OPTIMIZATION AND SENSITIVITY ANALYSIS FOR MULTIRESPONSE PARAMETER ESTIMATION IN SYSTEMS OF ORDINARY DIFFERENTIAL EQUATIONS

by

M. Guay

Department of Chemical Engineering
Queen's University
Kingston, Ontario, Canada K7L 3N6

and

D.D. McLean*

Department of Chemical Engineering
University of Ottawa
Ottawa, Ontario, Canada K1N 6N5

* *Author to whom correspondence should be addressed*

October, 1993

ABSTRACT

Methodology for the simultaneous solution of ordinary differential equations (ODEs) and associated parametric sensitivity equations using the Decoupled Direct Method (DDM) is presented with respect to its applicability to multiresponse parameter estimation for systems described by nonlinear ordinary differential equations. The DDM is extended to provide second order sensitivity coefficients and incorporated in multiresponse parameter estimation algorithms utilizing a modified Newton scheme as well as a hybrid Newton/Gauss-Newton optimization algorithm. Significant improvements in performance are observed with use of both the second order sensitivities and hybrid optimization method. In this work, our extension of the DDM to evaluate second order sensitivities and development of new hybrid estimation techniques provide ways to minimize the well-known drawbacks normally associated with second-order optimization methods and expand the possibility of realizing their benefits, particularly for multiresponse parameter estimation in systems of ODEs.

1. INTRODUCTION

Accurate and efficient calculation of parametric sensitivity coefficients (i.e., values of first and second partial derivatives of the state variables with respect to the model parameters) is essential for many aspects of parameter estimation in nonlinear models of processes in chemical engineering, especially determination of gradients for optimization and evaluation of parameter precision. In multiresponse parameter estimation, a mathematical model consisting of a system of simultaneous equations is fitted to a set of experimental observations on more than one response in order to estimate the unknown constants or parameters. Although a number of parameter estimation criteria and optimization techniques have been applied to this problem (e.g., Kalogerakis and Luus (1983); Biegler and Tjoa (1991)), the multiresponse estimation criterion of Box and Draper (1965) and the use of parametric sensitivities are of prime interest in this article. In this article, we present an efficient methods for calculating first and second order sensitivity coefficients in systems described by ordinary differential equations. Our objective was to couple this method with various optimization routines and investigate their performance in the optimization of the Box-Draper criterion for systems described by sets of simultaneous ordinary differential equations.

A number of approaches can be taken to optimize the Box-Draper criterion. For methods based on a quadratic approximation of the objective function in parameter space, such as the Gauss-Newton method, the gradient and Hessian matrix of the objective function with respect to the parameters must be evaluated. Evaluation of the Hessian matrix is a computationally demanding problem which often restricts the applicability of such a technique. Methods which facilitate evaluation of the Hessian matrix exist ; most attempt to approximate the Hessian matrix using only first order sensitivity coefficients (e.g., Bates and Watts (1985)). Analytically defined second order sensitivity coefficients are usually avoided, primarily because of the large increase in computational effort and storage. Alternatively, finite difference approximations may be used. Although easy to implement, they remain computationally intensive. Other methods, such as the quasi-Newton (or update) methods, seek to approximate second order parametric sensitivities by updating part or all of the Hessian matrix. These methods, while better than Gauss-Newton

methods for cases where second order sensitivities are needed, tend to be less efficient in normal situations. Robust hybrid methods (Biegler *et al.* (1991), Dennis *et al.* (1981)) have also been developed which combine the advantages of both approaches.

Where the model consists of a system of ordinary differential equations (ODEs), the evaluation of the gradient and Hessian matrix from sensitivity coefficients becomes computationally demanding. Typically, the solution of the ODEs is used for evaluating the sensitivity coefficients and, ultimately, the gradient and Hessian matrix. The accuracy of the numerical solution thus affects the estimate of these derivatives. A good solution method for the model and sensitivity coefficients is required. The method must also be efficient to be practical for the repetitive use required by parameter estimation. The decoupled direct method (DDM) for the solution of sensitivity coefficients (Dunker (1984)) has many advantages. In this article, the DDM algorithm of Leis and Kramer (1988), called ODESSA, was implemented in a multiresponse parameter estimation routine. Slight improvements to this algorithm were made with respect to error control. The algorithm was also extended to include solution of the second order sensitivity equations. This allowed efficient, more reliable evaluation of the Hessian matrix in the optimization routine.

Although the method for calculation of the model solution and parametric sensitivities is very important when modelling systems described by ODEs, the single most important factor is the choice of optimization method. In this study, four different approaches for estimation of parameters in models described by ODEs by optimization of the Box-Draper and/or least squares criteria were studied. Among the methods were a Gauss-Newton method, a full Newton method which includes second order terms, a hybrid method based on the method of Dennis *et al.* (1981) and a new Newton/Gauss-Newton hybrid method.

In Section 2, background concerning multiresponse parameter estimation in systems of ODEs along with methods used for the optimization of the Box-Draper criterion are given. Section 3 highlights the numerical solution of systems of ODEs and the calculation of first and second order sensitivity coefficients. In Section 4, four parameter estimation problems are solved.

The problems have been selected from both the literature and from our work in batch culture kinetics. The first example, a first order sequential reaction problem familiar to users of the Box-Draper criterion, was used to illustrate the possible advantages of including second order sensitivities. The second example involves a simple culture kinetics model involving only two responses and four parameters, the estimates of two of the parameters being highly correlated yielding a very poorly conditioned estimation problem. This example was used to assess the robustness (i.e., the ability to converge under easy and difficult estimation conditions) of the various approaches. The third problem involves the "DOW Test Problem" (Biegler *et al.* (1986)), a complex highly nonlinear chemical kinetics problem with four responses and eight parameters. The final example arises from our own study of batch culture kinetics and involves estimation of eleven parameters from observations on four responses. It carries difficulties common to the other examples and provides additional evidence upon which our conclusions can be based. The important findings are summarized in Section 5.

2. MULTIRESPONSE PARAMETER ESTIMATION IN SYSTEMS OF ODEs

2.1 General Multiresponse Model

Multiresponse parameter estimation consists of fitting a model to a set of M observed responses, y_{nm} (n=1,...,N), (m=1,...,M), measured at N experimental conditions ξ_n. The observed responses are modelled by

$$y_{nm} = f_m(\boldsymbol{\xi}_n,\boldsymbol{\theta}) + \varepsilon_{nm} \qquad (1)$$

where $f_m(\xi_n,\theta)$ is the expected value of the m^{th} response evaluated at the n^{th} set of experimental conditions, f_m is the m^{th} response function of known form, θ is a P-dimensional vector of unknown parameters to be estimated and ε_{nm} is the error term involved in observation of y_{nm}.

When the experimental conditions are fixed for a set of N observations, the expected responses depend only on θ. Thus the model can be written in matrix form as

$$\boldsymbol{Y} = \boldsymbol{H}(\boldsymbol{\theta}) + \boldsymbol{Z}(\boldsymbol{\theta}) \tag{2}$$

where **Y** is the N by M matrix of observations, **H**(θ) is the matrix of expected responses and **Z**(θ), the matrix of errors.

2.2 Systems Described by Ordinary Differential Equations

For systems described by ODEs, the vector of experimental conditions for the n^{th} run is written as

$$\boldsymbol{\xi}_n^T = (\boldsymbol{x}_n^T, t_n) \tag{3}$$

where $\mathbf{x}_n$ is a vector of values of independent variables (e.g., temperature or initial concentration of reactant) and t_n is the value of an additional independent integration variable, usually time. The M expected values of the responses are given by some known functions $\mathbf{h}_m(\mathbf{s}(t),\mathbf{x},\theta)$, m=1,...,M, which are functions of **s**(t), the K-dimensional vector of state variables, θ and **x**. The general multiresponse model for systems described by systems of ODEs is given by

$$y_{nm} = h_m(s(t_n),\boldsymbol{x}_n,\boldsymbol{\theta}) + \varepsilon_{nm} \tag{4}$$

The state variables are functions of **x**, θ and t, and are implicitly defined by the system of ODEs

$$\frac{d\boldsymbol{s}}{dt} = \boldsymbol{u}(t,\boldsymbol{x},\boldsymbol{s},\boldsymbol{\theta}) \tag{5a}$$

with initial conditions

$$\boldsymbol{s}(0) = \boldsymbol{s}_0(\boldsymbol{x},\boldsymbol{\theta}) \tag{5b}$$

2.3 Optimization of the Box-Draper Estimation Criterion

Assuming $\boldsymbol{\varepsilon}_{nm}$ is normally distributed such that

$$E[\varepsilon_{nm}] = 0 \tag{6a}$$

and

$$\begin{aligned} E[\varepsilon_{nm}\varepsilon_{rs}] &= \{\sigma_{ms}\} \quad , \quad n=r \\ &= 0 \quad , \quad n\neq r \end{aligned} \tag{6b}$$

(which means that the errors/responses are randomly distributed with mean zero and that responses from different run conditions are not correlated but responses for the same run have a fixed covariance matrix, Σ), Box and Draper (1965) showed that parameters chosen to minimize

$$d(\boldsymbol{\theta}) = |\boldsymbol{Z}^T\boldsymbol{Z}| \tag{7}$$

have highest posterior density. Using a Newton-type procedure to minimize $d(\theta)$ requires the evaluation of the gradient, γ, and Hessian matrix, Γ, of $d(\theta)$ with respect to θ. At the $(i+1)^{th}$ iteration, θ_{i+1} is calculated using the formula

$$\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i - \rho[\Gamma + \alpha\boldsymbol{I}]^{-1}\boldsymbol{\gamma} \tag{8}$$

where α is a correction term added to the diagonal elements of the Hessian matrix to ensure its positive definiteness, and ρ is a scalar between 0 and 1 chosen to ensure the step taken is acceptable (Dennis and Schnabel (1983)). α can be chosen, as in Bates and Watts (1985), as twice the absolute value of the most negative eigenvalue of the Hessian matrix. Alternatively, a trust region method (in which case $\rho = 1$), or a modified Cholesky decomposition (Dennis and Schnabel (1983)) can be used.

