PARAMETER IDENTIFICATION FOR AN ABSTRACT CAUCHY PROBLEM BY QUASILINEARIZATION

By

DENNIS W. BREWER (*University of Arkansas, Fayetteville, Arkansas*)

AND

JOHN A. BURNS AND EUGENE M. CLIFF

Virginia Polytechnic Institute and State University, Blacksburg, Virginia

Abstract. A parameter identification problem is considered in the context of a linear abstract Cauchy problem with a parameter-dependent evolution operator. Conditions are investigated under which the gradient of the state with respect to a parameter possesses smoothness properties which lead to local convergence of an estimation algorithm based on quasilinearization. Numerical results are presented concerning estimation of unknown parameters in delay-differential equations.

1. Introduction. During the past fifteen years considerable effort has been devoted to the problem of estimating unknown parameters in distributed parameter systems. The recent book by Banks and Kunisch [9] provides an excellent account of the progress made in the field. Many parameter estimation problems are best formulated as optimization problems (sometimes over infinite-dimensional "parameter spaces"), and algorithms are developed to minimize an appropriate cost function. Although there are several approaches to these problems, their infinite-dimensional nature requires that numerical approximations be introduced at some point in the analysis. Consequently, there are two basic classes of algorithms for optimization-based parameter estimation. The first type of algorithm, and the most frequently used for dynamic problems, is indirect and proceeds by initially approximating the dynamic equations (e.g., finite elements, finite differences, etc.) and then using optimization algorithms on the finite-dimensional problem. This approach is typified by the papers [1–6, 8, 10, 18].

The second more direct approach is based on the direct application of an optimization algorithm and employing numerical approximations at each step of the algorithm to compute the necessary solutions of the dynamic equations. This approach is used in [12, 13, 17, 19]. Both methods have advantages and disadvantages. Depending on the particular type of distributed parameter system, one method may out-perform the other.

Although we shall consider only the problem of identifying a finite number of parameters, the infinite-dimensional dynamic constraint enters into the optimization algorithm. Basically, the objective function from parameter space to $\mathbb{R}$ is a composition of a finite rank map with an operator (defining the dynamic constraint) on an infinite-dimensional space. Therefore, any method that requires gradients to be computed will have to deal with the differentiation of the infinite-dimensional constraint, i.e., the chain rule is needed. It is in this sense that the quasilinearization algorithm considered here has an "infinite-dimensional" nature.

Direct methods such as quasilinearization considered here are often limited by the fact that the dependence on unknown parameters of the solution to the infinite-dimensional dynamical equations may not be "smooth enough" to establish convergence of the algorithm. Indeed, some algorithms may not be properly defined without this necessary smoothness. Indirect methods avoid this difficulty and often lead to easily implemented algorithms. On the other hand, when direct methods can be applied it is sometimes possible to establish the convergence and the rates of convergence to the unknown optimal parameters (see [13, 19]).

This paper considers the dependence on an unknown parameter q of the solution of the linear abstract Cauchy problem

$$\begin{cases} \dot{x}(t) = A(q)x(t) + u(t), & 0 \le t \le T, \\ x(0) = x_0. \end{cases} \tag{1.1}$$

Our ultimate goal is to formulate and establish the convergence of a gradient-based parameter estimation algorithm applicable in this abstract setting.

This algorithm employs computation of the gradient $D_q x(t; q)$ of the solution of problem (1.1) with respect to the parameter. Conditions for the existence of this gradient are established in [11]. In Sec. 2 we review these conditions and the general setting for the remainder of the paper. Convergence of the algorithm requires certain smoothness properties of the gradient $D_q x(t; q)$ with respect to q. These properties are established in Sec. 3 and their applicability to a linear delay-differential equation is discussed in Sec. 4. In this example the delay is among the parameters so that in this setting the parameter dependence appears in unbounded terms of the evolution operator $A(q)$.

An abstract parameter estimation algorithm for a finite-dimensional parameter space using a discrete cost function is presented in Sec. 5. In Sec. 6 its convergence is established using the results of Sec. 3. In Sec. 7 we present several numerical examples which indicate the performance of the algorithm for delay and coefficient estimation in linear delay-differential equations. Additional examples may be found in [12]. Numerical testing and evaluation on a wider variety of parameter estimation problems will be undertaken in a subsequent paper.

