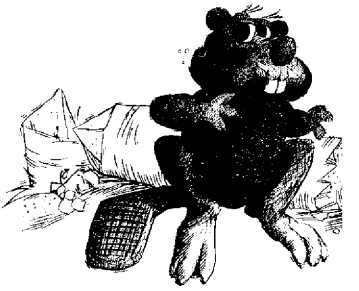
UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT



A Computational Approach to Parameter Estimation in Ordinary Differential Equations

James M. Varah

CS-82-58

December, 1982

A Computational Approach to Parameter Estimation in Ordinary Differential Equations

J.M. Varah

Computer Science Department
University of Waterloo
Waterloo, Ontario, Canada
N2L 3G1

ABSTRACT

We consider computational aspects of the problem of fitting parameters in differential equation models. We first review three basic techniques, which all involve least squares estimation. We then discuss the important problem of sensitivity analysis of the parameters, and give both a local, mathematical approach and a more global, computational approach. We then present two numerical examples, and present an alternative technique for the special case of the two-dimensional Lotka/Volterra equations.

1. INTRODUCTION

The general problem we are concerned with can be described as follows: we are given a system of differential equation

$$\underline{y}' = \underline{f}(t, \underline{y}, p) \tag{1.1}$$

$$\underline{y} = (y_1, \ldots, y_n) \ , \ p = (p_1, \ldots, p_m)$$

plus some data for the solution:

$$\underline{y}(t_i) = \underline{y}_i \ , \quad i = 1, \ldots, N$$

and we seek parameter values p so as to

$$\min_p \Phi(p) = \sum_{i=1}^{N} \sum_{j=1}^{n} (y_i(t_i; p) - y_{ij})^2. \tag{1.2}$$

The problem may also specify initial and/or boundary conditions $\underline{y}(t_0)$, $\underline{y}(t_f)$, with of course $t_0 \leq t_i \leq t_f$.

Such problems arise frequently in the modelling of various physical, chemical, and biological processes. Different algorithms have been proposed by many authors, and we review these in Section 2.

In addition to the minimization problem itself, one is normally also interested in the *sensitivity* of the parameters p to changes in the data. In

Section 3, we discuss the mathematical process of sensitivity analysis, and in addition describe a *computational* method of sensitivity analysis which we feel can make a valuable addition to the computational solution process. In Section 4 we present two numerical examples illustrating these ideas, and in Section 5 we present an alternative approach for the special case of the Lotka/Volterra equations.

2. THE METHODS

The basic methods in use presently are of three types:

(a) Initial Value (or Boundary Value) Technique:

where for any choice of p, we calculate $\Phi(p)$ by numerical integration using an initial value or boundary value method, assuming a full set of initial or boundary conditions is given. Thus with some initial choice of parameters p^0, we can proceed to minimize $\Phi(p)$ using some optimization technique, preferably one which does not require gradients explicitly. If a full set of initial or boundary conditions is not known, the set of parameters p must be extended to include these before the technique can be applied. This can cause extra difficulties; for example the initial estimate of the initial conditions may lead to a numerical integration which *diverges*, if the solution of the system (1.1) is sensitive to the choice of initial conditions.

Many authors use this method; we mention in particular van Domselaar and Hemker (1975), Bard (1974), and Benson (1979). The method can be extremely time-consuming because of the necessity of integrating the full set of equations for each choice of parameter value. Thus it must be used with caution: we suggest using it only when good estimates of the parameters are available (possibly from (b) below).

(b) Spline-Fitting Technique:

where we fit the data $\{y_{ij}\}$ by splines $s_j(t)$, $j = 1, \ldots, n$, and then find p to minimize the residual least squares function

$$\Phi_R(p) = \sum_{i=1}^{N} \sum_{j=1}^{n} (s_j'(t_i) - f_j(t_i, s, p))^2.$$

Of course, since this is a different function, its minimum will not in general be the same as for $\Phi(p)$, so this is really only a way of obtaining an approximate solution p. However, it seems to work well in practice, and has the advantage of not requiring costly numerical integrations of the DE system. We recommend this be used to get a good estimate of p, and this estimate *refined* by using method (a). References are Varah (1982a) and Swartz and Bremerman (1975). Notice also that when a full set of initial or boundary conditions is not known, the basic technique is the same; however, to estimate $\Phi(p)$, one still must integrate the DE using the given p, which requires a

minimization of Φ over the space of possible initial conditions.

(c) Explicit Integration:

where the DE system is explicitly integrated to functional form $F_j(\underline{y}, t, p) = 0$, $j = 1, ..., n$, and the least squares function $\Phi(p)$ computed explicitly by solving for each $\underline{y}(t_i)$. Of course, this is only possible for certain equations, but when it is, we recommend it; it avoids numerical integration completely, and we have found the results from this method more reliable in general. Again, if a full set of initial or boundary conditions is not known, one is left with integration constants to solve for; however this can be done using the data (see Section 5 for an example).

3. SENSITIVITY ANALYSIS

The problem (1.1) is really much more complicated than it may appear: in a particular problem the least squares surface $\Phi(p)$ may not be convex; there may be more than one local or global minimum; even if the minimum is unique, the surface may be very flat in some directions, so that some of the values p_i are indeterminate. All of the above can and do occur in the simplest problem of fitting parameters in linear constant coefficient systems; this problem is essentially equivalent to fitting data by a sum of exponentials, and examples and references are provided in Varah (1982a).

Thus we feel it is imperative that, along with any parameter estimates which are produced, estimates of how well-determined or how sensitive the parameters are be given as well. In what follows we shall describe both mathematical and computational approaches to this sensitivity analysis.