Bates and Watts (1985) developed a very appealing precision-related convergence criterion for this procedure of the form

where **C** is the Cholesky factor of Γ, s^2 is a scaling factor given by

$$\|C\delta\|/2s^2 < \varepsilon^2 \tag{9}$$

$$s^2 = |Z^T Z|/(N-P) \ , \tag{10}$$

δ is the parameter increment vector given by

$$\boldsymbol{\delta} = -\rho(\boldsymbol{\Gamma} + \alpha \boldsymbol{I})^{-1}\boldsymbol{\gamma} \tag{11}$$

and $\boldsymbol{\varepsilon}$ is the tolerance level, usually set to 0.001 (Bates and Watts (1985).

Expressions for the gradient and Hessian matrix of d(θ) have been obtained by Bates and Watts (1985). However, an alternative approach (Reilly, 1990) was used in this work based on the fact that the gradient of

$$\Phi = \frac{1}{2}\log(d(\boldsymbol{\theta})) \tag{12}$$

can be shown to be given by

$$\frac{\partial \Phi}{\partial \theta_p} = -\sum_{n=1}^{N} \left\{ \frac{\partial f(\boldsymbol{\theta}, t_n)^T}{\partial \theta_p} \boldsymbol{M}^{-1} \boldsymbol{z}_n \right\} \tag{13}$$

where $\mathbf{M} = \mathbf{Z}^T\mathbf{Z}$, $\mathbf{z}_n$ is the n^{th} M-dimensional row vector of $\mathbf{Z}$ and $\partial\mathbf{f}(\theta,t)/\partial\theta_p$ are first order sensitivities of the response function. They are obtained by differentiating eq. 5 with respect to θ_p and applying the chain rule to give

$$\frac{\partial f(\boldsymbol{\theta}, x, t)}{\partial \theta_p} = \sum_{k=1}^{K} \frac{\partial h(s(t), x, \boldsymbol{\theta})}{\partial s_k} \frac{\partial s_k}{\partial \theta_p} + \frac{\partial h(s(t), x, \boldsymbol{\theta})}{\partial \theta_p} \tag{14}$$

The derivatives $\partial s_k/\partial\theta_p$ are obtained by solving the sensitivity equations

$$\frac{d}{dt}\frac{\partial s}{\partial \theta_p} = \frac{\partial u}{\partial s}\frac{\partial s}{\partial \theta_p} + \frac{\partial u}{\partial \theta_p} \tag{15a}$$

with initial conditions

$$\frac{\partial \boldsymbol{s}(0)}{\partial \theta_p} = \frac{\partial \boldsymbol{s}_0}{\partial \theta_p} . \tag{15b}$$

For simplicity let $\mathbf{s_p} = \partial \mathbf{s}/\partial\theta_p$ and $\mathbf{J} = \{j_{mk}\} = \{\partial u_m/\partial s_k\}$ (m=1,...,K; k=1,...,K).

The Hessian matrix of $\Phi(\theta)$ is obtained by differentiating eq. 8 with respect to θ_q (q=1,...,P) to give

$$\frac{\partial^2 \Phi}{\partial\theta_p \partial\theta_q} = -\sum_{n=1}^{N} \left\{ \frac{\partial^2 \boldsymbol{f}^T}{\partial\theta_p \partial\theta_q} \boldsymbol{M}^{-1} \boldsymbol{z}_n - \frac{\partial \boldsymbol{f}^T}{\partial\theta_p} \boldsymbol{M}^{-1} \frac{\partial \boldsymbol{M}}{\partial\theta_q} \boldsymbol{M}^{-1} \boldsymbol{z}_n - \frac{\partial \boldsymbol{f}^T}{\partial\theta_p} \boldsymbol{M}^{-1} \frac{\partial \boldsymbol{f}}{\partial\theta_q} \right\} \tag{16}$$

where

$$\begin{aligned} \frac{\partial^2 \boldsymbol{f}^T}{\partial\theta_p \partial\theta_q} = & \sum_{k=1}^{K} \left(\frac{\partial^2 \boldsymbol{h}}{\partial\theta_p \partial s_k} \frac{\partial s_k}{\partial\theta_q} \right) + \frac{\partial^2 \boldsymbol{h}}{\partial\theta_p \partial\theta_q} \\ & + \sum_{k=1}^{K} \left\{ \frac{\partial s_k}{\partial\theta_p} \sum_{k^*=1}^{K} \left(\frac{\partial^2 \boldsymbol{h}}{\partial s_k \partial s_{k^*}} \frac{\partial s_{k^*}}{\partial\theta_q} \right) + \frac{\partial \boldsymbol{h}}{\partial s_k} \frac{\partial^2 s_k}{\partial\theta_p \partial\theta_q} \right\} \end{aligned} \tag{17}$$

$\partial^2 s_k/\partial\theta_p\partial\theta_q$ is evaluated by solving the second order sensitivity equations obtained by differentiating eq. 15 with respect to θ_q to give

$$\begin{aligned} \frac{d}{dt} \frac{\partial^2 \boldsymbol{s}}{\partial\theta_p \partial\theta_q} = & \left(\sum_{i=1}^{k} \left(\frac{\partial^2 \boldsymbol{u}}{\partial \boldsymbol{s} \partial s_i} \frac{\partial s_i}{\partial\theta_q} \right) + \frac{\partial^2 \boldsymbol{u}}{\partial \boldsymbol{s} \partial\theta_q} \right) \frac{\partial \boldsymbol{s}}{\partial\theta_p} + \frac{\partial \boldsymbol{u}}{\partial \boldsymbol{s}} \frac{\partial^2 \boldsymbol{s}}{\partial\theta_p \partial\theta_q} \\ & + \sum_{i=1}^{K} \frac{\partial^2 \boldsymbol{u}}{\partial\theta_p \partial s_i} \frac{\partial s_i}{\partial\theta_q} + \frac{\partial^2 \boldsymbol{u}}{\partial\theta_p \partial\theta_q} \end{aligned} \tag{18a}$$

with initial conditions,

$$\frac{\partial^2 \boldsymbol{s}(0)}{\partial\theta_p \partial\theta_q} = \frac{\partial^2 \boldsymbol{s}_0}{\partial\theta_p \partial\theta_q}. \tag{18b}$$

Again let $\mathbf{s}_{pq} = \partial^2\mathbf{s}/\partial\theta_p\partial\theta_q$.

The gradient and Hessian matrix of d(θ) are given by

$$\gamma_p(\boldsymbol{\theta}) = \frac{\partial d(\boldsymbol{\theta})}{\partial\theta_p} = -\frac{2}{N}|\boldsymbol{M}|\frac{\partial\Phi}{\partial\theta_p} \tag{19}$$

and

$$\Gamma_{pq}(\boldsymbol{\theta}) = \frac{\partial^2 d(\boldsymbol{\theta})}{\partial\theta_p\partial\theta_q} = -\frac{2}{N}\left(\gamma_q(\boldsymbol{\theta})\frac{\partial\Phi(\boldsymbol{\theta})}{\partial\theta_p} + |\boldsymbol{M}|\frac{\partial^2\Phi(\boldsymbol{\theta})}{\partial\theta_p\partial\theta_q}\right) \tag{20}$$

When second order sensitivity coefficients are evaluated, eq. 16 is used directly to obtain the Hessian matrix of d(θ) in order to estimate θ using the modified Newton scheme given by eq. 8. From our experience, this yields a very robust optimization method.

Assuming that there is no second order effect of the parameters on the model then the Hessian matrix of Φ becomes

$$\left(\frac{\partial^2\Phi}{\partial\theta_p\partial\theta_q}\right)_{GN} = \sum_{n=1}^{N}\left\{\frac{\partial\boldsymbol{f}^T}{\partial\theta_p}\boldsymbol{M}^{-1}\frac{\partial\boldsymbol{M}}{\partial\theta_q}\boldsymbol{M}^{-1}\boldsymbol{z}_n + \frac{\partial\boldsymbol{f}^T}{\partial\theta_p}\boldsymbol{M}^{-1}\frac{\partial\boldsymbol{f}}{\partial\theta_q}\right\} \tag{21}$$

where GN denotes the Gauss-Newton approximation which requires only the first order sensitivity coefficients. This expression can be used in eq. 8 to give a generalized modified Gauss-Newton method similar to the one developed by Bates and Watts (1985). However, they used an algorithm based on a QR decomposition of **Z** to obtain the derivatives of Φ. In this work, analytical expressions, eq. 13, 16 and 21, were used directly. In our experience, these expressions proved to be more efficient then the QR decomposition method. Although, it has been reported for the single response case that the normal equations may not be as numerically stable as the QR decomposition method for calculating an approximate Hessian matrix for optimization of the least squares criterion (Moré, 1978), this behaviour was never encountered for the optimization of the Box-Draper criterion.