2. The general setting. The application of quasilinearization to parameter estimation requires knowledge of the derivative of the state with respect to the unknown parameter. This topic is addressed in [11]. In this section we review the framework used there to obtain differentiability and establish notation to be used in the remainder of this paper.

Let P be an open subset of a normed linear space $\mathbb{P}$ with norm $|\cdot|$, and let X be a Banach space with norm $\|\cdot\|$. For every $q \in P$ let $A(q)$ be a linear operator on $D(A(q))$ in X. Throughout this paper we assume

(H1) $A(q)$ generates a strongly continuous semigroup $S(t; q)$ on X;

(H2) $D(A(q)) = D$ is independent of q;

(H3) $\|S(t; q)x\| \leq Me^{\omega t}\|x\|$, $x \in X$, $t \geq 0$, $q \in D$, for some constants M and ω independent of q, x, and t.

Fix $T > 0$ and $u \in L^1(0, T; X)$. Define $Q(t; q) = \int_0^t S(t - s; q)u(s)\,ds$ for $q \in P$, $0 \leq t \leq T$. Note that if Eq. (1.1) has a strong solution then it is given by the formula $x(t) = S(t; q)x_0 + Q(t; q)$ for $0 \leq t \leq T$.

In applications of this theory, it is useful to consider just those terms of $A(q)$ in which the parameter appears. To this end we write $A(q) = A + B(q)$ where A and $B(q)$ both have domain D and A is independent of q. Concerning $B(q)$ we assume

(H4)   for every $q, q_0 \in P$, there is a constant K such that

$$\int_0^T \|B(q)S(t; q_0)x\|\,dt \leq K\|x\| \quad \text{for all } x \in D.$$

In Sec. 4 we discuss an example in which an unbounded operator $B(q)$ satisfies (H4). This hypothesis does imply, however, that the linear mapping $x \to B(q)S(\cdot; q_0)x$ is bounded as a mapping from D into $L^1(0, T; X)$. Let $F(q, q_0)$ denote the bounded linear extension of this operator to X. Let $\|\cdot\|_1$ denote the norm in $L^1(0, T; X)$. Concerning F we assume

(H5)   there is closed subspace Y of X such that

(i) $F(q, q_0)x_0 \in L^1(0, T; Y)$ for $q, q_0 \in P$, and

(ii) for every $q_0 \in P$ and $\varepsilon > 0$, there exists $\delta > 0$ such that $\|F(q, q_0)y - F(q_0, q_0)y\|_1 \leq \varepsilon\|y\|$ for $y \in Y$ and $|q - q_0| \leq \delta$.

The analogue of F for the function $Q(t; q)$ is the mapping $G(q, q_0)$ from $L^1(0, T; D)$ into $L^1(0, T; X)$ defined by

$$[G(q, q_0)w](t) = \int_0^t B(q)S(t - s; q_0)w(s)\,ds.$$

By (H4) it follows that G can be extended to a bounded linear mapping on $L^1(0, T; X)$, so that, in particular, $G(q, q_0)u$ is defined as an element of $L^1(0, T; X)$. In addition we assume

(H6) $G(q, q_0)u \in L^1(0, T; Y)$ for $q, q_0 \in P$ where Y denotes the subspace required by (H5).

3. Parameter dependence. In this section we deduce smoothness properties of the solution $x(t; q) = S(t; q)x_0 + Q(t; q)$ with respect to q. These properties

are derived from similar properties of $F(q, q_0)$ and $G(q, q_0)$ which are operators related to $A(q)$. These results will be used in Sec. 5 to prove convergence of the parameter estimation algorithm. Throughout this section, $T > 0$, $x_0 \in X$, and $u \in L^1(0, T; X)$ are fixed as given in Eq. (1.1). The symbol D_q denotes Fréchet differentiation with respect to q. These results are given as a series of lemmas whose proofs are at the end of this section.

LEMMA 3.1. Suppose (H1)–(H5) hold. In addition, suppose that for a given $q^* \in P$

(H7) $F(q, q_0)x_0$ is Fréchet differentiable with respect to q at q_0 for every $q_0 \in P$.

