A New Algorithm for the Estimation of Parameters in Ordinary Differential Equations

M. HWANG and J. H. SEINFELD

Department of Chemical Engineering
California Institute of Technology, Pasadena, California 91109

A new computational algorithm for the estimation of parameters in ordinary differential equations from noisy data is presented. The algorithm is computationally faster than quasilinearization because of the reduction of the number of ordinary differential equations that must be solved at each iteration. A modification is also presented to remove ill-posedness. The algorithm is illustrated on a simple example.

The estimation of parameters in a mathematical model from noisy experimental data is an important problem in process analysis and adaptive control. Let us consider the case in which the model consists of a set of ordinary differential equations. The estimation of parameters in ordinary differential equations has received much attention (*1* to *3*). Various numerical techniques, such as quasilinearization (*3* to *5*) and steepest descent (*6*) have been suggested and demonstrated in the literature for this problem.

When the observation process is continuous in time, the necessary conditions for minimization of the objective function result in a two-point boundary value problem, the solution of which is often highly sensitive to the initial guesses of the unspecified boundary conditions. When the observations are made at discrete times, the usual quasilinearization routine is an initial value problem. However, there are $(n + l)^2$ matrix differential equations involved, where n and l are the state and parameter vector dimensions, respectively, limiting the usefulness of the algorithm to small values of n and l. Finally, if the system is ill-posed, that is, when large changes in parameter values cause only small changes in the objective function, convergence of any algorithm is extremely unstable.

The objective of this paper is to present a new computational algorithm for the estimation of parameters in ordinary differential equations which alleviates the difficulties noted above. Specifically, we present an algorithm which, in the continuous measurement case, does not involve a two-point boundary value problem and which, in the discrete measurement case, requires fewer equations to be integrated per iteration than quasilinearization. Also, we show how ill-posedness may be simply removed. The only requirement that must be satisfied is that for the given set of error-free measurements there exists a unique set of parameters, that is, the system is observable with respect to the parameter vector.

PROBLEM STATEMENT

We consider a system governed by

$$\dot{x}(t) = f(t, x, p) \tag{1}$$

$$x(0) = x_0 \tag{2}$$

where x is an n-vector and p is an l-vector of constant parameters. The observations of the system are related to the state by

$$y(t) = h(t, x) + (\text{errors}) \qquad 0 \leq t \leq T \tag{3}$$

where y is an m-vector of observations. The problem is to find the value of p which minimizes a least square objective function. If the observations are taken continuously in time t, the objective function is

$$I(p) = \int_0^T ||y(t) - h(t, x(t; p))||^2_{Q(t)}\, dt \tag{4}$$

where the norm

$$||y(t) - h(t, x(t; p))||^2_{Q(t)} = [y(t) - h(t, x(t; p))]^T Q(t) [y(t) - h(t, x(t; p))] \tag{5}$$

and $x(t; p)$ denotes the solution of Equations (1) and (2). If the observations are made only at discrete times, $t_1, t_2, \ldots, t_s$, then I is given by

$$I(p) = \sum_{i=1}^{s} ||y(t_i) - h(t_i, x(t_i; p))||^2_{Q(t_i)} \tag{6}$$

Since Equation (4) can be reduced to Equation (6) by replacing $Q(t)$ by $Q(t)\,\delta(t - t_i)$, where $\delta(t - t_i)$ is the Dirac delta function, we will henceforth consider the objective function of Equation (4). We assume that f and h have continuous first-order partial derivatives with respect to their arguments. Finally, we assume that the error free system, no errors in Equation (3), is observable with regard to p, that is, there exists a unique value of the parameter vector p^* at which $I(p^*) = 0$.