In cases where the second order effects are not negligible in eq. 16 and where second order sensitivity coefficients are not evaluated, the generalized Gauss-Newton scheme may not perform well and an update method must be used to approximate the matrix

$$\boldsymbol{A} = \sum_{n=1}^{N} \frac{\partial^2 \boldsymbol{f}_n^T}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \boldsymbol{M}^{-1} \boldsymbol{z}_n.$$

where $\partial^2 \mathbf{f}_n{}^T / \partial\theta\partial\theta^T$ is an MxPxP array with multiplication being performed with respect to the first dimension. Dennis *et al.* (1981) developed a successful update formula for nonlinear least squares. This method can easily be extended to the Box-Draper criterion for multiresponse parameter estimation. The Gauss-Newton approximation of the Hessian matrix is calculated using eq. 21 at each iteration and the second order term, **A**, is updated at each iteration using the formula

$$\begin{aligned} \boldsymbol{A}^{(i+1)} = \tau^{(i)} \boldsymbol{A}^{(i)} + & \frac{\left(\tilde{\boldsymbol{\nu}}^{(i)} - \boldsymbol{A}^{(i)} \boldsymbol{\delta}^{(i)}\right) \boldsymbol{\nu}^{(i)T} + \boldsymbol{\nu}^{(i)} \left(\tilde{\boldsymbol{\nu}}^{(i)} - \boldsymbol{A}^{(i)} \boldsymbol{\delta}^{(i)}\right)^T}{\boldsymbol{\delta}^{(i)T} \boldsymbol{\nu}^{(i)}} \\ & - \frac{\boldsymbol{\delta}^{(i)T} \left(\tilde{\boldsymbol{\nu}}^{(i)} - \boldsymbol{A}^{(i)} \boldsymbol{\delta}^{(i)}\right)}{\boldsymbol{\delta}^{(i)T} \boldsymbol{\nu}^{(i)}} \boldsymbol{\delta}^{(i)} \boldsymbol{\delta}^{(i)T} \end{aligned} \tag{22}$$

where

$$\boldsymbol{\delta}^{(i)} = \boldsymbol{\theta}^{(i+1)} - \boldsymbol{\theta}^{(i)}$$

$$\boldsymbol{\nu}^{(i)} = \boldsymbol{\gamma}^{(i+1)} - \boldsymbol{\gamma}^{(i)}$$

$$\tilde{\boldsymbol{\nu}}^{(i)} = \sum_{n=1}^{N} \left(\frac{\partial \boldsymbol{f}_n \left(\boldsymbol{\theta}^{(i+1)}\right)^T}{\partial \boldsymbol{\theta}} - \frac{\partial \boldsymbol{f}_n \left(\boldsymbol{\theta}^{(i)}\right)^T}{\partial \boldsymbol{\theta}} \right) \boldsymbol{M}^{-1(i+1)} \boldsymbol{z}_n^{(i+1)}$$

$$\tau^{(i)} = \min\left(1, \frac{\left| \boldsymbol{\delta}^{(i)T} \tilde{\boldsymbol{\nu}}^{(i)} \right|}{\left| \boldsymbol{\delta}^{(i)T} \boldsymbol{A}^{(i)} \boldsymbol{\delta}^{(i)} \right|} \right).$$

Therefore, at the $(i+1)^{th}$ iteration, the Hessian matrix of Φ is given by

$$\frac{\partial^2 \Phi^{(i+1)}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} = \left(\frac{\partial^2 \Phi^{(i+1)}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right)_{GN} + \boldsymbol{A}^{(i+1)} \tag{23}$$

Initially, $\mathbf{A}^{(0)}$ is set to zero such that the first step is essentially a Gauss-Newton step. It is then updated using eq. 22. The method proposed in this work is a modification of the hybrid method of Dennis *et al.* (1981) (DGW method) which allows switching from the Gauss-Newton scheme to the update scheme, eqs. 22 and 23, depending on the progress of the optimization.

The switching rules employed in the proposed method were inspired by Dennis *et al.*(1981). At each iteration, the adequacy with which the current scheme models the objective function is checked. To do this, the actual change in the objective function is compared to the change predicted by the current optimization scheme. If the inequality

$$0.5 \leq \left| \frac{d(\boldsymbol{\theta}^{(i)}) - d(\boldsymbol{\theta}^{(i-1)})}{q^{(i)} - d(\boldsymbol{\theta}^{(i-1)})} \right| \leq 1.5 \tag{24}$$

is true, where $q^{(i)}$ is the predicted value of the objective function based on the current scheme, then the current scheme is a good candidate for the following iteration. The alternative scheme is then considered. If eq. 24 is also true for the alternative scheme, then the alternative scheme is used for the next iteration unless the inequality

$$\left| \frac{q^{(i)} - d(\boldsymbol{\theta}^{(i)})}{q_a^{(i)} - d(\boldsymbol{\theta}^{(i)})} \right| \leq 1.0 \tag{25}$$

is true. $q_a^{(i)}$ is the predicted value of the objective function for the alternative scheme. If eq. 24 is not true for both schemes, then eq. 25 is used to decide which scheme is to be used. If eq. 24 is not true for one of the schemes, then the other is used in the following iteration. In addition to these checks, the switching procedure retains the current scheme if good progress is being made, such that

$$\left| \frac{d(\boldsymbol{\theta}^{(i+1)}) - d(\boldsymbol{\theta}^{(i)})}{d(\boldsymbol{\theta}^{(i)})} \right| \geq 0.2. \tag{26}$$

A change of scheme is also considered if

$$\left| \frac{d(\boldsymbol{\theta}^{(i+1)}) - d(\boldsymbol{\theta}^{(i)})}{d(\boldsymbol{\theta}^{(i)})} \right| \leq 0.0001 \tag{27}$$

unless the Bates-Watts convergence criterion is less then 0.001 in which case the current scheme is kept for the next iteration. This is mainly because a small value of this criterion indicates the proximity of an optimum.

The rules, eq. 24 and 25, are essentially the same as those developed by Dennis *et al.* (1981). Eq. 24 is a measure of the closeness of the quadratic approximation to the actual determinant surface. The range of values used in this equality was increased relative to those used by Dennis *et al.* in order to account for the nonlinear behaviour of the Box-Draper criterion. Eq. 25 compares the forecasted change in the Box-Draper criterion for each scheme. The choice of critical value for this inequality is self-explanatory. Eqs. 26 and 27 measures the relative change in the objective function at the $(i+1)^{th}$ iteration. A relative change greater than 20% was chosen as a significant indication that the current scheme performs adequately. Simlarly, a relative change less than 0.1% warrants the use of the current scheme for the $(i+1)^{th}$ iteration.

3. CALCULATION OF PARAMETRIC SENSITIVITIES IN SYSTEMS OF ODES

3.1 Solution of Systems of ODEs

Consider Gear's predictor-corrector method for the solution of eq. 6. It is given by

$$s_{l+1}^{(0)} = \sum_{i=0}^{k-1} \bar{\alpha}_i s_{l-i} + h\beta_0 \dot{s}_l \tag{28a}$$

$$s_{l+1} = \sum_{i=0}^{k-1} \alpha_i s_{l-i} + h\beta_1 \dot{s}_{l+1} \tag{28b}$$

The superscript (0) is used to denote the predicted value of $\mathbf{s}_{l+1}$. $\bar{\alpha}_i$, α_i, β_0 and β_1 result from fitting a k-degree polynomial to **s**(t).

In most efficient codes for the solution of systems of ODEs, predictor-corrector schemes such as Adams' formulae and Gear's backward difference formulae (BDF) are implemented using Nordsieck vectors to efficiently store the solution history in terms of a Taylor series expansion of each variable $\mathbf{s}_l$ at the l^{th} integration step. They take the general form

$$U_l = \left[s_l \, , \; h\dot{s}_l \, , \; \frac{h^2}{2!}\frac{d^2 s_l}{dt^2} \, , \ldots , \; \frac{h^k}{k!}\frac{d^k s_l}{dt^k} \right]^T \tag{29}$$

Nordsieck vectors are advantageous. First, they allow a more efficient predictor step. Second, the simultaneous change of integration step size and order is facilitated. Third, they permit use of more than one algorithm within one implementation of a multistep algorithm (for more details, see Kramer and Leis (1988)).

At each integration step, the solution of eq. 5 can be advanced by solving the equation

$$s_{l+1} = s_{l+1}^{(0)} + h\beta_1 \left(\dot{s}_{l+1} - \dot{s}_{l+1}^{(0)} \right) \tag{30}$$

As described by Kramer and Leis (1988), let $\mathbf{b}_l$ be defined such that

$$\begin{aligned} s_{l+1} &= s_{l+1}^{(0)} + \beta_1 b_l \\ h\dot{s}_{l+1} &= h\dot{s}_{l+1}^{(0)} + b_l \end{aligned} \tag{31}$$

$\mathbf{b}_l$ must then be determined such that, $G_r(\mathbf{b}_l)$ (r= 1,...,K), the resulting residuals of the system, are equal to zero where

$$G(\beta_l) = hu(s_{l+1}^{(0)} + \beta_1 b_l , t_{l+1}) - h\dot{s}_{l+1}^{(0)} - b_l \quad (32)$$

This is usually done through a Newton-Raphson scheme. The value of $\mathbf{b}_l$ obtained is used to calculate $\mathbf{s}_{l+1}$.