For brevity, let $DF(q_0)$ denote $D_q[F(q, q_0)x_0]|_{q=q_0}$ for $q_0 \in P$. In addition, suppose

(H8) $DF(q)$ is strongly continuous in q at q^*, that is, for each $h \in \mathbb{P}$, the mapping $q \to DF(q)h$ from P into $L^1(0, T; X)$ is continuous at q^*.

Then for each $t \in [0, T]$, $S(t; q)x_0$ is Fréchet differentiable with respect to q at every $q \in P$ and $D_q[S(t; q)x_0]$ is strongly continuous with respect to q at q^*.

LEMMA 3.2. Suppose (H1)–(H6) hold and, in addition, suppose that for a given $q^* \in P$

(H9) $G(q, q_0)u$ is Fréchet differentiable with respect to q at q_0 for every $q_0 \in P$.

Again denoting this derivative by $DG(q_0)$ for $q_0 \in P$, assume

(H10) $DG(q)$ is strongly continuous in q at q^*.

Then for $t \in [0, T]$, $Q(t; q)$ is Fréchet differentiable with respect to q at every $q \in P$ and $D_q[Q(t; q)]$ is strongly continuous in q at q^*.

LEMMA 3.3. Suppose (H1)–(H5) and (H7) hold and, in addition, suppose

(H11) $F(q, q^*)$ is locally Lipschitz continuous in q at q^*, uniformly for $y \in Y$, that is, there exist constants $K_1, \delta_1 > 0$ such that

$$\|F(q, q^*)y - F(q^*, q^*)y\|_1 \leq K_1 |q - q^*| \|y\|$$

whenever $|q - q^*| < \delta_1$ and $y \in Y$.

Moreover, assume that

(H12) $DF(q)$ is strongly locally Lipschitz continuous with respect to q at q^*. That is, for each $h \in \mathbb{P}$, there are constants $K, \delta > 0$ such that

$$\|DF(q)h - DF(q^*)h\| \leq K |q - q^*|$$

for $|q - q^*| \leq \delta$.

Then $D_q[S(t; q)x_0]$ is strongly locally Lipschitz continuous with respect to q at q^* for every $t \in [0, T]$.

LEMMA 3.4. Suppose (H1)–(H6), (H9)–(H10) hold and, in addition, suppose

(H13) $DG(q)$ is strongly locally Lipschitz continuous with respect to q at q^*.

Then $D_q[Q(t; q)]$ is strongly locally Lipschitz continuous with respect to q at q^* for every $t \in [0, T]$.

Although the assumptions (H1)–(H13) are rather technical, we shall see that they can easily be verified for delay systems even in the case that the unknown parameter

is the delay itself. Therefore, the results presented here remove the limitations placed on the perturbation $B(q)$ in papers [13, 16].

For completeness we now present the proofs of Lemma 3.1–Lemma 3.4. However, these proofs make use of the basic results found in [11], and in order to keep the length of the proofs reasonable, we assume that the reader has [11] in hand.

Proof of Lemma 3.1. It is shown in [11] that (H1)–(H5), (H7) imply that $D_q[S(t;q)x_0]$ exists for $q \in P$. Furthermore, it is given by the formula

$$D_q[S(t;q)x_0]h = \int_0^t S(t-s;q)[DF(q)h](s)\,ds, \qquad h \in \mathbb{P}. \tag{3.1}$$

Therefore, by substitution, we obtain

$$D_q[S(t;q)x_0]h - D_q[S(t;q^*)x_0]h \tag{3.2}$$

$$= \int_0^t [S(t-s;q) - S(t-s;q^*)]([DF(q)h](s))\,ds$$

$$+ \int_0^t S(t-s;q^*)([DF(q)h](s) - [DF(q^*)h](s))\,ds.$$

Let $\varepsilon > 0$ be given and let $C = Me^{\omega t}$. It can be shown (see [11, proof of Theorem 1]) that for all $x \in X$

$$\|S(t;q)x - S(t;q^*)x\| \le C\|F(q,q^*)x - F(q^*,q^*)x\|_1. \tag{3.3}$$

Combining Eq. (3.3) with (H5)(ii) shows that for some $\delta_1 > 0$

$$\|S(t,q)y - S(t;q^*)y\| \le \varepsilon C\|y\|, \qquad 0 \le t \le T, \ y \in Y,$$