A. Mathematical Approach

This technique involves first of all obtaining an estimate of the Hessian matrix $H(p, y)$ of the function $\Phi(p)$ at the minimum point $p = p^*$. This is done by integrating the *sensitivity equations*: if we define $Z_{ij} = \dfrac{\partial y_i}{\partial p_j}$, the Z_{ij} satisfy the linear equation

$$\frac{d}{dt}\left(\frac{\partial y_i}{\partial p_j}\right) = \frac{\partial f_i}{\partial p_j} + \sum_k \frac{\partial f_i}{\partial y_k}\,\frac{\partial y_k}{\partial p_j} \tag{3.1}$$

or

$$Z' = G + J\,Z \ ,$$

which can be integrated for any value of p and $\underline{y}(t)$ assuming initial (or boundary) conditions are known, say $Z(t_0)$. Of course with fixed initial conditions $y(t_0)$, this is no problem: we can set $Z(t_0) = 0$. However if we do not have a complete set of fixed initial or boundary conditions, it is not clear

what to use for $Z(t_0)$, and it may be necessary to assume some probabilistic model for $Z(t)$. In the fortuitous case that the original equations can be explicitly integrated (see 2(c) above), the situation is much simpler: differentiating $F_i(t, y, p) = 0$ with respect to p_j gives

$$\sum_k \frac{\partial F_i}{\partial y_k} \frac{\partial y_k}{\partial p_j} + \frac{\partial F_i}{\partial p_j} = 0$$

or

$$J Z + G = 0 ,$$

which can be solved explicitly for $Z(t)$.

Once $Z(t)$ has been computed for the minimum $p = p^*$ and solution $\underline{y}(t)$, the Hessian of $\Phi(p)$ is given by

$$H_{kl}(p, y) = \sum_{i=1}^{N} \sum_{j=1}^{n} Z_{jl}(t_i) Z_{jk}(t_i) + \sum_{i=1}^{N} \sum_{j=1}^{n} (y_j(t_i) - y_{ij}) \frac{\partial^2 y_j(t_i)}{\partial p_k \partial p_l} . \quad (3.2)$$

In practice, the second term is usually ignored in the calculation of H; this is really only justified for problems with small residual or problems which are nearly linear.

Recently, efforts have been made (in the setting of general nonlinear least squares problems) to use both terms, the first as a measure of parameter - dependent sensitivity, and the second as a measure of intrinsic sensitivity. See Ramsin and Wedin (1977) and Bates and Watts (1980). We take the view here that we are given a particular model (1.1) with both kinds of sensitivity present, and that any results on sensitivity should be for this fixed model.

Once the Hessian has been found, there are two ways of using it to estimate parameter sensitivity; both ways depend basically on the ill-condition of H. Assume for simplicity of notation that $n = 1$. The first method allows p to vary from p^* (the minimum point) as long as the least squares function $\Phi(p)$ does not change appreciably: that is,

$$\Phi(p) - \Phi(p^*) \leq \epsilon . \quad (3.3)$$

This is the approach used by Bard (1974, pg. 71); who shows that if we assume local quadratic behaviour behaviour for $\Phi(p)$ near p^*, then (3.3) leads to the ellipsoid in p-space

$$\frac{1}{2}(p - p^*)^T H(p - p^*) \leq \epsilon .$$

Thus the greatest uncertainty in $(p - p^*)$ is $\frac{2\epsilon}{\lambda_{\min}(H)}$, in the direction of the eigenvector corresponding to $\lambda_{\min}(H)$. This method is also used by van Domselaar and Hemker (1975), who manage to define ϵ very precisely by assuming a normal distribution for the errors in the data $\{y_i\}$, giving $\epsilon = \frac{N}{N-m} \Phi(p^*) F_\alpha(m, N-m)$, where $F_\alpha(m, N-m)$ is the upper α probability point of the Fisher distribution with m and $N-m$ degrees of

freedom.

A second approach which uses the Hessian is to allow a perturbation in each data value $y_i \to y_i(1+\epsilon_i)$, with ϵ_i measuring the noise or uncertainty in the data. Then we can argue that instead of minimizing $\Phi(p) = \sum_1^N (y(t_i,p)-y_i)^2$ to get $p = p^*$, we could just as well minimize

$$\Phi_\epsilon(p) = \sum_1^N (y(t_i,p)-y_i(1+\epsilon_i))^2$$

to get $p = p_\epsilon^* = p^* + \delta_\epsilon$. To measure δ_ϵ, we can proceed as follows:

$$\Phi_\epsilon(p) = \Phi(p) - 2\sum_1^N \epsilon_i y_i (y(t_i)-y_i) + O(\|\underline{\epsilon}\|_2^2)$$

so at the minimum $p = p_\epsilon^*$,

$$0 = \frac{\partial}{\partial p_k}(\Phi_\epsilon(p)) = \frac{\partial}{\partial p_k}(\Phi(p)) - 2\sum_1^N \epsilon_i y_i \frac{\partial}{\partial p_k}(y(t_i,p)-y_i) + O(\|\underline{\epsilon}\|_2^2) \ .$$

Recall that $\frac{\partial}{\partial p_k}(y(t_i)) = Z_k(t_i)$, which is singly subscripted here because we are assuming $y(t)$ has only one component (for notational simplicity only). Thus we can write the above equation as

$$0 = 2\sum_1^N (y(t_i,p)-y_i)Z_k(t_i) - 2\sum_1^N \epsilon_i y_i Z_k(t_i) + O(\|\underline{\epsilon}\|)_2^2) \ .$$

Now since $p = p_\epsilon^* = p^* + \delta_\epsilon$,

$$y(t_i,p) = y(t_i,p^*) + (\underline{Z}(t_i))^T \underline{\delta}_\epsilon + O(\|\underline{\delta}_\epsilon\|_2^2) \ .$$

Substituting this in the above equation and keeping only first order terms gives

$$H\underline{\delta}_\epsilon = \underline{b} \tag{3.4}$$

where

$$H = \sum_1^N (\underline{Z}(t_i))^T \underline{Z}(t_i) \ , \quad \underline{b} = \sum_1^N \epsilon_i y_i \underline{Z}(t_i) \ .$$

Notice that H, which here is only the first term of the Hessian, can be written as J^TJ, with J the $N \times m$ Jacobian matrix $J_{ik} = Z_k(t_i)$. Thus $\underline{\delta}_\epsilon$ is the least squares solution to the overdetermined system

$$J\underline{\delta}_\epsilon = \underline{\eta} \tag{3.5}$$

where $\eta_i = \epsilon_i y_i$. If we use the singular value decomposition $J = U\Sigma V^T$ (notice $H = J^TJ = VD^2V^T$) then again the largest deviations in p are in the direction of the smallest eigenvector of H.