A NEW ALGORITHM

Assume we have an initial guess $p^{(0)}$ which generates a trajectory of Equations (1) to (3) denoted by $x^{(0)}$. To describe the trajectory of Equations (1) to (3) generated by $p^{(0)} + \delta p^{(0)}$ we can linearize Equation (1) about $x^{(0)}$ if $\delta p^{(0)}$ is chosen such that for some $\eta > 0$

$$||x^{(1)}(t) - x^{(0)}(t)|| < \eta \tag{7}$$

Then we can write the following unique perturbation equations (*7, 8*)

$$\delta \dot{x}^{(0)} = f_x^{(0)} \delta x^{(0)} + f_p^{(0)} \delta p^{(0)} \tag{8}$$

$$\delta x^{(0)}(0) = 0 \tag{9}$$

$$\delta y^{(0)} = h_x^{(0)} \delta x^{(0)} \tag{10}$$

Correspondence concerning this paper should be addressed to J. H. Seinfeld.

where $f_x^{(0)}$ denotes $\left[\frac{\partial f}{\partial x}\right]_{x^{(0)},p^{(0)}}$ etc.

Substituting the solution of Equations (8) and (9) into Equation (10), we obtain

$$\delta y^{(0)}(t) = \theta^{(0)}(t)\,\delta p^{(0)} \tag{11}$$

where

$$\theta^{(0)}(t) = h_x^{(0)} D^{(0)}(t) \tag{12}$$

$$D^{(0)}(t) = \int_0^t \Phi^{(0)}(t,\tau) f_p^{(0)}(\tau)\, d\tau \tag{13}$$

and the fundamental matrix $\Phi(t,\tau)$ satisfies

$$\frac{\partial \Phi^{(0)}(t,\tau)}{\partial t} = f_x^{(0)}\,\Phi^{(0)}(t,\tau) \tag{14}$$

$$\Phi^{(0)}(t,t) = I \text{ (identity matrix)} \tag{15}$$

It can be easily shown that $D^{(0)}(t)$ is the matrix of sensitivity coefficients, $(\partial x/\partial p)_{x^{(0)},p^{(0)}}$, which satisfies

$$\dot{D}^{(0)} = f_x^{(0)} D^{(0)} + f_p^{(0)} \tag{16}$$

$$D^{(0)}(0) = 0 \tag{17}$$

Finally, $\delta p^{(0)}$ can be evaluated from

$$\delta p^{(0)} = K^{(0)}(T)^{-1} \int_0^T \theta^{(0)}(t)^T Q(t) \delta y^{(0)}(t)\, dt \tag{18}$$

where

$$K^{(0)}(T) = \int_0^T \theta^{(0)}(t)^T Q(t) \theta^{(0)}(t)\, dt \tag{19}$$

If $K^{(0)}(T)$ is nonsingular, there exists a unique perturbation $\delta p^{(0)}$ corresponding to the perturbation $\delta y^{(0)}$. This implies that the system of Equations (1) to (3) is locally observable with regard to $p^{(0)}$ at $(x^{(0)}, p^{(0)})$ (8).

The objective of the parameter estimation method is to use Equation (18) to improve an initial guess $p^{(0)}$ and subsequent values of p denoted $p^{(i)}$. However, in order to use Equation (18) we must pay special attention to $\delta y^{(i)}$. Theoretically, $\delta y^{(i)}$ is given by $y^{(i+1)} - y^{(i)}$, however, in any iteration $y^{(i+1)}$ is unknown because $p^{(i+1)}$ is unknown. Thus, to use Equation (18) we need to find a way of computing $\delta y^{(i)}$. Let us replace $\delta y^{(i)}$ by

$$\delta y^{(i)} = \epsilon\,[y(t) - h(t, x^{(i)})] \tag{20}$$

for some $\epsilon > 0$ such that Equation (7) is satisfied. Combining Equations (18) and (20) we obtain for the i^{th} iteration

$$\delta p^{(i)} = \epsilon\, K^{(i)}(T)^{-1} \int_0^T \theta^{(i)}(t)^T Q(t)\,[y(t) - h(t, x^{(i)})]\, dt \tag{21}$$

Equation (21) provides the basic algorithm to update $p^{(i)}$.