Since the value of $\mathbf{b}_l$ is proportional to the local truncation (Gear, (1971)), its magnitude can be compared with some user-defined tolerance in order to control the error involved with the numerical solution as it proceeds. It is used to evaluate the vector of local errors $\mathbf{e}_0$. The error test, which involves comparison of the magnitude of the error with the user-defined tolerance, is generally performed by taking a norm, usually Euclidean, of the error vector. In a number of efficient codes, the integration step size and order are selected in order to satisfy the inequality

$$\|E_0\| \leq 1.0 \quad (33)$$

where $\mathbf{E}_0$ is the K-dimensional vector of weighted local error. The i^{th} element of this vector is obtained by dividing the i^{th} element of $\mathbf{e}_0$ by the i^{th} element of the vector of error weights, $\mathbf{w}_0$. The weights represent relative and/or absolute error tolerances.

A very popular implementation of this procedure is the LSODE program (Hindmarsch (1980)). LSODE uses variable order, variable step-size, predictor-corrector formulae for the integration of initial value stiff and nonstiff systems of ODEs. It uses Nordsieck vectors and implements Gear's BDF and Adams-Moulton formulae.

3.2 Decoupled Direct Method(DDM) for Calculating Sensitivities

In the decoupled direct method (Dunker (1984)), a predictor-corrector method using Gear's backward difference formulae or Adams formulae is used to solve eq. 5. Information generated by the solution is then used in the solution of eq. 15. In DDM, the solution of eq. 15 is decoupled from that of eq. 5 but is conducted simultaneously. This means that each step taken in solving eq. 5 is followed by the solution of eq. 15 over the same step (Kramer and Leis (1988)).

The approach is simple. For example, time derivatives for the sensitivity coefficients can be approximated by a one-sided difference formula between $\mathbf{s}_{p,l+1}$ and $\mathbf{s}_{p,l}$ over the time interval h. The approximation can be substituted into eq. 15 to yield an explicit formula for $\mathbf{s}_{p,l+1}$ of the form

$$\boldsymbol{s}_{p,l+1} = \left[\boldsymbol{I} - h\boldsymbol{J}_{l+1}\right]\boldsymbol{s}_{p,l} + h\frac{\partial \boldsymbol{f}_{l+1}}{\partial \theta_p} \tag{34}$$

In the simultaneous solution of the model and the sensitivity equations, the model solution, $\mathbf{s}_{l+1}$, is used to calculate $\mathbf{J}_{l+1}$, $\partial\mathbf{u}/\partial\theta_p$ and $\mathbf{s}_{p,l+1}$ at each step. To improve the approximation, a k-step linear difference equation can be used. This method was first applied to the problem of stiff ODEs by Dunker (1984).

Recently, a very efficient algorithm based on the ODE solving routine LSODE, called ODESSA, was developed by Kramer and Leis (1988). ODESSA incorporates Gear's BDF and Adams' formulae for initial value problems using the Nordsieck arrays. At each step, the solution of the state equations is handled first. As given above, $\mathbf{b}_l$ is determined using a modified Newton scheme on the first and second rows of the Nordsieck vectors. $\mathbf{b}_l$ is used to evaluate the local truncation error for the model solution, $\mathbf{e}_0$. The error must be such that the inequality in eq. 33 is observed. If it is not then the step is repeated with, for example, a smaller step size. A successful error test initializes the solution of the sensitivity equations.

For the solution of eq. 15, use of eqs. 30, 31 and 32 in terms of sensitivities gives an explicit formula for the sensitivities

$$\mathbf{s}_{p,l}=\left[\mathbf{I}-h\beta_1\mathbf{J}_l\right]^{-1}\left(\mathbf{s}_{p,l}^{(0)}-h\beta_1\mathbf{s}_{p,l}^{(0)}+h\beta_1\frac{\partial \mathbf{u}_l}{\partial \theta_p}\right) \tag{35}$$

where $\mathbf{J}_l$ is evaluated at the current solution conditions, t_l, $\mathbf{s}_l$. As discussed by Kramer and Leis (1988), $\mathbf{J}_l$ should be evaluated at the predicted value, $\mathbf{s}_l^{(0)}$, at t_l. Considerable computational gains are thus realized by reducing the number of evaluations of the Jacobian matrix.

The error control strategy for the solution of eq. 15 is similar to that for the solution of eq. 5. The local truncation error for the sensitivities, $\mathbf{e}_p$ (p=1,...,P), is assumed to be proportional to the difference between the predicted, $\mathbf{s}_p^{(0)}$,and the corrected value, $\mathbf{s}_p$, (p=1,...,P) of the sensitivities. The errors are subject to the following P inequalities

$$\frac{1}{K}\sum_{k=1}^{K}\left(\frac{e_{p,k}}{w_{p,k}}\right)^2 \le 1.0 \qquad (p=1,...,P) \tag{36}$$

In both cases, eq. 33 and eq. 36, failure of the error test causes the step to be repeated starting with a new step size and, possibly, a new integration order.

We have found that an alternative error control strategy significantly improves the performance of the DDM. This strategy, developed by Shampine (1988) for one-step ODE codes, has been shown to improve the dependency of the local error of the solution on the requested tolerance. It involves definition of a new internal tolerance as a function of the requested (or external) tolerance and of the integration order which takes the form

$$\mu = \tau^{(k+1)/k} \tag{37}$$

where μ is the internal tolerance, τ is the requested tolerance and k is the integration order. Although application of this method to multistep codes has not been rigorously investigated in

the literature, our results showed an improvement similar to that achieved for one-step codes. For DDM, the implementation of this strategy resulted in a net improvement in both the local error dependency on external tolerance and the computational efficiency of the method.

Failure of the error test initiates a change in step size and order as described for the model solution. The step is then repeated starting with prediction of the model solution. If the step is repeated more then three times, $\mathbf{J}_l$ and $\partial \mathbf{u}_l / \partial \theta_p$ are re-evaluated directly from the analytical formulae. If the error test is successful, the Nordsieck arrays are updated and the step is repeated. This can be preceded by selection of a new integration step size and order in a manner similar to LSODE using the following equations

$$\begin{aligned}
\alpha &= \frac{1}{1.2}\left(\frac{1}{\max_j \|\hat{E}_j\|}\right)^{1/(k+1)} \quad , \quad j=0,\ldots,P \quad (\textit{order } k) \\
\alpha &= \frac{1}{1.3}\left(\frac{1}{\max_j \|\hat{E}_j\|}\right)^{1/k} \quad , \quad j=0,\ldots,P \quad (\textit{order } k-1) \\
\alpha &= \frac{1}{1.4}\left(\frac{1}{\max_j \|\hat{E}_j\|}\right)^{1/(k+2)} \quad , \quad j=0,\ldots,P \quad (\textit{order } k+1)
\end{aligned} \tag{38}$$

where α is a factor that multiplies the current step size to generate the new step size for the respective integration orders, and $\hat{\mathbf{E}}_j$ are the K-dimensional vector of weighted error vectors for the sensitivities just as defined in eq.33. The local error vectors for the sensitivities, $\mathbf{e_p}$ are weighted by the corresponding values of the internal tolerances given by $\boldsymbol{\mu}_p=\{\mu_{i,p}\} = \{w_{i,p}^{(k+1)/k}\}$. The maximum value of α is used as the criterion for integration order and step size.

3.3 Calculation of Second Order Sensitivity Coefficients using DDM

An extension of this method can be used to calculate second order sensitivities very accurately. The formula is obtained by differentiating the explicit formula for $\partial \mathbf{s}/\partial \theta_p$ (i.e. eq. 35) with respect to θ_q yielding

This extension allows the simultaneous solution of the model and of the first and second order

$$\left[I - h\beta_1 J_l\right]\left(\frac{\partial^2 s_l}{\partial\theta_p\partial\theta_q}\right) \approx \frac{\partial^2 s_l^{(0)}}{\partial\theta_p\partial\theta_q} - h\beta_1\frac{d}{dt}\left(\frac{\partial^2 s_l^{(0)}}{\partial\theta_p\partial\theta_q}\right)$$
$$+ h\beta_1\left[\sum_{k=1}^{K}\frac{\partial^2 u_l}{\partial s\partial s_k}\frac{\partial s_{kl}}{\partial\theta_q} + \frac{\partial^2 u_l}{\partial\theta_p\partial\theta_q}\right]\frac{\partial s_l}{\partial\theta_p} \qquad (39)$$
$$+ h\beta_1\sum_{k=1}^{K}\left(\frac{\partial^2 u_l}{\partial\theta_p\partial s_k}\frac{\partial s_{kl}}{\partial\theta_q}\right) + h\beta_1\frac{\partial^2 u_l}{\partial\theta_p\partial\theta_q}$$

sensitivity equations.