In a practical situation, we will only have a *bound* for each $|\eta_i|$. However, using $J = U\Sigma V^T$, we obtain

$$(\delta_\epsilon)_k = \sum_{k=1}^m V_{ik} \frac{u_i}{\sigma_i}$$

where $u = U^T \eta$. Thus we get the bound

$$| (\delta_\epsilon)_k | \leq \| \eta \|_2 \sum_{k'=1}^{m} \frac{| V_{ik} |}{\sigma_i} \tag{3.6}$$

which we have found useful in practice. One can actually store J and compute the SVD to avoid the extra ill-conditioning in H, but it is probably more practical to keep only H which is only $m \times m$, and use its eigenvector decomposition to get (3.6).

B. Computational Approach

Although the above techniques can indicate when parameters are sensitive, a serious drawback is that they are only *local* techniques. Often it is much more enlightening to have a more *global* picture of the relative changes in $\Phi(p)$ as p varies. One way of doing this is to plot contours or surfaces of $\Phi(p)$ projected into two-dimensional planes; that is, we vary just two parameters (p_{i_1}, p_{i_2}) holding the others fixed, and measure $\Phi(p)$ for these parameters varying between given limits. We have built this facility into our parameter estimation program, and find that, although this is time-consuming, it provides a very useful tool for gaining insight into parameter sensitivity of a particular model. We will show some results of this technique in the examples following. See also the surface plots in Varah (1982a).

4. TWO EXAMPLES

The first of our examples is a pharmacology problem modelling the absorption of ethanol in the bloodstream (Ralston et al (1979)) which we also mentioned earlier (Example D of Varah (1982a)). The equation modelling ethanol concentration is

$$\frac{dz}{dt} = \theta - \frac{\theta_2 z}{\theta_3 + z} \quad , t \leq 2$$

$$\frac{dz}{dt} = -\frac{\theta_2 z}{\theta_3 + z} \quad , t > 2$$

with $z(0) = 0$ and $0 \leq t \leq 8.5$. The data are given in both the above references, and the qualitative behaviour of $z(t)$ is that it rises almost linearly to $t = 2$ and then decays back to zero.

One can of course integrate numerically and use techniques (a) or (b) to minimize a least squares residual, but it is possible to integrate explicitly, as we indicated in Varah (1982a): for $t \leq 2$, solve

$$\theta_1 (e^x - 1) - \theta_2 x = \frac{(\theta_1 - \theta_2)^2 t}{\theta_3} t$$

for x, and set $z(t) = \dfrac{\theta_1\theta_2(e^x-1)}{\theta_1-\theta_2}$. (This involves, for each t, solving a nonlinear equation for x which has two roots. Since we need $z(t) > 0$, we take the positive root if $\theta_2 < \theta_1$, and the negative root if $\theta_2 > \theta_1$. Even the case $\theta_2 = \theta_1$ can be handled by using limits: $z(t) \to \theta_1 t$ as $\theta_2 \to \theta_1$.) for $t > 2$, solve

$$\theta_3 x + z(2)(e^x - 1) = (2-t)\theta_2$$

For x and set $z(t) = z(2)e^x$.

This problem has a unique minimum near $\theta^* = (.557, .221, .151)$ with $\sqrt{\Phi(\theta^*)} = 0.300$. However the least squares surface is very flat in the θ_2/θ_3 plane, with a narrow trough running roughly along the line $\theta_3 = 2.5\theta_2 - 0.4$. We present a surface plot of this in Figure 1, with $0.1 \le \theta_2 \le 2.0$, $0.1 \le \theta_3 \le 5.0$. Thus the parameters are very sensitive in this direction, with other values

$$\begin{aligned} \theta^{(1)} &= (.58, .78, 1.54) , \quad \sqrt{\Phi} = .3236 \\ \theta^{(2)} &= (.607, 1.29, 2.76) , \quad \sqrt{\Phi} = .3243 . \\ \theta^{(3)} &= (.61, 2.03, 4.63) , \quad \sqrt{\Phi} = .3286 \end{aligned}$$