whenever $|q - q^*| \le \delta_1$. In particular, putting $y = [DF(q)h](s) \in Y$ by (H5)(i) we obtain

$$\|[S(t-s;q) - S(t-s;q^*)][DF(q)h](s)\| \le \varepsilon C\|[DF(q)h](s)\|$$

for $|q - q^*| \le \delta_1$, a.e. $s \in (0, T)$. Since $DF(q)h$ is continuous at q^*, there exist constants $K_2, \delta_2 > 0$ such that

$$\|DF(q)h\|_1 \le K_1 \quad \text{for } |q - q^*| \le \delta_2.$$

Combining these estimates shows that the first term in Eq. (3.2) is bounded by $\varepsilon C K_2$ if $|q - q^*| \le \min(\delta_1, \delta_2)$.

Using (H8) it is easy to see that there exists $\delta_3 > 0$ such that the second term in Eq. (3.2) is bounded by εC for $|q - q^*| \le \delta_3$. These estimates complete the proof of Lemma 3.1.

Proof of Lemma 3.2. By Theorem 3 of [11], $D_q[Q(t;q)]$ exists for $q \in P$ and

$$D_q[Q(t;q)] - D_q[Q(t;q^*)] = \int_0^t [S(t-s;q) - S(t-s;q^*)][DG(q)(s)]\,ds$$

$$+ \int_0^t S(t-s;q^*)[(DG(q))(s) - (DG(q^*))(s)]\,ds \tag{3.4}$$

where u has been suppressed in the notation. Since $DG(q) \in L^1(0, T; Y)$ for $q \in P$ by (H6), the proof follows exactly as in the proof of Lemma 3.1.

Proof of Lemma 3.3. Let $\varepsilon > 0$ be given. By Eq. (3.3) and (H11) there exists $\delta_1 > 0$ such that

$$\|S(t;q)y - S(t;q^*)y\| \leq CK_1 \|y\| \, |q - q^*|$$

for $y \in Y$ and $|q - q^*| \leq \delta_1$. Since $DF(q)h \in L^1(0, T; Y)$ by (H5)(i), we have, as in the proof of Lemma 2.1, that the first term of Eq. (3.2) is bounded by $K_1 K_2 |q - q^*|$ for $|q - q^*| \leq \min(\delta_1, \delta_2)$. An estimate of the same form is obtained easily for the second term of Eq. (3.2) using (H12). These estimates complete the proof of Lemma 3.3.

Proof of Lemma 3.4. Since $DG(q)u \in L^1(0, T; Y)$ by (H6), the proof follows exactly as in the proof of Lemma 3.3 using Eq. (3.4) in place of Eq. (3.2).

4. Application to delay-differential equation. In this section we apply the framework of the previous sections to the linear delay-differential equation

$$\begin{cases} \dot{x}(t) = a_0 x(t) + \sum_{k=1}^{n} a_k x(t - q_k) + u(t), \\ x(0) = \eta, \\ x_0 = \varphi. \end{cases} \tag{4.1}$$

Let $\mathbb{P} = \mathbb{R}^n$; fix $r > 0$; and let $P = \{q = (q_1, q_2, \ldots, q_n) : 0 < q_k < r$ for $k = 1, 2, \ldots, n\}$. In Eq. (4.1), $\eta \in \mathbb{R}$, $a_k \in \mathbb{R}$, $k = 0, 1, \ldots, n$, $\varphi \in L^1(-r, 0)$ with norm denoted by $\|\varphi\|_1$, $u \in L^1(0, T)$, and $x_t(s) = x(t + s)$ for $t \geq 0$, $-r \leq s \leq 0$. By a solution of Eq. (4.1) we mean a function x that is absolutely continuous on $[0, T]$ and satisfies Eq. (4.1) almost everywhere on $(0, T)$.

Following the construction in [14], we take $X = \mathbb{R} \times L^1(-r, 0)$ with norm $\|(\eta, \varphi)\| = |\eta| + \|\varphi\|_1$ and define for $q \in P$ an operator $A(q)$ on

$$D = \{(\eta, \varphi) \in X : \varphi \text{ is abs. cont. on } [-r, 0], \ \dot{\varphi} \in L^1(-r, 0), \ \text{and} \ \varphi(0) = \eta\}$$

by

$$A(q)(\eta, \varphi) = \left(a_0 \varphi(0) + \sum_{k=1}^{n} a_k \varphi(-q_k), \dot{\varphi} \right).$$

Then it is well known that $A(q)$ generates a strongly continuous semigroup $S(t;q)$ on X satisfying $S(t;q) = (y(t), y_t)$ where $y(t) = y(t;q)$ denotes the solution of Eq. (4.1) with $u = 0$. It is a consequence of standard results that (H1)–(H3) hold in this setting.

