	FPD No. of iterations	FPD Converged objective function
5	4	0.37×10^{-8}	4	0.19×10^{-6}
10	2	0.49×10^{-6}	2	0.54×10^{-6}
25	4	0.18×10^{-5}	Failed to converge	
50	6	0.45×10^{-8}	Failed to converge	
75	Failed to converge		Failed to converge	
100	5	0.53×10^{1}*	Failed to converge	
150	5	0.36×10^{-6}	Failed to converge	

* Converged to 126.

EXAMPLE II. THE SECOND-ORDER UNDAMPED OSCILLATOR

Recently Hwang and Seinfeld (1972) presented this example for the purpose of illustrating the Gauss-Newton algorithm. This example has only one unknown parameter, and in their formulation only $x_1(t)$ was observed. Thus this problem is one of limited state observability. As shown by Hwang and Seinfeld (see their Figure 1) the objective function possesses multiple local minima with a true minimum at π^2. The system is described by the following set of linear ordinary differential equations:

$$\frac{dx_1}{dt} = x_2 \tag{7}$$

$$\frac{dx_2}{dt} = -k_1 x_1 \tag{8}$$

The following set of initial conditions was assumed:

$$x_1(0) = 0 \quad x_2(0) = \pi \tag{9}$$

In the present study data were generated at 10 unequally spaced points between $0 < t \leq 1$ for a parameter value of $k_1 = \pi^2$.

Again, as shown in Table 1, the MGN method has a much larger region of convergence. But the MGN method cannot overcome the difficulties associated with local minima. Note that for initial parameter values of 75 and 100, convergence to the best value of the parameter is not achieved. But for an initial guess of 150, convergence to π^2 was achieved because the initial gradient directed the search toward π^2 and the local minimum at approximately 125 is passed entirely. Table 1 also shows that when the initial parameter values were quite close to the true value, the FPD method took about the same number of iteration steps as the MGN method. This is not surprising because when close to the minimum, both methods compute and use the same inverse Hessian and thus converge in a similar manner as indicated by Equation (1).

The effect of noisy data upon the MGN method was assessed for an initial parameter value of $k_1 = 25$, and $0\% \leq \sigma \leq 50\%$. In all cases about 4 iterations were required to achieve convergence, and the parameter rapidly converged to approximately π^2.

CONCLUSIONS AND DISCUSSION

In this paper two examples were studied in detail. Other reacting systems have been studied such as the pyrolytic dehydrogenation of benzene (Hougan and Watson, 1948) and the oxygenation of Hemoglobin (Gibson, 1959). Estimation results from these other systems confirm those presented here.

Based upon the results from this study it appears that the MGN algorithm is a very powerful tool for the solution of parameter estimation problems. These confirm and hopefully significantly extend results presented by Bard (1970). Compared with the FPD algorithm the MGN method is computationally faster when distant from the minimum. It has a large region of convergence, works well with noisy data, and with limited state observability. On the other hand, it is important to emphasize that these conclusions are merely based upon computational results, and while it is hoped these conclusions are generally true, general validity cannot be guaranteed. In addition, with any of these methods one cannot guarantee that a global minimum has been reached. This is one of the most vexing problems associated with finding extremals of nonlinear problems.

ACKNOWLEDGMENT

Acknowledgment is made to the donors of The Petroleum Research Fund, administered by The American Chemical Society, for support of this research. In addition, the aid of L. Johnson in performing some of the digital computations is gratefully appreciated.

NOTATION

$\underline{\underline{A}}$ = premultiplication matrix
A, B, C, D, E = reacting species and concentrations
g = gradient vector
k = superscripted iteration index
k_1, k_2 = parameters
q = parameter vector
t = time
x_1, x_2 = state variables

Greek Letters

α = one dimensional search constant
π = 3.14159
σ = percent standard deviation

LITERATURE CITED

Bard, Y., "Comparison of Gradient Methods for the Solution of Nonlinear Parameter Estimation Problems," *Siam. J. Numerical Analysis*, 7, 157 (1970).

Bekey, G. A., "System Identification–An Introduction and a Survey," *Simulation*, 15, 151 (1970).

Fletcher, R., and M. J. D. Powell, "A Rapidly Convergent Descent Method for Minimization," *Computer J.*, 6, 163 (1963).

Gibson, Q. H., "The Kinetics of Reactions Between Haemoglobin and Gases," *Prog. Biophys. Biophys. Chem.*, 9, 1 (1959).

Hartley, H. O., "The Modified Gauss-Newton Method for the Fitting of Nonlinear Regression Functions by Least Squares," *Technometrics*, 3, 269 (1961).

Hougan, O., and K. M. Watson, *Chemical Process Principles*, Vol. 3, Wiley, New York (1948).

Hwang, M., and J. H. Seinfeld, "A New Algorithm for the Estimation of Parameters in Ordinary Differential Equations," *AIChE J.*, 18, 90 (1972).

Seinfeld, J. H., "Identification of Parameters in Partial Differential Equations," *Chem. Eng. Sci.*, 24, 65 (1969).

———., "Nonlinear Estimation Theory," *Ind. Eng. Chem.*, 62, 32 (1970).

Manuscript received November 20, 1972; note accepted November 22, 1972.