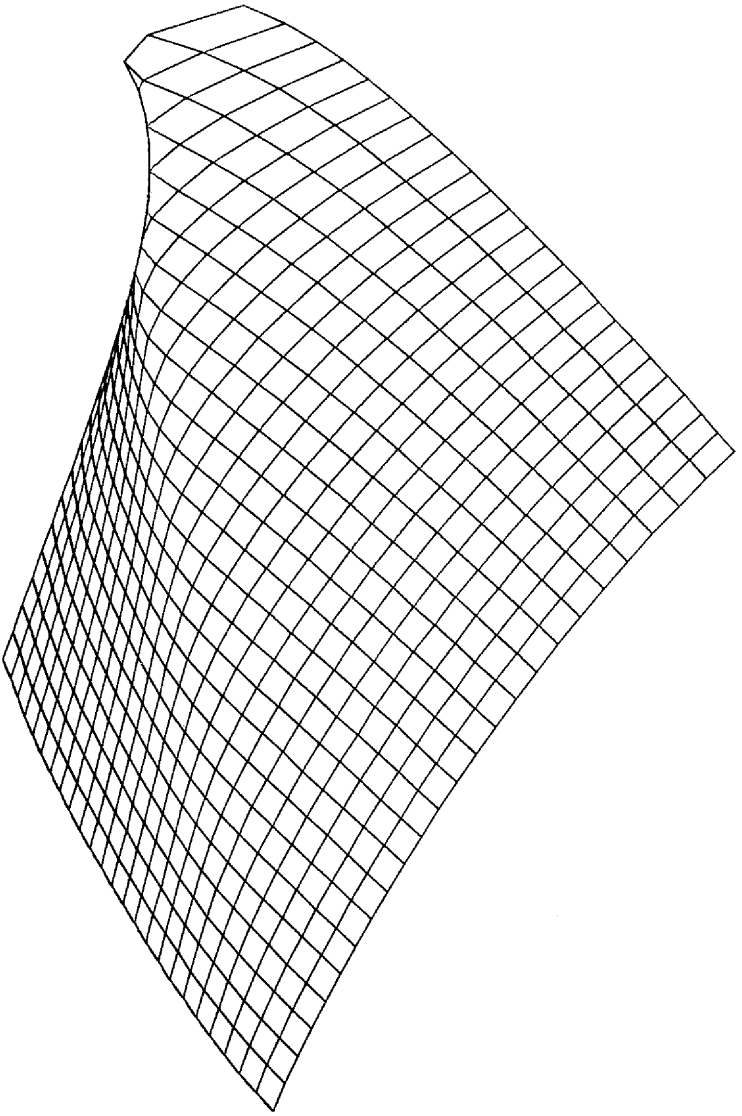


Figure 1

The first of these was found as the minimum using the spline least squares technique (b) in Varah (1982a).

However, this sensitivity, readily apparent from the surface plot, is not apparent from solving the sensitivity equations and finding the Hessian; the singular values of J, measured at $\theta = \theta*$, are $\sigma_i = .24, 5.4, 30$. However, at $\theta = \theta^{(1)}$, we get $\sigma_i = .034, 1.2, 18$ and the sensitivity is more noticeable.

As a second example (and we thank Robert Miura for the reference), we use the Hodgkin-Huxley (1952) equations modelling nerve action potentials, as presented in Hafner et al (1981). The model equations (for the squid axon) are

$$\left.\begin{aligned} \frac{dS_1}{dt} &= -[p_1 S_2^3 S_4 (S_1 + 115) + p_2 S_3^4 (S_1 - 12) + p_3 (S_1 + 10.6)] \\ \frac{dS_2}{dt} &= \alpha_m - (\alpha_m + \beta_m) S_2 \\ \frac{dS_3}{dt} &= \alpha_n - (\alpha_n + \beta_n) S_3 \\ \frac{dS_4}{dt} &= \alpha_h - (\alpha_h + \beta_h) S_4 \ . \end{aligned}\right\} \quad (4.1)$$

Here $S_1(t)$ denotes the membrane potential and $S_2(t), S_3(t), S_4(t)$ are excitation variables; the α's and β's are rate constants varying with the membrane potential as follows:

$$\begin{aligned} \alpha_m &= 0.1(25 + S_1) / (\exp*(0.1(S_1 + 25)) - 1) \\ \beta_m &= 4 \exp*(S_1 / 18) \\ \alpha_n &= 0.01 (S_1 + 10) / \exp*(0.1(S_1 + 10)) - 1) \\ \beta_n &= 0.125 \exp*(S_1 / 80) \\ \alpha_n &= 0.07 \exp*(0.05 S_1) \\ \beta_n &= 1 / (\exp*(0.1(S_1 + 30)) + 1) \ . \end{aligned}$$

The parameters p_1, p_2, p_3 represent maximum conductivity for the three different ionic currents present. Initial values for the state variables are $S_1(0) = 0$, $S_2(0) = 0.1$, $S_3(0) = 0.2$, $S_4(0) = 1.0$ (note: $S_2(0)$ and $S_3(0)$ are reversed in Hafner et al). For realistic values of the parameters, the membrane potential decreases from zero to a minimum rather sharply, increases to a slightly positive value, and then decays to zero over a time interval of about 10 milliseconds.

Hafner et al generate 100 data points by simulation: they fix the parameters at appropriate values $(p_1 = 120, p_2 = 36, p_3 = 0.3)$, and integrate the equations (4.1) at these parameter values, using the numerical values obtained as data after adding random normal noise. They then try to recover the original parameter values by minimizing a least squares function $\Phi(p)$. We are more interested in parameter sensitivity: therefore we derive data for $S_1(t)$ rather arbitrarily (as in Table 1), and once we find a minimum near the above parameter values, we

measure deviation in $\Phi(p)$ as we vary p two components at a time in a region around the minimum.

t	0.5	1.0	1.5	2.0	2.5	3.0	3.5
$S_1(t)$	-5.0	-10.	-30	-100	-100	-70	-50
t	4.0	4.5	5.0	5.5	6.0	7.0	8.0
$S_1(t)$	-30	0	10	10	10	10	10

Table 1

Because of the computational time involved in the numerical integrations, we did not produce surface plots as we did in the previous example. Indeed we merely computed contours in the three planes p_1/p_2, p_1/p_3, p_2/p_3. The results however were interesting: although p_1 and p_2 seem well-determined, the value of p_3 appears to be rather sensitive; in the p_1/p_3 plane, the value of $\Phi(p)$ varied by less than 10% when p_3 was changed by 50%.

5. THE LOTKA/VOLTERRA EQUATIONS: AN ALTERNATIVE APPROACH

Now consider the two variable Lotka/Volterra model

$$\left.\begin{aligned} x'(t) &= rx - \alpha xy \\ y'(t) &= -sy + \beta xy \end{aligned}\right\} \tag{5.1}$$

with positive parameters $p = (\alpha, \beta, r, s)$, and positive data values (x_i, y_i), $i = 1, \ldots, N$. This is a common model in ecology and other areas (see e.g. Clark (1976, pg. 194)). If the data are given as $(x(t_i), y(t_i))$, one can proceed as above, treating the problem as a two-variable system of equations and forming $\Phi(p)$ as in (1.2). However, we would like to present an alternative procedure for parameter estimation for this equation which may be more meaningful in some cases, and may also be used for other two-dimensional autonomous systems.