For $q = (q_1, \ldots, q_n)$ and q_0 in P, $(\eta, \varphi) \in X$, and $w \in L^1(0, T)$ it follows that in this example the mappings F and G of Sec. 3 are given by

$$F(q, q_0)(\eta, \varphi) = \left(\sum_{k=1}^{n} a_k y(\cdot - q_k; q_0), 0 \right) \tag{4.2}$$

and

$$[G(q, q_0)w](t) = \left(\sum_{k=1}^{n} a_k z(t - q_k; q_0), 0 \right) \tag{4.3}$$

for a.e. $t \in (0, T)$ where $z(t; q)$ denotes the solution of Eq. (4.1) with $u = w$ and $(\eta, \varphi) = (0, 0)$. It is shown in [11] that these mappings satisfy (H4)–(H6) with the closed subspace $Y = \mathbb{R} \times \{0\}$. It is also shown in [11] that F and G satisfy the differentiability hypotheses (H7) and (H9) for $(\eta, \varphi) = x_0 \in D$ and $q, q_0 \in P$. Furthermore, their Fréchet derivatives are given by

$$[DF(q)h](t) = \left(-\sum_{k=1}^{n} a_k \dot{y}(t - q_k; q)h_k, 0 \right) \tag{4.4}$$

and

$$[DG(q)h](t) = \left(-\sum_{k=1}^{n} a_k \dot{z}(t - q_k; q)h_k, 0 \right) \tag{4.5}$$

for $q \in P$, $h = (h_1, \ldots, h_n) \in \mathbb{R}^n$, where $y(t; q)$ is the solution of Eq. (4.1) with $u = 0$ and $z(t; q)$ is the solution of Eq. (4.1) with $(\eta, \varphi) = (0, 0)$.

It remains to establish conditions under which (H8), (H10)–(H13) are satisfied.

LEMMA 4.1. Fix $q^* = (q_1^*, \ldots, q_n^*) \in P$ and $x_0 \in D$. Then $F(q, q^*)x_0$ as defined by Eq. (4.2) satisfies (H11).

Proof. In Sec. 5 of [11] it is shown that there are constants C_2 and $\delta_2 > 0$ such that

$$\|F(q^* + h, q^*)(\eta, 0) - F(q^*, q^*)(\eta, 0)\|_1 \leq C_2|h| \, \|(\eta, 0)\|$$

for $h \in \mathbb{R}^n$, $\eta \in \mathbb{R}$, $|h| \leq \delta_2$. Here we define $|h| = \sum_{k=1}^{n} |h_k|$. This estimate is equivalent to (H11) with $Y = \mathbb{R} \times \{0\}$.

LEMMA 4.2. Suppose $x_0 = (\eta, \varphi) \in D$. Then $DF(q)$ as given by Eq. (4.4) satisfies (H8). Moreover, if, in addition, $\dot{\varphi}$ is of bounded variation on $[-r, 0]$, then $DF(q)$ satisfies (H12).

Proof. Let $A_m = \max_k |a_k|$ and $|h| = \max_k |h_k|$. Then we obtain the estimate

$$\|DF(q)h - DF(q^*)h\|_1 \leq A_m|h| \sum_{k=1}^{n} \int_0^T |\dot{y}(t - q_k; q) - \dot{y}(t - q_k; q^*)| \, dt \tag{4.6}$$

$$+ A_m|h| \sum_{k=1}^{n} \int_0^T |\dot{y}(t - q_k; q^*) - \dot{y}(t - q_k^*; q^*)| \, dt.$$