Solution of the second order sensitivity equations is performed similarly to the solution of eq. 15 in section 3.2. The calculation of the second order sensitivities is initiated after a successful error test of the first order sensitivities. Prediction, correction and error tests are performed sequentially. By analogy to the K inequalities of the error test for the solution of eq. 15, the error test for the solution of the second order sensitivity equations consists of K inequalities where the 2-norm of each P by P matrix (i.e., the spectral radius, ρ, of each matrix) of local errors is compared to a weighted measure of tolerance. The test takes the form

$$\rho(S_k) \le 1.0 \qquad (k=1,\ldots,K) \qquad (40)$$

where

$$S_k = \left\{\frac{e_{pq,k}}{\mu_{pq,k}}\right\} \qquad (k=1,\ldots,K).$$

The change of integration order is performed in a manner similar to eq. 37. The test is simply augmented by the K measures of weighted errors of eq. 40. In general, we have found that change of integration order or step size is primarily due to errors in the calculation of first order sensitivity coefficients. The calculation of second order derivatives involves a significant increase in computing effort and storage. However, it should be noted that the resulting K P by P matrices are symmetric and that the solution can thus be restricted to K upper triangular matrices.

This reduces the computational effort and the storage requirements.

4. RESULTS AND DISCUSSION

The performance of four methods for multiresponse parameter estimation in systems described by sets of simultaneous ODEs based on methodologies described in the previous two sections were investigated.

The first method was a modification of the generalized Gauss-Newton method of Bates and Watts (1985), called the GN method. The second was a hybrid method based on the DGW update formula discussed in section 2.4, called the HYB method. Both methods are based on first order approximations of the Hessian matrix. The third method was a full modified Newton method which utilizes second order sensitivity coefficients in the evaluation of the Hessian matrix, called the N method. The fourth method was a new hybrid method which allows switching between the Newton method and the Gauss-Newton method called the N/GN method. The switching rule for this method depends solely on the magnitude of the Bates-Watts convergence criterion. Since this criterion tends to decrease in the proximity of a local minimum, it can be used to indicate that the optimization routine is near convergence. Under such conditions, the Gauss-Newton method generally performs very well and second order terms are not needed in the evaluation of the Hessian matrix. In this study, the hybrid method switched from the modified Newton method to the generalized GN method when the Bates-Watts convergence criterion became less than or equal to 0.05.

Four estimation problems were used to test the behaviour of these four methods. For each of these problems, the CPU time required to perform the estimation on a SUN SPARCstation IPC, the number of iterations (NI) and the number of objective function evaluations (NOF) were recorded to evaluate and compare the performance of the four methods. For two of the examples, different sets of initial parameter estimates were used. For all cases, the numerical solution of the ODEs was obtained at a requested relative tolerance of 10^{-6}. The Bates-Watts convergence criterion was required to be less than or equal to 10^{-6}.

4.1 Example 1

To demonstrate the possible gains that can be made by including second order sensitivities for evaluating the gradient and Hessian matrix of d(θ) in multiresponse parameter estimation a frequently used example involving estimation of two rate constants for the first order sequential reaction

$$A \xrightarrow{\theta_1} B \xrightarrow{\theta_2} C$$

was employed. The data and problem statement have been given by Box and Draper (1965). The model is given by

$$\begin{aligned} \frac{dc_A}{dt} &= -\theta_1 c_A \\ \frac{dc_B}{dt} &= \theta_1 c_A - \theta_2 c_B \\ \frac{dc_C}{dt} &= \theta_2 c_B \end{aligned} \tag{42}$$

where c_A, c_B and c_C are the molar concentrations of A, B and C respectively.

As in most iterative schemes, the outcome and the rate of convergence of the estimation algorithm are highly dependent on the initial guesses for the values of the parameters. In the present case the performance of the four multiresponse estimation methods, expressed in terms of the number of iterations, number of objective-function evaluations and CPU time, was tested using different sets of initial estimates. The choice of each set of initial estimates was arbitrary. Tables 1, 2, 3 and 4 give the results obtained for the GN, HYB, N and N/GN methods respectively.

Only the N and N/GN methods were able to perform the estimation for the four sets of initial estimates. Results in Table 1 show that the GN method converged satisfactorily only for initial estimates $(0.4,0.6)^T$ and $(1.0,2.0)^T$. The HYB method performed well but failed to converge in one case. This suggests a relatively poor robustness of the Gauss-Newton method. The introduction of second order information in the evaluation of the Hessian matrix by either the DGW update formula or the evaluation of sensitivity coefficients was clearly advantageous in cases where initial estimates were poor (i.e., the first, third and fourth cases). The N and N/GN methods were clearly more robust and performed very efficiently. The HYB method did not perform as well, indicating the beneficial effects of using second order sensitivities. The approximation to the matrix **A** in eq. 16 by the update formula thus appears to differ from the actual **A** matrix evaluated in the N and N/GN methods. This is generally expected of update formulae (Dennis and Schnabel (1983)).

4.2 Example 2

Herbert's model is a common and simple description of batch culture kinetics of microorganisms in the presence of a growth limiting substrate and is given by

$$\begin{aligned} \frac{dX}{dt} &= \frac{\mu_{max} XS}{K_S + S} - k_d X \\ \frac{dS}{dt} &= -\frac{1}{Y_S}\frac{\mu_{max} XS}{K_S + S} \end{aligned} \tag{43}$$

where X is the concentration of viable cells, S is the substrate concentration, μ_{max} is the specific growth rate constant, K_s is the substrate saturation constant, k_d is the death rate constant and Y_s is the yield of biomass per unit mass of substrate. A common problem encountered in the estimation of parameters in this type of model is the high correlation between the estimates of μ_{max} and K_s. This often causes common estimation procedures to fail or yield meaningless results. It thus provides a good test of robustness for such algorithms with respect to their ability to converge efficiently for poorly conditioned systems.

Data generated from eq. 43 were employed to estimate the four parameters using the four estimation methods. The initial conditions for X and S were $0.5x10^{3}$ cells/mL and $4.0x10^{-1}$ µg/mL, respectively. μ_{max}, K_s, k_d and Y_s were assigned the values 0.5, 1.5, 0.1 and 5.0, respectively. X and S were assigned variances of 1.5 and 0.75 respectively with a correlation of -0.75. The data were then used to estimate the four parameters of eq. 43 with the four estimation methods. The estimation was performed for two sets of initial parameter estimates; the first set had initial estimates of 1.0, 4.0, 1.0 and 1.0 for μ_{max}, K_s, k_d and Y_s respectively and the second set, values of 5.0, 45.0, 0.2 and 6.0. The results are given in Tables 5 and 6.

For the first case, only the N, N/GN and HYB methods converged. The converged estimates were 6.03, 52.21, 0.14 and 6.50 for μ_{max}, K_s, k_d and Y_s, respectively. In Figure 1, the predicted responses are plotted against time along with the observed responses. The scatter in the data is considerable. For the second case, all four methods converged to the same estimates obtained with the first set of initial guesses. For the GN and HYB methods, convergence was not reached with respect to the Bates and Watts convergence criterion but with respect to the relative change in the objective function, which became less than 10^{-9}.

In this example, the correlation between the estimates of μ_{max} and K_s was nearly 1.0. It was therefore impossible to estimate them independently. Moreover, as noted in the above results, it was not possible to recover the parameter values used for the simulation. However, consistent results were obtained for the N and N/GN methods. Interestingly, there was little change in computing required in the two cases for these two methods. The GN method failed to converge in the first case. The HYB method yielded parameter estimates identical to those obtained with the N and N/GN methods. However, as shown in Table 5, it could not efficiently reach the optimum in the first case. This indicated that use of second order sensitivity coefficients in evaluation of the Hessian matrix did convey an increased robustness to the N and N/GN methods. The main effect of including second order information was improvement in conditioning of the Hessian matrix. For example, it was observed that the condition number of the Hessian matrix calculated at the initial parameter estimates for the first case was approximately 10^{8} for the GN and HYB methods and 10^{5} for the N and N/GN methods. This

was the main source of difficulty for the GN and HYB methods in the first case. As shown in Table 6, there was little difference in computing time among the four methods. Apparently, the increase in computing time required to evaluate the second order sensitivity coefficients was almost entirely compensated by the requirement for fewer iterations, resulting from the improved conditioning of the Hessian matrix. The increased robustness was due to the more accurate evaluation of the Hessian matrix in the N and N/GN methods. By using these methods, problems of convergence related to the high correlation between two parameters and large scatter in the data were overcome. It must be emphasized that the improved robustness of the estimation routine in no way improved the very poor quality of the parameter estimates. This task remains a challenge for the model builder through collection of more informative data or, potentially, use of an alternative model form.