Recapitulating, we have determined a relation between a perturbation in the parameter vector and a perturbation in the observations. This relation is Equation (18). Then, assuming that the perturbation in the observations can be approximated by Equation (20) for a small enough ϵ, we obtained the key relation Equation (21). From a computational point of view, the new algorithm is used as follows:

1. Select ϵ, $Q(t)$ and an initial guess $p^{(0)}$.
2. Integrate Equations (1) and (2) to generate $x^{(0)}$.
3. Integrate Equations (16) and (17) and compute $\theta^{(0)}$ from Equation (12). Compute $K^{(0)}(T)$ from Equation (19) and $\delta p^{(0)}$ from Equation (21).
4. Continue steps 2 and 3 until $||\delta p^{(i)}||$ is less than a preset criterion.

Actually Equation (21) can be derived in a different way. At any step in the iteration Equation (4) can be written as

$$I(p) = \int_0^T ||y - h^{(i)} - (h - h^{(i)})||^2_{Q(t)}\, dt \tag{22}$$

Linearizing h about $h^{(i)}$

$$h - h^{(i)} = h_x^{(i)} D^{(i)} \delta p^{(i)} + 0(\delta p^{(i)2}) \tag{23}$$

Substituting Equation (23) into (22) and neglecting second-order terms

$$I(\delta p^{(i)}) = \int_0^T ||y - h^{(i)} - h_x^{(i)} D^{(i)} \delta p^{(i)}||^2_{Q(t)}\, dt \tag{24}$$

Using the stationary condition

$$\frac{\partial I}{\partial (\delta p^{(i)})} = 0 \tag{25}$$

and solving for $\delta p^{(i)}$, we obtain Equation (21) with $\epsilon = 1$.

RELATION OF THE NEW ALGORITHM TO QUASILINEARIZATION

A relevant question is what is the relation of Equation (21) to the quasilinearization approach for estimation of parameters in ordinary differential equations. We can show that the extension of the quasilinearization method applied to the case of discrete measurements will yield Equation (21) with proper choices of the boundary conditions for the homogeneous and particular solutions of the linearized state equation. This is demonstrated as follows.

If we adjoin to Equation (1) the l relations $\dot{p} = 0$ and define the $(n + l)$-vector $z = (x^T, p^T)^T$ then $z(t)$ satisfies

$$\dot{z} = g(t,z) \;\; ; \;\; z(0) = (x_0^T, p^T)^T \tag{26}$$

Linearizing Equation (26) about the i^{th} iterate z

$$\dot{z}^{(i+1)} = g_z^{(i)} z^{(i+1)} + g^{(i)} - g_z^{(i)} z^{(i)} \tag{27}$$

The solution of this equation can be written

$$z^{(i+1)}(t) = \Lambda^{(i)}(t)\,\alpha^{(i+1)} + \psi^{(i)}(t) \tag{28}$$

where the $(n + l) \times (n + l)$ matrix $\Lambda^{(i)}$ satisfies

$$\dot{\Lambda}^{(i)} = g_z^{(i)} \Lambda^{(i)} \tag{29}$$

and the $(n + l)$-vector $\psi^{(i)}$ satisfies

$$\dot{\psi}^{(i)} = g_z^{(i)} \psi^{(i)} + g^{(i)} - g_z^{(i)} z^{(i)} \tag{30}$$

The objective function is

$$I = \int_0^T ||y - h(\Lambda^{(i)}\,\alpha^{(i+1)} + \psi^{(i)})||^2_{Q(t)}\, dt \tag{31}$$

which is to be minimized with respect to $\alpha_k^{(i+1)}$, $k = n + 1, \ldots, n + l$. If the initial conditions of Equations (29) and (30) are taken as

$$\lambda_{kk}^{(i)}(0) = \begin{cases} 0 & k \leq n \\ 1 & k > n \end{cases} \tag{32}$$

$$\psi_k^{(i)}(0) = \begin{cases} x_{0k} & k \leq n \\ 0 & k > n \end{cases} \tag{33}$$

then Equation (28) becomes

$$x^{(i+1)}(t) = D^{(i)}(t) p^{(i+1)} + \psi^{(i)}(t) \tag{34}$$

Also Equations (30) and (33) yield

$$\psi^{(i)} = x^{(i)}(t) - D^{(i)}p^{(i)} \tag{35}$$

Substituting Equations (34) and (35) into (31), expanding h about $h^{(i)}$, minimizing with respect to $p^{(i+1)} - p^{(i)}$, we obtain Equation (21) with $\epsilon = 1$. The important point is that by the formulation of Equations (26) to (34), using the notions of quasilinearization, Equation (21) can be obtained.