Now from Eq. (4.1) we obtain

$$\int_0^T |\dot{y}(t - q_k; q) - \dot{y}(t - q_k; q^*)|\, dt \le \int_0^T |\dot{y}(t; q) - \dot{y}(t; q^*)|\, dt$$

$$\le A_m \sum_{j=1}^n \int_0^T |y(t - q_j; q) - y(t - q_j^*; q^*)|\, dt$$

$$\le A_m \sum_{j=1}^n \int_0^T |y(t - q_j; q) - y(t - q_j; q^*)|\, dt$$

$$+ A_m \sum_{j=1}^n \int_0^T |y(t - q_j; q^*) - y(t - q_j^*; q^*)|\, dt$$

$$\le A_m \sum_{j=1}^n \int_0^T |y(t; q) - y(t; q^*)|\, dt$$

$$+ A_m \sum_{j=1}^n \int_0^T |y(t - q_j; q^*) - y(t - q_j^*; q^*)|\, dt.$$

Now since $y(t; q) = S(t; q)x_0$ is differentiable with respect to q, it is not difficult to show that there are constants β and δ such that

$$\int_0^T |y(t; q) - y(t, q^*)|\, dt \le \beta |q - q^*| \tag{4.8}$$

whenever $|q - q^*| \le \delta$. Combining Eqs. (4.7) and (4.8) with Eq. (4.6) yields

$$\|DF(q)h - DF(q^*)h\|_1 \le A_m^2 |h| n \beta |q - q^*| \tag{4.9}$$

$$+ A_m^2 |h| n \sum_{k=1}^n \int_0^T |y(t - q_k; q^*) - y(t - q_k^*; q^*)|\, dt$$

$$+ A_m |h| \sum_{k=1}^n \int_0^T |\dot{y}(t - q_k; q^*) - \dot{y}(t - q_k^*; q^*)|\, dt.$$

Since $(\eta, \varphi) \in D$, we have y and $\dot{y}$ in $L^1(-r, T)$. Therefore, the integral terms in Eq. (4.9) approach zero as $q \to q^*$ and (H8) holds. If $\dot{\varphi}$ is of bounded variation on $[-r, 0]$, then y and $\dot{y}$ are of bounded variation on $[-r, T]$. By [15, Theorem 2.1.7(b)], this implies that the integral terms in Eq. (4.9) are $O(|q - q^*|)$ as $q \to q^*$ so that (H12) holds.

LEMMA 4.3. Suppose $u \in L^1(0, T)$. Then $DG(q)$ as defined by Eq. (4.5) satisfies (H10). Moreover, if, in addition, u is of bounded variation on $[0, T]$, then $DG(q)$ satisfies (H13).

   Proof. Using Eq. (4.5) in place of Eq. (4.4), one obtains the estimate (4.9) with y replaced by z. Now if $u \in L^1(0, T)$ then z and $\dot{z}$ are in $L^1(-r, T)$, so that (H10) holds. Similarly, if u is of bounded variation on $[0, T]$, then z and $\dot{z}$ are of bounded variation on $[-r, T]$, so that (H13) is satisfied.

5. The algorithm. In this section we define an estimation algorithm over a finite-dimensional parameter space based on quasilinearization and establish local convergence using the results of Sec. 3. In particular, we assume that the parameter space $\mathbb{P}$ is $\mathbb{R}^n$ with canonical basis e_i, $i = 1, 2, \ldots, n$. This algorithm can also be cast in a separable Hilbert space as in [17].

Given $x_0 \in D$ and $q \in P \subset \mathbb{R}^n$, a strong solution of Eq. (1.1) is given by $S(t; q)x_0 + Q(t; q)$. Here we have used the notation of Sec. 2. Let C be a bounded linear mapping from X into $\mathbb{R}^l$ and define

$$\gamma(t; q) = C[S(t; q)x_0 + Q(t; q)].$$

The parameter estimation algorithm is related to the following optimization problem.

PROBLEM 5.1. Let $\bar{y}_j \in \mathbb{R}^l$, $j = 1, 2, \ldots, m$, be data values taken at times $t_j \in [0, T]$, $j = 1, 2, \ldots, m$, respectively. For $q \in P$ define the quadratic cost function

$$J(q) = \sum_{j=1}^{m} |\gamma(t_j; q) - \bar{y}_j|^2.$$

Find $q^* \in P$ such that $J(q^*) \leq J(q)$ for all $q \in P$.