The equations (5.1) can be explicitly integrated to the implicit function form

$$g(x, y) = r \log y - \alpha y + s \log x - \beta x - C = 0 \tag{5.2}$$

which defines a simple closed curve in the positive x–y quadrant for any positive values of the parameters α, β, r, s. As the integration constant C increases, the solution curve shrinks, finally to the single point $x = s/\beta$, $y = r/\alpha$ for $C = r(\log(r/\alpha)-1)+s(\log(s/\beta)-1)$. Thus, for given C, α, β, r, s, we can use (5.2) to find the solution curve numerically by first finding the limiting x-values (the two x-solutions of (5.2) for $y = r/\alpha$), and then for each x between these limits, finding the two corresponding y-values by solving (5.2) as a function of y for this x-value.

However, our object of course is to find that choice of parameters for which the above closed curve best fits the given data in some sense. Since the integration constant C only affects the overall size of the closed curve, we define it in terms of the other parameters and the data as follows:

$$C = \frac{1}{N}\sum_i (r \log y_i - \alpha y_i + s \log x_i - \beta x_i) \quad , \tag{5.3}$$

so that the equation (5.2) holds in the mean for the given data. This choice guarantees that we have a nontrivial curve, and moreover that the centre of mass of the data $(x^*, y^*) = (\frac{1}{N}\sum_i x_i, \frac{1}{N}\sum_i y_i)$ is inside the curve, since

$$\begin{aligned} g(x^*, y^*) &= r \log(\tfrac{1}{N}\textstyle\sum y_i) - \frac{\alpha}{N}\sum_i y_i + s \log(\frac{1}{N}\sum x_i) - \frac{\beta}{N}\sum x_i - C \\ &= r\left[\log\left(\frac{\sum y_i}{N}\right) - \frac{\sum \log y_i}{N}\right] + s\left[\log(\frac{\sum x_i}{N}) - \frac{\sum \log x_i}{N}\right] \\ &\geq 0 \end{aligned}$$

using the arithmetic-geometric mean inequality. Notice that the interior of the curve is described by $g(x, y) > 0$. Of the other parameters, one is merely a scaling factor, which we fix by setting $\alpha = 1$, leaving free parameters β, r, s. In our experience, fitting all three parameters from the data is too ill-conditioned, and we normally fix β and fit only the two parameters r and s. Notice also that the centre of the closed solution curve $(x, y) = (s/\beta, y/\alpha)$ will match the centre of mass of the data (x^*, y^*) when $r = \alpha y^*$, $s = \beta x^*$, and these values are usually good initial estimates for the "best" r and s.

To define the "best" values of the parameters, we need a least squares function like $\Phi(p)$ which measures the deviation of the closed curve from the data points. We define this as follows, with thanks to David Kirkpatrick: let d_i be the distance from (x_i, y_i) to the solution curve along the ray from the centre of mass (x^*, y^*). One needs to solve a simple scaler nonlinear equation to find the intersection points of the ray with the curve; then d_i is the smaller of the distances. We then define

$$\Phi(r, s; \beta) = \frac{1}{N}(\sum_i d_i^2)^{1/2} \quad . \tag{5.4}$$

Each d_i will be small if the point (x_i, y_i) is close to the solution curve, and if all are close, the curve will fit the data well.

With this choice of Φ, we can proceed to find best choices of r and s for a fixed β, and to give surface plots measuring the sensitivity of the parameters. Our first example is for the Barnes data given by van Domselaar and Hemker (1975), and also used in Varah (1982a). We fix $\beta = 1$ and give in Figures 2 and 3 the solution curves and data for $(r, s) = (1.0, 2.0)$ and $(r, s) = (0.8, 2.0)$. Then we minimize $\Phi(r, s; 1)$ giving $(r, s) = (.377, .832)$ with the solution shown in Figure 4. To show that the results are very insensitive to the choice of β (and thus that β can not be well-determined by the data), we tried various values of β, and for each we minimized Φ over r and s. The results are shown in Table 2; the minimum value of Φ changes only very slightly with β, and as well the corresponding solution curves are very similar.

β	best r	best s	$\Phi(r, s; \beta)$
0.3	.371	.238	.0272
0.4	.375	.323	.0266
0.5	.377	.407	.0264
0.6	.379	.492	.0265
0.7	.379	.577	.0267
0.8	.379	.663	.0270
0.9	.379	.748	.0274
1.0	.378	.832	.0277
1.1	.377	.917	.0282