4.3 Example 3

This example was formulated by the Dow Chemical Company to illustrate parameter estimation encountered in industry (see Biegler *et al.*(1986)). The model describes the kinetics of an isothermal batch reactor system. As in Biegler (1991), the following formulation due to Villadsen *et al.* (see Biegler *et al.*(1981)) was studied:

$$
\begin{aligned}
\frac{dy_1}{dt} &= -k_2 y_1 y_2 \frac{-2x_3 + x_4 + 2y_1 - y_2 + y_3}{y_1 + \beta_1(-x_3 + x_4 + y_1 - y_2) + \beta_2 y_3} \\
\frac{dy_2}{dt} &= -k_1 y_2 (x_2 + 2x_3 - x_4 - 2y_1 + y_2 - y_3) \\
&\quad - k_2 y_1 y_2 \frac{-2x_3 + x_4 + 2y_1 - y_2 + y_3}{y_1 + \beta_1(-x_3 + x_4 + y_1 - y_2) + \beta_2 y_3} \\
\frac{dy_3}{dt} &= k_1 (x_3 + y_1 - y_3)(x_2 + 2x_3 - x_4 - 2y_1 + y_2 - y_3) \\
&\quad + k_{-1}\beta_1(-x_3 + x_4 + y_1 - y_2) + k_2 y_1 y_2 \frac{-2x_3 + x_4 + 2y_1 - y_2 + y_3}{y_1 + \beta_1(-x_3 + x_4 + y_1 - y_2) + \beta_2 y_3} \\
&\quad - \frac{1}{2} k_{-1} \beta_2 y_3 \frac{-2x_3 + x_4 + 2y_1 - y_2 + y_3}{y_1 + \beta_1(-x_3 + x_4 + y_1 - y_2) + \beta_2 y_3}
\end{aligned}
\tag{44}
$$

where

$$\beta_1 = \frac{K_1}{K_2}, \quad \beta_2 = \frac{K_3}{K_2}, \quad k_1 = k_{10}\exp\left(\frac{E_1}{R}\left(\frac{1}{x_1} - \frac{1}{T_0}\right)\right)$$

$$k_2 = k_{20}\exp\left(\frac{E_2}{R}\left(\frac{1}{x_1} - \frac{1}{T_0}\right)\right), \quad k_{-1} = k_{-10}\exp\left(\frac{E_{-1}}{R}\left(\frac{1}{x_1} - \frac{1}{T_0}\right)\right)$$

x_1, the temperature of the system, x_2, the catalyst concentration, x_3, the initial concentration of y_1, and x_4, the initial concentration of y_2, are the operating conditions of the system. T_0 is a reference temperature (= 340.15 Kelvin). y_1, y_2 and y_3 are the state variables. β_1, β_2, k_{10}, k_{20}, k_{-10}, E_1, E_2 and E_{-1} are the model parameters.

The data for this parameter estimation problem are given in Biegler *et al*. (1986). It consists of three experimental runs taken at varying temperatures. Measurements of the concentration of four species are given. The first three variables correspond to the state variables y_1, y_2 and y_3 respectively. The fourth response variable is given by

$$y_4 = y_{1o} - y_1 - y_3 \tag{45}$$

where y_{1o} is the initial concentration of y_1.

Parameter estimates which minimize the Box-Draper estimation criterion and a Least-Squares criterion (trace of $\mathbf{Z}^T\mathbf{Z}$, which equals the sum of squared residuals for both responses) were obtained using the four optimization methods. The Least Squares criterion was used to assess the robustness of the methods to the estimation criterion. A logarithmic transformation of the parameters was used. This is a common strategy employed in parameter estimation which often results in an improvement in convergence properties of the optimization. Its beneficial effects are often due to the fact that it attenuates effects due to the difference in magnitude between parameter values and reduces the nonlinearity of the estimation problem. Measurements of responses y_1 and y_4 were used to obtain the parameter estimates given in Table 7 for the Box-

Draper and Least-Squares criteria. The logarithm of the resulting parameter estimates are listed.

Performance indicators for the four optimization methods are given in Tables 8 and 9. The N and N/GN methods converged in fewer iterations than the GN and HYB methods. It was very interesting to note that the GN and HYB methods were greatly affected by the change in objective function whereas the N and N/GN methods were not. A considerable increase in the number of iterations and computing time was observed in the optimization of the Box-Draper criterion with the GN and HYB methods. The poor conditioning of the Hessian matrix obtained with the GN and HYB methods was again improved by using the second order information in the N and N/GN methods.

The results listed in Tables 8 and 9 were indicative of important features of the N and N/GN methods. It was observed that the evaluation of second order sensitivity coefficients produced a significant increase in computational requirements. This should be expected if we consider that the N and N/GN methods require solution of an additional M*P*P (192, in this case) second order sensitivity equations. Clearly, if a second order method were selected in the present case, improvements in performance of the optimization routines would only be obtained with an increase in CPU time. This fact was well demonstrated in Table 9 and in the past examples where the GN and HYB methods required less computational effort in easier problems. However, this increase was actually smaller than expected considering the computational task. It is doubtful that such good performance could be realistically achieved with a conventional ODE solver.

4.4 Example 4

This is an example which illustrates the type of estimation problem encountered in the study of batch culture kinetics. It is a typical model in which biomass concentration is related to the substrate uptake by a Monod relationship. The model was inspired by an investigation of batch experiments of *Eschirichia coli* in the presence of glucose. Two products, acetate and ethanol, were considered important factors in the growth of the microorganism. The model is given by

$$
\begin{aligned}
\frac{dX}{dt} &= \mu_{maxS}\left(\frac{XS}{K_S+S}\right) + \mu_{maxA}\left(\frac{XA}{K_A+A}\right)\left(\frac{K_{IS}}{K_{IS}+S}\right) - k_dX \\
\frac{dS}{dt} &= -\frac{\mu_{maxS}}{Y_S}\left(\frac{XS}{K_S+S}\right) \\
\frac{dA}{dt} &= \frac{\mu_{maxS}}{Y_AY_S}\left(\frac{XS}{K_S+S}\right) - \frac{\mu_{maxA}}{Y_{AA}}\left(\frac{XA}{K_A+A}\right)\left(\frac{K_{IS}}{K_{IS}+S}\right) \\
\frac{dE}{dt} &= \frac{\mu_{maxS}}{Y_E}\left(\frac{XS}{K_S+S}\right) - k_{evap}E
\end{aligned}
\tag{46}
$$

where X is the concentration of biomass expressed in cells per unit volume, S is the concentration of glucose, A is the concentration of acetate and E is the concentration of ethanol. μ_{maxS}, μ_{maxA}, K_S, K_A, K_{IS}, Y_S, Y_A, Y_{AA}, Y_E, k_d, k_{evap} are the model parameters. This model was used to simulate six batch experiments. A fixed covariance matrix, which reflected the scatter and the correlation between the response variables observed in actual experimental data, was used to generate the simulated data. Typical variances for X, S, A and E were 800 (counts/mL)^2, 0.003 (μg/mL)^2, 0.03 (μg/mL)^2 and 0.01 (μg/mL)^2 respectively. The correlation matrix used in the simulation was

$$
\boldsymbol{\rho} = \begin{bmatrix} 1.00 & -0.75 & 0.25 & 0.20 \\ -0.75 & 1.00 & -0.35 & -0.40 \\ 0.25 & -0.35 & 1.00 & 0.05 \\ 0.20 & -0.40 & 0.05 & 1.00 \end{bmatrix}.
$$

Table 10 summarizes the initial conditions, duration and number of observations for each run.

The four optimization algorithms were used to determine the values of the model parameters which optimize the Box-Draper criterion. The logarithms of the actual values of the parameters used for simulating the experiments, initial estimates, and the estimates obtained with the N method are shown in Table 11. Figure 2 shows the resulting fit for run 1. Performance

indicators for the four optimization methods are given in Table 12.

All four optimization methods converged. The N/GN, GN and HYB methods yielded parameter estimates similar to those listed in Table 11. The N and N/GN methods achieved faster convergence in terms of the number of iterations but required greater computing time. Again, the increase was expected considering the required solution of an additional 220 (=(4)(11)(10)/2) simultaneous ODEs. Although this increase in CPU time was significant, it was not prohibitive. On the other hand, although the GN and HYB methods required less CPU time, their overall performance was not entirely desirable, each method requiring an exceedingly large number of iterations to reach convergence. Both methods displayed superlinear convergence.

The difference in performance of the first order methods and the second order methods was primarily due to the nonlinearity experienced by this model under changes in parameter values. Because the GN and HYB depend solely on the validity of a first order approximation of the model with respect to the parameters, they are very susceptible to the nonlinear effects expected in this case. The poor convergence properties displayed by these methods was due to their inability to overcome these effects. Nevertheless, the HYB method performed better than the GN method indicating that the DGW approximation of second order terms in the Hessian matrix was clearly beneficial.

5. SUMMARY

The Decoupled Direct Method (DDM) for the simultaneous solution of ODEs and the associated first order parametric sensitivity equations was extended to include solution of the second order parametric sensitivity equations. The extended DDM was then used to develop a full Newton scheme for multiresponse parameter estimation in systems of ODEs. A hybrid method was also developed which allowed switching between this Newton scheme and a generalized Gauss-Newton scheme. The switching procedure was based on the magnitude of the Bates-Watts convergence criterion (Bates and Watts (1985)).

Both methods were compared to a generalized Gauss-Newton method and also to a hybrid method which incorporated the DGW update method and the generalized Gauss-Newton method. This latter method was based on a modification of the single response nonlinear least squares algorithm of Dennis *et al.* (1981).

In comparing the four estimation methods as applied to four test examples it was observed that methods which incorporated second order sensitivity coefficients in the evaluation of the Hessian matrix were more robust and reliable. The modified full Newton method and the hybrid (Newton/Gauss-Newton) method successfully converged despite problems with high correlation between parameters, large residuals and poor initial estimates. The use of a hybrid method incorporating the DGW update formula to estimate the Hessian matrix was also observed to be robust but not to the extent observed for the full second order methods. Of the two methods employing second order sensitivities, the new hybrid method was significantly more efficient.