The new algorithm has two important advantages over ones proposed to date. First, for continuous measurements in time, a two-point boundary value problem need not be solved. Second, for discrete measurements, there is a savings in computing time in using Equations (12) and (16) to (19) over Equation (28). The savings in time results from not having to solve Equations (29) and (30), only Equation (16).

Finally, let us examine the properties of the proposed scheme based on Equation (21) in the vicinity of a minimum of I, either local or global. For this purpose we rewrite Equation (24) as

$$I(\delta p^{(i)}) = \int_0^T ||y - h^{(i)}||^2_{Q(t)}\, dt$$

$$- 2\int_0^T \theta^{(i)}(t)^T Q(t)\,[y - h^{(i)}]\, dt\, \delta p^{(i)} + ||\delta p^{(i)}||^2_{K^{(i)T}} \tag{36}$$

or equivalently

$$I(\delta p^{(i)}) = \left|\left| \delta p^{(i)} - K^{(i)}(T)^{-1}\int_0^T \theta^{(i)}(t)^T Q(t) \right.\right.$$

$$\left.\left. [y - h^{(i)}]\, dt \right|\right|^2_{K^{(i)T}} + \int_0^T ||y - h^{(i)}||^2_{Q(t)}\, dt$$

$$- \left|\left| \int_0^T \theta^{(i)}(t)^T Q(t)\,[y - h^{(i)}]\, dt \right|\right|^2_{K^{(i)}(T)^{-1}} \tag{37}$$

Equations (36) and (37) represent an approximation to the I surface by a quadratic function of $\delta p^{(i)}$ in the neighborhood of $p^{(i)}$. The value of I corresponding to $p^{(i)}$ is the second term on the right-hand side of Equation (37) and can be denoted $I^{(i)}$. The minimum of $I(\delta p^{(i)})$ occurs at $\delta p^{(i)}$ given by Equation (21) with $\epsilon = 1$. Since the third term on the right-hand side of Equation (37) is positive, the value of I at the minimum of the quadratic approximation is decreased from $I^{(i)}$ by that amount. Normally, $p^{(i)}$ will not be in the neighborhood of a true minimum of I, so that Equation (21) provides a small step in the direction of decreasing I by approximating the actual I surface by the quadratic form of Equation (24). If, however, the minimum of $I(\delta p^{(i)})$ occurs at an actual minimum of I, then Equation (37) is exact to $0(\delta p^{(i)2})$ and the condition which holds at the minimum is

$$\int_0^T \theta^{(i)}(t)^T Q(t)\,[y - h^{(i)}]\, dt = 0 \tag{38}$$

The essential difficulty with convergence to a local minimum is embodied in Equation (38). Since Equation (38) holds at a local minimum, $\delta p^{(i)}$ becomes zero when computed from Equation (21). This is, of course, the problem encountered with all gradient-type methods. However, from Equation (38) we can propose a trial and error method of probing for the global minimum of I. Local minima in I will depend for their existence and location on the weighting matrix $Q(t)$ as well as on the nature of the system. The usual procedure to certify that one has found the global minimum of I is to start the algorithm from a number of different initial guesses $p^{(0)}$. If the dimension of p is large, this can be quite tedious. However, if the weighting matrix $Q(t)$ is changed, the location and number of local minima will, in general, change. Therefore, we can start the iterations for our chosen $Q(t)$ and continue until $||\delta p^{(i)}||$ is less than some criterion. Then, we change $Q(t)$ such that the time-weighting of the observations is different. If the algorithm had stopped at a local minimum, the change of $Q(t)$ should begin the algorithm moving again. Of course, whether the algorithm actually proceeds to another minimum after $Q(t)$ is changed depends on the depth and curvature of the local minimum. If the algorithm had stopped at the global minimum of I, $||\delta p^{(i)}||$ should remain small even after $Q(t)$ has been changed. In the example we illustrate this procedure.