Table 2

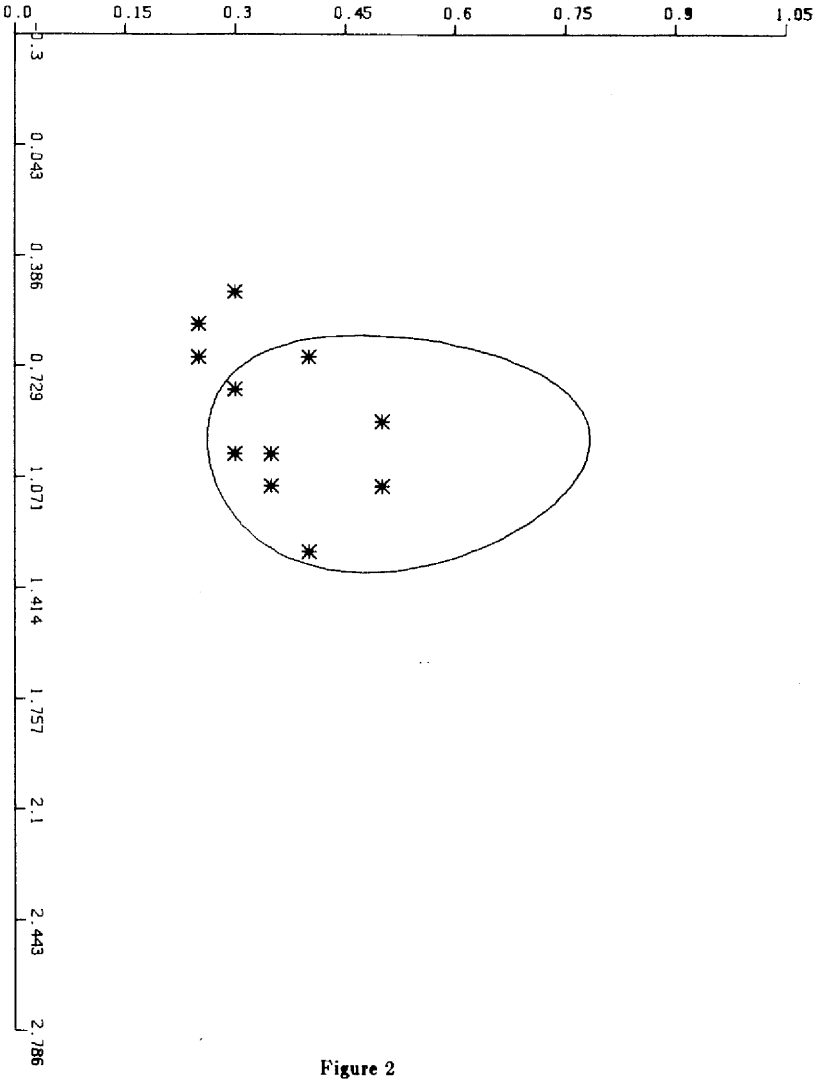

0.0
0.15
0.3
0.45
0.6
0.75
0.9
1.05
-0.3
0.043
0.386
0.729
1.071
1.414
1.757
2.1
2.443
2.786

Figure 2

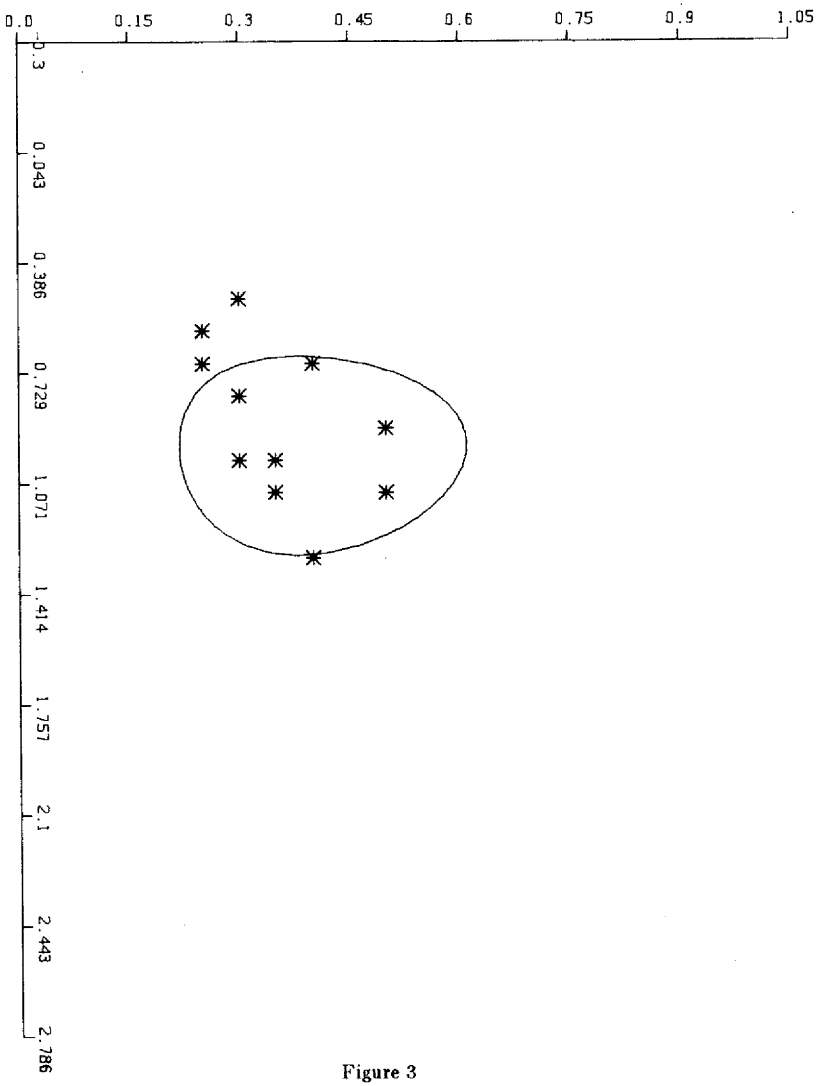

0.0
0.15
0.3
0.45
0.6
0.75
0.9
1.05
0.3
0.043
0.386
0.729
1.071
1.414
1.757
2.1
2.443
2.786