There are well-known drawbacks to requiring second-order derivatives in parameter estimation algorithms. Additional coding for calculating the derivatives, increased computational time and storage are prime factors which increase dramatically with the size of the problem. Some of these problems can be circumvented. For example, the use of symbolic computation packages greatly reduced the amount of coding required. In fact, with packages such as MAPLE, only coding of the model equations was required. Our extension of the DDM to evaluate second-order sensitivities and development of new hybrid estimation techniques provide ways to minimize these drawbacks and expand the possibility of realizing the benefits of second order methods, particularly for multiresponse parameter estimation in systems of ODEs. It is clear that the efficient and accurate evaluation of second order sensitivity coefficients enhances convergence. In addition, this allows evaluation of the extent of nonlinearity that the model experiences upon variation of the model parameters. In particular, we are currently exploring for the evaluation of curvature measures similar to those developed by Bates and Watts (1988) for complex nonlinear multiresponse regression models. Such curvature measures have been shown to be useful diagnostic tools in nonlinear regression analysis for single responses.

As a final, supplemental comment, we note that many of the problems in multiresponse parameter estimation arise as a consequence of limitations in the model building strategy. In particular, starting with too large or complex a model form with limited information in the data inevitably leads to problems in estimation. We strongly recommend beginning with a simplified model with informative data, collected from carefully designed experiments, followed by subsequent modification of the model and further data collection until an adequate fit is achieved. Suitable transformation of the parameters can also improve the estimation. With such a strategy, a simple first-order estimation method will usually suffice. Nevertheless, situations do arise where second-order methods are required making hybrid methods like those developed here valuable.

Acknowledgments - The financial support of the National Science and Engineering Research Council of Canada and the Ontario Graduate Scholarship Program are gratefully acknowledged.

NOMENCLATURE

Scalars

$d(\theta)$ Box-Draper estimation criterion

h integration step-size

k integration order

K number of state variables

M number of response variables

N number of observations

q predicted value of the objective function based on the quadratic approximation for the optimization scheme used

t independent variable (usually time)

α term used to make the Hessian matrix sufficiently positive definite

$\alpha_i, \alpha_i, \beta_i$ constants resulting from the fit of k-degree polynomial to the solution of system of ODEs

ρ step-size constant in eq. 8 or spectral radius in eq. 40

ε critical value of the Bates-Watts convergence criterion

Arrays

A PxP matrix containing second order derivatives in the evaluation of the Hessian matrix

b K-dimensional accumulated error vector used to advance the solution of a system of ODEs

C upper triangular Cholesky factor of Γ

e K-dimensional vector of local error estimates

E KxPxP array of local error estimates

F = $\{f_{nm}\}$ = NxM matrix of predicted variables

H = $\{h_{nm}\}$ = NxM matrix of predicted variables

I identity matrix

J Jacobian of a K-dimensional system of ODEs

s K-dimensional vector of state variables

$\mathbf{s}_0$ K-dimensional vector of initial conditions for the state variables

u K-dimensional vector of time derivatives of the state variables

U Kxk matrix of Nordsieck vectors

w K-dimensional vector of error weights

x vector of experimental conditions (independent variables), excluding t_n

$\mathbf{x}_n$ vector of experimental conditions, excluding t_n, for the n^{th} run

Y $= \{y_{nm}\}$ = NxM matrix of observed responses

Z $= \{\boldsymbol{\varepsilon}_{nm}\}$ = NxM matrix of residuals

γ gradient vector of the parameter estimation criterion with respect to the parameters

δ P-dimensional parameter increment vector

θ P-dimensional vector of parameters

$\boldsymbol{\mu}$ K-dimensional vector of internal error weights

ν,ν P-dimensional perturbation vectors used in the update of **A**

ξ_n n^{th} vector of experimental conditions

Σ $= \{\sigma_{rs}\}$ = MxM covariance matrix for the responses

Γ Hessian matrix of the parameter estimation criterion with respect to the parameters

Superscripts

$^{(i)}$ i^{th} iteration of an iterative scheme

^{T} transpose of a matrix or vector

Subscripts

$_{n,r}$ denotes the n^{th} integration step and the r^{th} state variable of a vector

$_{nm}$ denotes the element in the n^{th} row and m^{th} column of a matrix

Abbreviations

NI number of iterations

NOF number of objective function evaluations

REFERENCES

Bates, D M, D G Watts, Multiresponse Estimation With Special Application to Linear Systems of Differential Equations, *Technometrics* **27**, 329-339 (1985).

Biegler, L T, I B Tjoa, Simultaneous Solution and Optimization Strategies for Parameter Estimation of Differential-Algebraic Equation Systems, *Ind. Eng. Chem. Res.* **30**, 376-385 (1991).

Biegler, L T, J J Damiano and G E Blau, Nonlinear Parameter Estimation: A Case Study Comparison, *A.I.Ch.E. Journal* **32**, 29-45 (1986).

Box, G E P , N Draper, The Bayesian Estimation of Common Parameters, *Biometrika* **52**, 355-365 (1962).

Dennis, J E, D M Gay and R E Welsch, An Adaptive Nonlinear Least-Squares Algorithm, *ACM Trans. Math. Soft.* **7**, 348-368 (1981).

Dennis, J E, R B Schnabel, *Numerical Methods of Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall, Englewood Cliff, NJ (1983).

Dunker, A M, The Decoupled Direct Method of Calculating Sensitivity Coefficients in Chemical Kinetics, *J. Chem. Phys.* **81**, 2385-2393 (1984).

Gear, C W, *Numerical Initial Value Problems in Ordinary Differential Equations*. Prentice-Hall, Englewood Cliff, NJ (1971).

Hindmarsh, A C, LSODE and LSODI, Two Initial Value Ordinary Differential Equation Solvers, *ACM-SIGNUM NEWSLETTER* **15**, 10-11 (1980).

Kalogerakis, N, R Luus, Improvement of Gauss-Newton Method for Parameter estimation through the Use of Information Index, *Ind. Eng. Chem. Fund.* **22**, 436-445 (1983).

Kramer M A, J R Leis, Simultaneous Solution and Sensitivity Analysis of ODE Systems, *ACM Trans. Math. Soft.* **14**, 45-60 (1988).

Kramer M A, J R Leis, Sensitivity Analysis of Systems of Differential and Algebraic Equations, *Comp. Chem. Eng.* **9**, 93-96 (1985).

Moré, J J, The Levenberg-Marquardt Algorithm: Implementation and Theory, In *Lecture Notes in Mathematics no.630 Numerical Analysis*. G Watson, Ed., Springer-Verlag, New York, 105-116 (1978).

Reilly, P M, Personal Communication, Dept. of Chemical Engineering, University of Waterloo, Waterloo, Canada (1990).

Shampine, L F, Tolerance Proportionality In ODE Codes, *Mathematical Computing* **39**, 239-251 (1988).

TABLE 1

Estimation Efficiency Using the GN Method for Example 1

INITIAL ESTIMATES		NI(NOF)	CPU secs	CONVERGED ESTIMATES	
0.04	0.06	83(86)	117.1	0.208	0.495
0.4	0.6	7(7)	1.3	0.208	0.495
2.0	3.0	Convergence Failed			
1.0	2.0	9(9)	5.0	0.208	0.495

TABLE 2

Estimation Efficiency Using the HYB Method for Example 1

INITIAL ESTIMATES		NI(NOF)	CPU secs	CONVERGED ESTIMATES	
0.04	0.06	17(17)	3.5	0.208	0.495
0.4	0.6	7(7)	1.4	0.208	0.495
2.0	3.0	Convergence Failed			
1.0	2.0	9(9)	5.2	0.208	0.495

TABLE 3

Estimation Efficiency Using the N Method for Example 1

INITIAL ESTIMATES		NI(NOF)	CPU secs	CONVERGED ESTIMATES	
0.04	0.06	13(13)	3.4	0.208	0.495
0.4	0.6	7(7)	2.0	0.208	0.495
2.0	3.0	11(11)	9.4	0.208	0.495
1.0	2.0	9(9)	7.3	0.208	0.495

TABLE 4

Estimation Efficiency Using the N/GN Method for Example 1

INITIAL ESTIMATES		NI(NOF)	CPU SECS	CONVERGED ESTIMATES	
0.04	0.06	13(13)	3.2	0.208	0.495
0.4	0.6	7(7)	1.8	0.208	0.495
2.0	3.0	11(11)	9.4	0.208	0.495
1.0	2.0	9(9)	6.8	0.208	0.495

TABLE 5

Estimation Efficiency for Example 2
Initial estimates: μ_{max} = 1.0, K_s = 4.0, Y_s = 1.0, k_d = 1.0

METHOD	CPU secs	NI(NOF)	BD
N	11.3	16(33)	325.1
N/GN	6.5	14(28)	325.1
GN	Convergence Failed		
HYB	62.0	135(137)	325.1

TABLE 6

Estimation Efficiency for Example 2
Initial estimates: μ_{max} = 5.0, K_s = 45.0, Y_s = 6.0, k_d = 0.2

METHOD	CPU secs	NI(NOF)	BD
N	8.3	11(11)	325.1
N/GN	7.6	11(11)	325.1
GN	6.3	15(15)	325.1
HYB	6.6	15(15)	325.1

TABLE 7

Initial Estimates and Estimates based
on the Least-Squares and Box-Draper Criteria
for the Full Newton Method for Example 3