TREATMENT OF ILL-POSED PARAMETER ESTIMATION PROBLEMS

In this section we briefly consider the case when the problem is ill-posed, that is, when large changes in the parameter values cause only small changes in the objective function, resulting in the $I(p)$ surface being very flat. When this situation occurs, any numerical scheme becomes extremely sensitive to various computational errors, such as truncation and rounding errors.

To remove these computational difficulties, we will employ the following theorem proved by Klinger (*10*).

Theorem

If A is normal, that is $AA^* = A^*A$, where A^* is the conjugate transpose of the matrix A, then for all $\sigma > 0$

$$[A + \sigma(A^*)^{-1}]\, x = b \tag{39}$$

is better conditioned than $Ax = b$ in terms of the P-condition number unless $P(A) = 1$ where

$$P(A) = \left| \frac{\max_i |\lambda_i|}{\min_i |\lambda_i|} \right| \tag{40}$$

and λ_i are the eigenvalues of A. Applying this theorem to Equation (21), we obtain

$$\delta p^{(i)} = (K^{(i)}(T)^T K^{(i)}(T) + \sigma I)^{-1} K^{(i)}(T)^T \cdot$$

$$\int_0^T \theta^{(i)}(t)^T Q(t)\,[y(t) - h(t, x^{(i)})]\, dt \tag{41}$$

where $K^{(i)}(T)$ is always normal. When the problem is ill-posed $K^{(i)}(T)$ becomes nearly singular. (Note that K cannot be singular because of the observability assumption.) The use of Equation (41) instead of Equation (21) can remove the inaccuracy in computing $\delta p^{(i)}$ due to computational errors as a result of K being almost singular. If we choose a large value for σ, compared to K^TK, the magnitude of $\delta p^{(i)}$ will be decreased. A more complete treatment of ill-posed linear problems is given by Franklin (*11*).

EXAMPLE

We wish to estimate p for the system

$$\dot{x}_1 = x_2 \tag{42}$$

$$\dot{x}_2 = -px_1$$

$$x_1(0) = 0\ ;\ x_2(0) = \pi$$

$$y(t) = x_1(t) + (\text{errors}) \qquad 0 \leq t \leq 1 \tag{43}$$

We assume that the true value of p is $\pi^2 = 9.8696044$.

The curve of $I = \int_0^1 (y(t) - x_1(t))^2\,dt$ versus p is shown in Figure 1 for error-free $y(t)$. There is a local minimum at about $p = 117$ and, in fact, other local minima not shown at larger values of p.

First, using Equation (21) with $Q(t) = 1$ and $\epsilon = 1$, the iterations converged to the local minimum at $p = 116.1$ from initial guesses of 58 and 100. This case was run with error-free observations. Its purpose was to confirm the inability of an algorithm based on Equation (21) with no change of $Q(t)$ to avoid a local minimum.

Next, still considering error-free measurements, the new algorithm was used with $Q(t) = \delta(t - 0.1)$. Thus, only the single measurement at $t = 0.1$ was used as the performance index, representing a severe test of the method. The progress of the iterations are shown below.

Iteration Number	$p^{(i)}$
0	220
1	157.41
2	114.43
19	9.870083
20	9.869604

Now we proceed to the case in which the measurements y(t) are corrupted with error. The error was produced by

$$y(t) = (1 + 0.2G(0, 1))\ \sin\ \pi t \qquad (44)$$

where $G(0, 1)$ is a normally distributed random variable with mean zero and standard deviation one. We already know that with $Q(t) = 1$, the iterations are likely to converge to the local minimum for initial guesses higher than 58. Thus, in order to test the effect of choice of $Q(t)$ on convergence to the local minimum, we chose

$$Q(t) = \begin{cases} 10 & t \leq 0.3 \\ 0.1 & t > 0.3 \end{cases} \qquad (45)$$