Figure 3

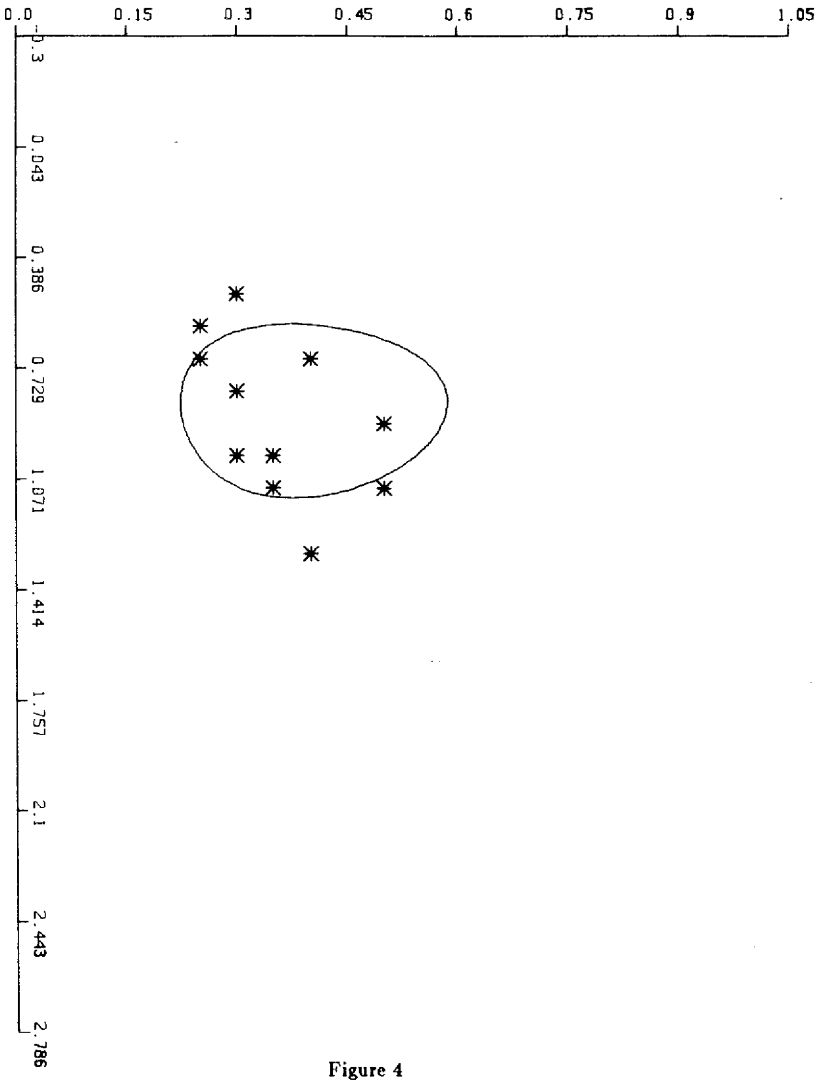

0.0
0.15
0.3
0.45
0.6
0.75
0.9
1.05
-0.3
0.043
0.386
0.729
1.071
1.414
1.757
2.1
2.443
2.786

Figure 4

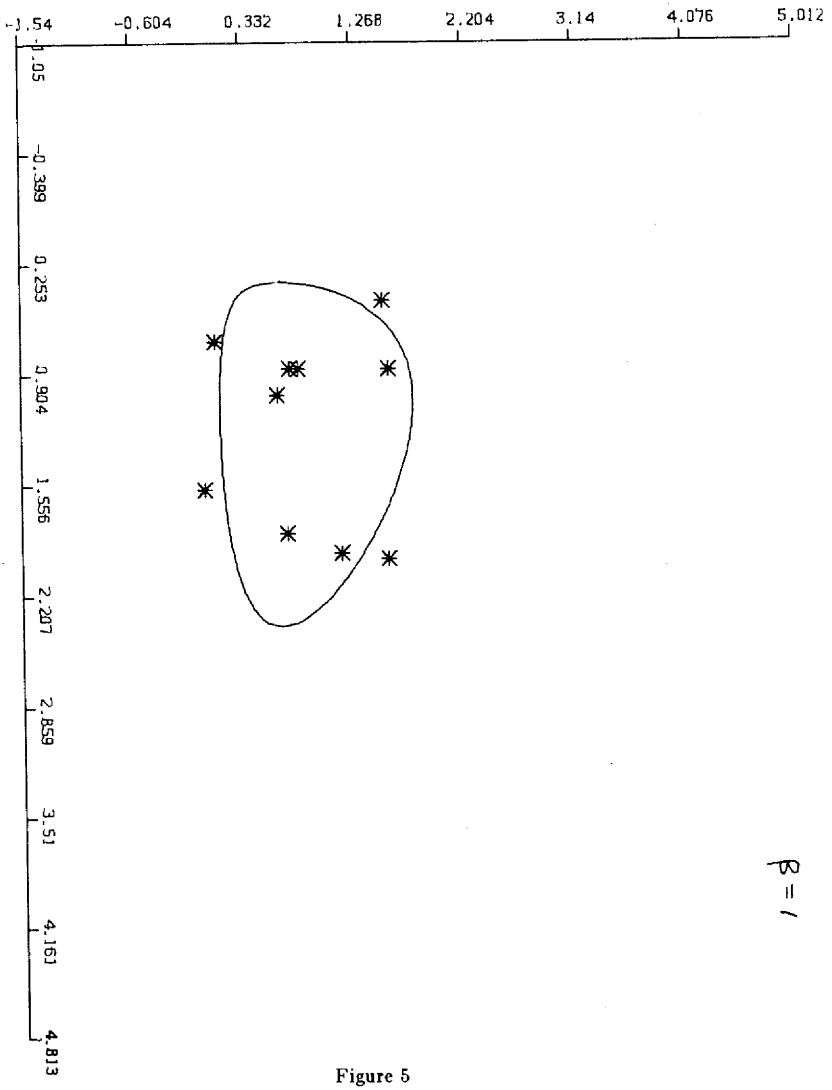

-1.54
-0.604
0.332
1.268
2.204
3.14
4.076
5.012
-1.05
-0.399
0.253
0.904
1.556
2.207
2.859
3.51
4.161
4.813