Parameters	Initial ln(Estimates)	Box-Draper ln(Estimates)	Least-Squares ln(Estimates)
k_{10}	1.0	1.3400	1.1767
k_{20}	1.0	1.1057	1.0885
k_{-10}	6.4	4.5572	4.6949
β_1	0.0	-1.7634	-1.9284
β_2	0.0	-2.4131	-2.5927
E_1	9.9	8.6016	8.7173
E_2	9.9	9.1632	9.1536
E_{-1}	9.9	9.4142	9.4095

TABLE 8

Estimation Efficiency for Example 3
using the Least Squares Criterion (LS)

METHOD	CPU secs	NI(NOF)	LS
N	345.2	15(16)	0.258
N/GN	279.9	16(17)	0.258
GN	155.5	27(27)	0.258
HYB	158.2	27(27)	0.258

TABLE 9

Estimation Efficiency for Example 3
using the Box-Draper Criterion (BD)

METHOD	CPU secs	NI(NOF)	BD
N	407.5	18(19)	0.0109
N/GN	356.8	19(20)	0.0109
GN	826.5	127(127)	0.0109
HYB	270.3	38(43)	0.0109

TABLE 10

Initial Conditions, Duration (in hours) and
Number of Observations for each Run
in Example 4

Run #	X_0	S_0	A_0	E_0	Duration	# Obs.
1	57.0	0.037	0.089	0.0071	72.0	14
2	77.0	0.036	0.145	0.0116	72.0	14
3	20.0	0.1	1.0	0.01	43.5	16
4	105.0	0.2	0.0	0.02	40.0	16
5	50.0	0.2	2.0	0.02	30.0	15
6	200.0	0.2	2.0	2.0	21.0	9

TABLE 11

Actual Parameter Values, Initial Estimates and Estimates Obtained with the Full Newton Method for Example 4

Parameters	ln(Actual Values)	ln(Initial Estimates)	ln(Estimates)
μ_{maxS}	-0.22314	0.1	-0.23402
μ_{maxA}	-1.6094	-2.0	-1.5951
k_d	-2.6173	-3.0	-2.5703
Y_S	1.9661	2.0	1.9613
Y_A	-0.90422	-1.0	-0.90792
Y_{AA}	2.3026	2.0	2.3118
Y_E	-0.33648	0.1	-0.35931
k_{evap}	-3.4483	-4.0	-3.4457
K_S	-3.3524	-4.0	-3.5393
K_A	-0.69150	-1.0	-0.71981
K_{IS}	1.9760	1.0	2.5277

TABLE 12

Estimation Efficiency for Example 4 using the Box-Draper Criterion (BD)

METHOD	CPU secs	NI(NOF)	BD
N	4359.4	20(21)	91.413
N/GN	3801.6	20(21)	91.413
GN	3498.8	106(109)	91.413
HYB	2887.3	86(89)	91.413

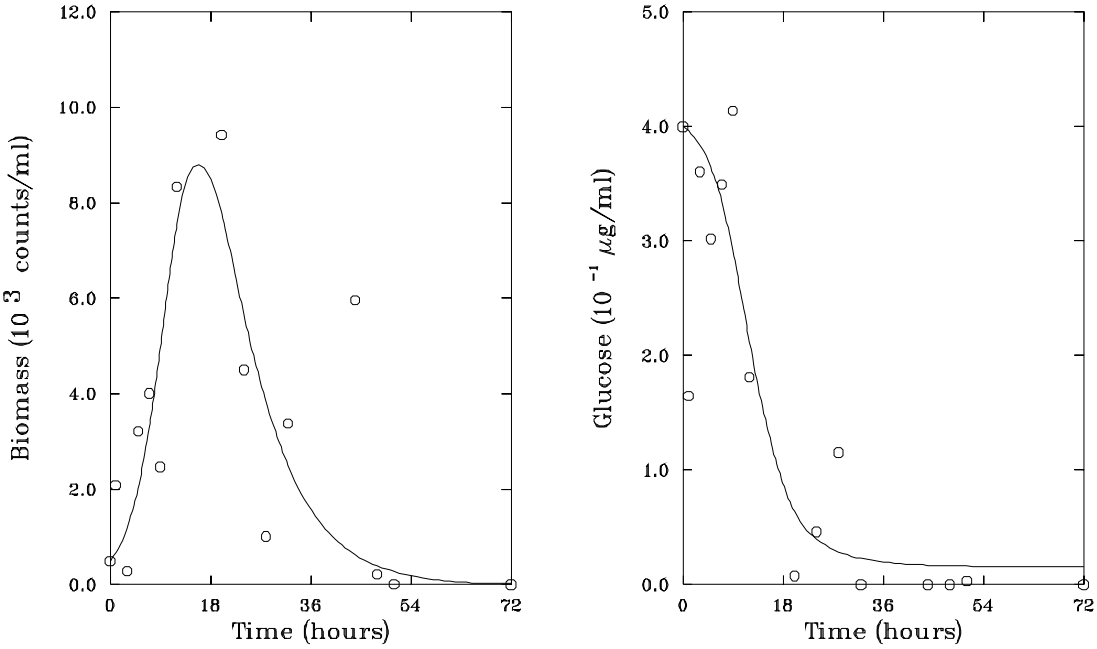

Figure 1 - Predicted (curve) and observed responses (empty circles) for Example 2.

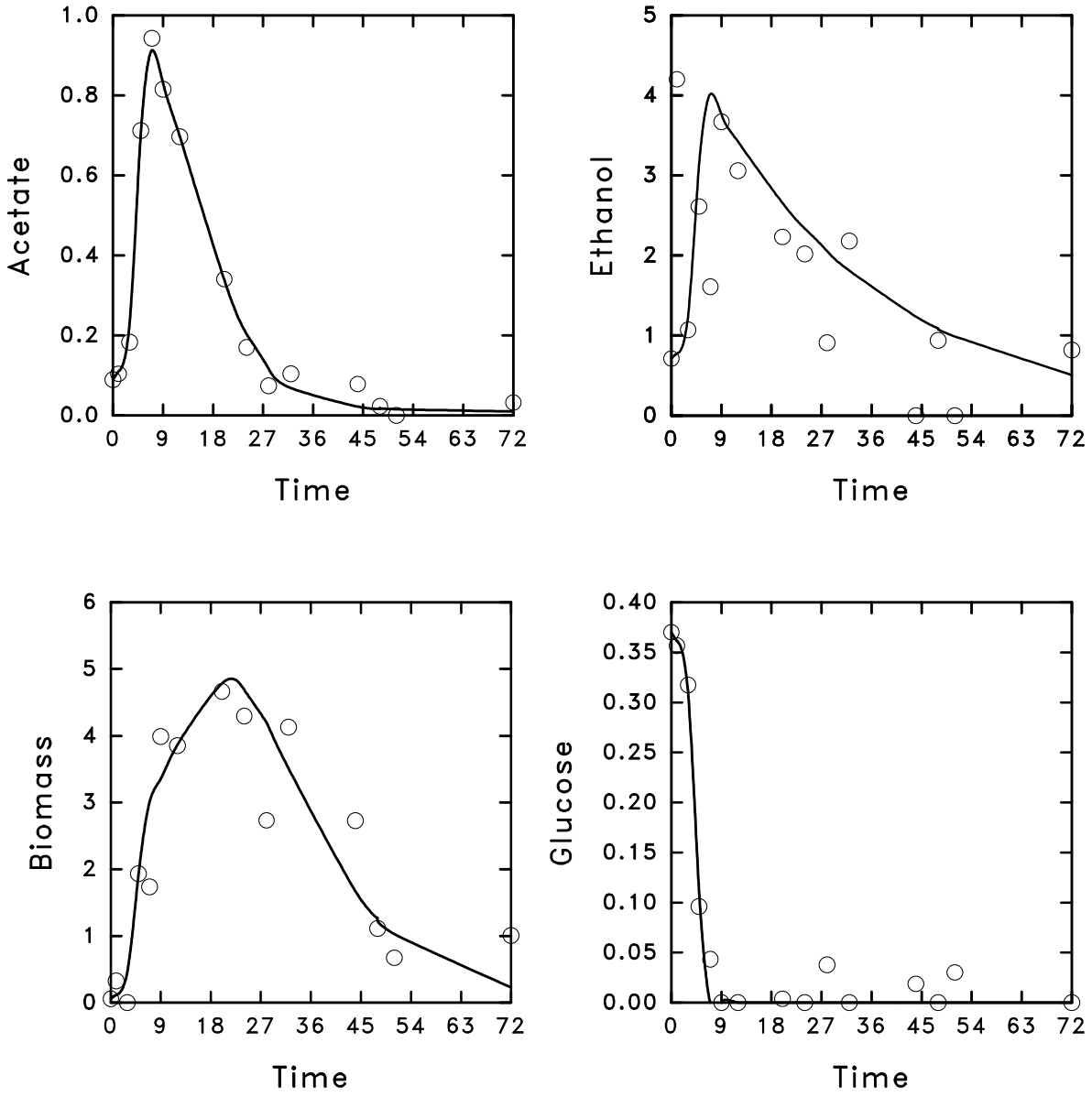

Figure 2 - Plot of observed (·) and predicted (———) responses for Run 1 in Example 4.