The course of the iterations for $p^{(0)} = 220$ and 33 were

Iteration Number	$p^{(i)}$	$p^{(i)}$ (error-free)
0	220	33
1	111.53	30.59
2	74.75	26.36
13	9.75854	9.869648
14	9.75851	9.869604

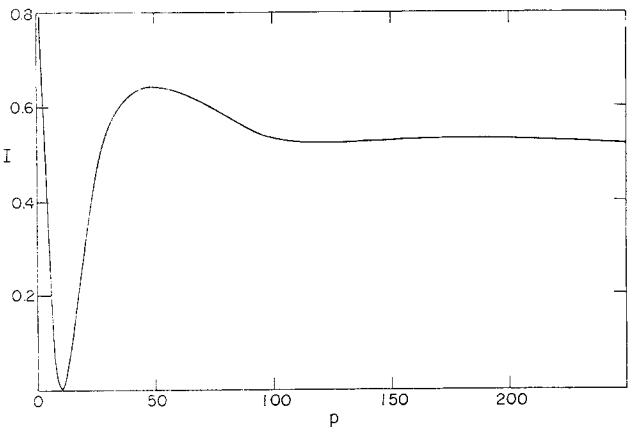

Fig. 1. Least square objective function for example.

SUMMARY

A new algorithm for estimation of parameters in ordinary differential equations has been presented. The algorithm has the following features:

1. For continuous measurements in time, the solution of a two-point boundary value problem is avoided.
2. For discrete measurements fewer differential equations than in the common quasilinearization algorithm need be integrated.
3. Ill-posedness can be alleviated.
4. In searching for the global minimum of the objective function, a simple change in the weighting matrix $Q(t)$ may be faster than trying a number of differential initial parameter guesses.

ACKNOWLEDGMENT

This work was supported by National Science Foundation Grant GK-10136.

NOTATION

A = arbitrary matrix
b = arbitrary constant vector
D = $(n \times l)$ sensitivity matrix
f = n-dimensional vector function
g = $(n + l)$-dimensional vector function
h = m-dimensional vector function
I = identity matrix or performance index
K = $(l \times l)$-observability matrix
p = l-dimensional constant parameter
$P(A)$ = conditional number of matrix A
Q = $(m \times m)$ weighting matrix
s = number of measurement points
t, T = time variables
x = n-dimensional state vector
y = m-dimensional observation vector
z = $(n + l)$-dimensional vector

Greek Letters

α = $(n + l)$-dimensional constant vector
ϵ = constant
η = constant
θ = $(m \times l)$ matrix
λ = eigenvalues of matrix A
Λ = $(n + l) \times (n + l)$ transition matrix
σ = constant
Φ = $(n \times n)$ transition matrix
ψ = $(n + l)$ dimensional vector

LITERATURE CITED

1. Seinfeld, J. H., *Ind. Eng. Chem.*, **1**, 32 (1969).
2. Nieman, R. E., D. G. Fisher, and D. E. Seborg, *Int. J. Control*, **13**, 2, 209 (1971).
3. Lee, E. S., "Quasilinearization and Invariant Imbedding," Academic Press, New York (1968).
4. Leondes, C. T., G. Paine, *J. Optimization Theory Applic.*, **5**, 310 (1968).
5. *Ibid.*, **6**, 395 (1968).
6. Hestenes, M. R., *ibid.*, **5**, 303 (1969).
7. ———., and T. Guinn, *ibid.*, **2**, 87 (1968).
8. Hwang, M., and J. H. Seinfeld, *ibid.*, in press.
9. Franklin, J. N., "Matrix Theory," Prentice-Hall, Englewood Cliffs, N. J. (1968).
10. Klinger, A., *J. Optimization Theory Applic.* **2**, 117 (1968).
11. Franklin, J. N., *J. Math. Anal. Applic.*, **31**, 682 (1970).

Manuscript received March 29, 1971; revision received May 24, 1971, paper accepted June 3, 1971.