Figure 5

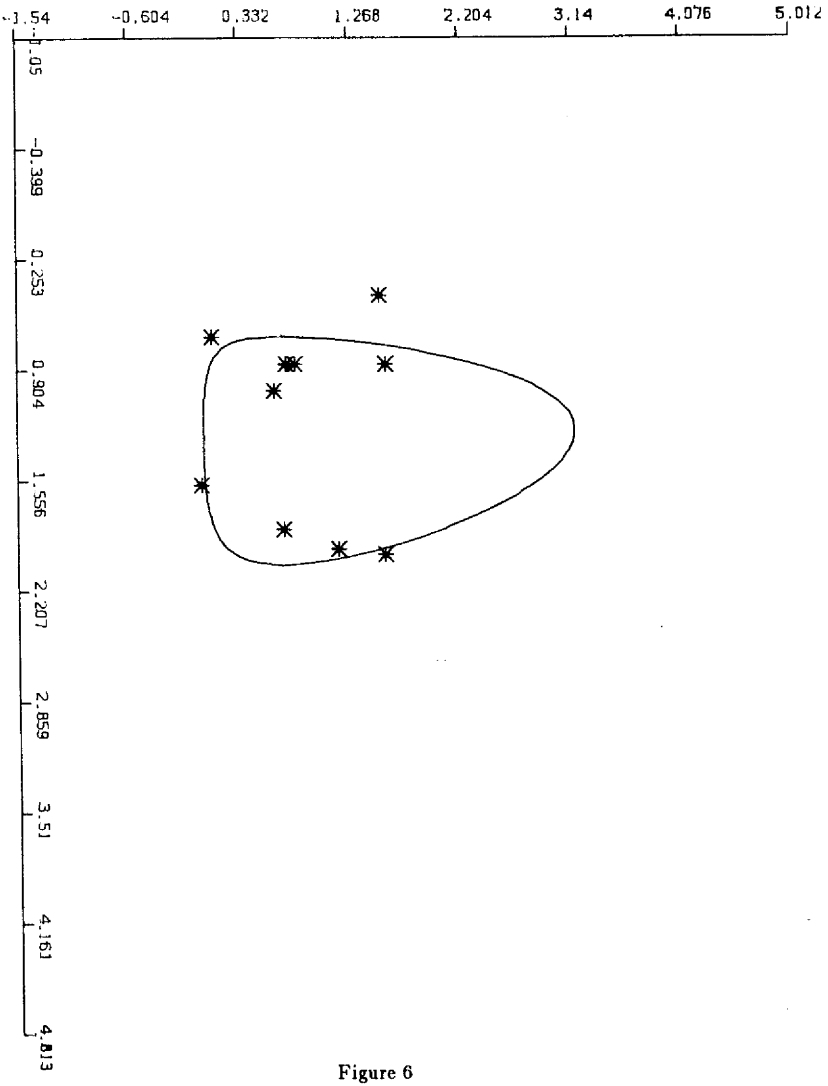

-1.54
-0.604
0.332
1.268
2.204
3.14
4.076
5.012
-0.05
-0.399
0.253
0.904
1.556
2.207
2.859
3.51
4.161
4.813

Figure 6

Now consider the data set in Table 3, obtained from an ellipse in the x–y plane by adding random noise at the 0.5 level in both x and y coordinates.

x	y	x	y
.87	.82	.87	.74
.71	.12	.47	1.53
1.96	1.17	1.58	.02
1.84	.72	1.99	1.57
1.03	.64	.87	1.58

Table 3

In Figures 5 and 6 we give the optimal solutions for $\beta = 1$ (for which $r = .67$, $s = 1.07$, $\Phi = .116$) and $\beta = 8$ ($r = .71$, $s = 10.1$, $\Phi = .071$). In this case, as β increases, the minimum $\Phi(r, s; \beta)$ slowly decreases, but the corresponding solution curves change drastically, giving larger and larger y-values at the top of the curve. The reason is the large *gap* in the data at the top, which might be of practical relevance if this were a model of a predator/prey cycle; population measurements might not be possible at certain times of the year for example. In any case, there are no data points to keep the curve bounded in this region, and in fact for large β, slightly smaller values of $\Phi(r, s; \beta)$ occur for these solution curves. For much the same reason, with a large fixed value of β, the surface plot of $\Phi(r, s; \beta)$ is very flat in the r-direction. Although the solution curves are very different again, the corresponding values of Φ do not change much as r changes. Thus this "gap" problem is much more ill-conditioned than the first data set.

6. REFERENCES

Y. Bard (1975). *Nonlinear Parameter Estimation.* Academic Press, New York.

D.M. Bates and D.G. Watts (1980). Relative Curvature Measures of Nonlinearity. *J. Royal Stat. Soc.* (Series B), *42*, 1-25.

M. Benson (1979). Parameter Fitting in Dynamic Models. *Ecol. Mod. 6*, 97-115.

C. Clark (1976). *Mathematical Bioeconomics.* Wiley-Interscience, New York.

D. Hafner, U. Borchard, O. Richter, and M. Neugebauer (1981). Parameter

Estimation in Hodgkin-Huxley Type Equations for Membrane Action Potentials in Nerve and Heart Muscle. *J. Theor. Biol. 91*, 321-345.

A.L. Hodgkin and A.F. Huxley (1952). A Quantitative Description of Membrane Current and Its Application to Conduction and Excitation in Nerve. *J. Phsiol. 117*, 500-544.

M.L. Rolston, R.I. Jennrich, P.F. Samson, and F.K. Uno (1979). Fitting Models Defined by Differential Equations. *Proc. 12th Annual Symposium on the Interface of Computer Science and Statistics.* J. Gentleman, ed.

H. Ramsin and P.A. Wedin (1977). A Comparison of Some Algorithms for the Nonlinear Least Squares Problem. *BIT 17*, 72-90.

J. Swartz and H. Bremerman (1975). Discussion of Parameter Estimation in Biological Modelling: Algorithms for Estimation and Evaluation of the Estimates. *J. Math. Biol. 1*, 241-257.

B. van Domselaar and P.W. Hemker (1975). Nonlinear Parameter Estimation in Initial Value Problems. Report NW18/75, *Math. Centrum*, Amsterdam.

J.M. Varah (1982a). A Spline Least Squares Method for Numerical Parameter Estimation in Differential Equation. *SISSC 2*, 20-46.

J.M. Varah (1982b). On Fitting Exponentials by Nonlinear Least Squares. Tech. Rep. 82-2, Computer Science Department, University of British Columbia, Vancouver, Canada.