The quasilinearization method defines a recursive algorithm whose fixed point is a local solution of Problem 5.1. A more complete exposition is given in [7]. Given an initial guess $q_0 \in P$, define

$$q_{k+1} = f(q_k), \qquad k = 0, 1, 2, 3, \ldots,$$

where

$$f(q) = q - [D(q)]^{-1}b(q),$$

$$D(q) = \sum_{j=1}^{m} M^{\mathrm{T}}(t_j; q)M(t_j; q),$$

$$b(q) = \sum_{j=1}^{m} M^{\mathrm{T}}(t_j; q)[\gamma(t_j; q) - \bar{y}_j],$$

and the matrix $M(t; q)$ has its ith column $M^i(t; q)$ given by

$$M^i(t; q) = C D_q[S(t; q)x_0 + Q(t; q)]e_i, \qquad i = 1, 2, 3, \ldots, n.$$

LEMMA 5.1. Suppose the hypotheses of Lemmas 3.1 and 3.2 are satisfied. Then $M(t_j; q)$ is continuous in q at q^*.

Proof. This is a direct consequence of Lemmas 3.1 and 3.2 and the above definitions.

LEMMA 5.2. Suppose the hypotheses of Lemmas 3.3 and 3.4 are satisfied. Then there exist constants $\alpha, \delta > 0$ such that

$$|M(t_j; q) - M(t_j; q^*)| \leq \alpha|q - q^*|$$

for $|q - q^*| \leq \delta$, $j = 1, 2, \ldots, m$.

Proof. This is a direct consequence of Lemmas 3.3 and 3.4 and the above definitions.

Note that although the smoothness results of the previous sections hold for an infinite-dimensional parameter, the implementation of the solution of Problem 5.1 by this method is limited to finitely many parameters. In fact, a simple rank argument is used in [17] to show that if the number of parameters, n, exceeds the number of data values, ml, then the matrix $D(q)$ is singular. In [17] a pseudoinverse is proposed as a means of solving the underdetermined problem.

We can now prove the following convergence results. These results are typical of quasilinearization methods, and the proofs given here are in the same spirit as those in [7]. As is usual for this method, whether the hypotheses of the following theorems hold for a particular example depends on the nature of the data. We obtain superlinear convergence when there is an exact fit to data (Theorem 5.1) and linear convergence in the presence of error (Theorem 5.2).

THEOREM 5.1. Suppose the hypotheses of Lemmas 3.1 and 3.2 are satisfied. Moreover, assume $[D(q^*)]^{-1}$ exists, $f(q^*) = q^*$, and $J(q^*) = 0$. Then for every $\varepsilon > 0$, there exists $\delta > 0$ such that

$$|f(q) - f(q^*)| \le \varepsilon |q - q^*|$$

for $|q - q^*| \le \delta$. In particular, there is a neighborhood $\mathscr{U}$ of q^* such that $q_k \to q^*$ as $k \to \infty$ whenever $q_0 \in \mathscr{U}$.

Proof. Note that $f(q^*) = q^*$ implies that $b(q^*) = 0$, or

$$\sum_{j=1}^{m} M^{\mathrm{T}}(t_j; q^*)[\gamma(t_j; q^*) - \overline{y}_j] = 0. \tag{5.1}$$

Therefore

$$f(q) - f(q^*) = D(q)^{-1}[D(q)(q - q^*) - b(q)]$$

$$= D(q)^{-1}\left[\sum_{j=1}^{m} M^{\mathrm{T}}(t_j; q)[M(t_j; q)(q - q^*) - (\gamma(t_j; q) - \overline{y}_j)]\right]$$

$$= D(q)^{-1}\sum_{j=1}^{m} M^{\mathrm{T}}(t_j; q)[M(t_j; q) - M(t_j; q^*)](q - q^*)$$

$$- D(q)^{-1}\sum_{j=1}^{m} M^{\mathrm{T}}(t_j; q)[\gamma(t_j; q) - \gamma(t_j; q^*) - M(t_j; q^*)(q - q^*)]$$

$$- D(q)^{-1}\sum_{j=1}^{m} M^{\mathrm{T}}(t_j; q)[\gamma(t_j; q^*) - \overline{y}_j].$$

Therefore, using Eq. (5.1), we have that

$$
f(q) - f(q^*) = D(q)^{-1} \sum_{j=1}^{m} M^{\mathrm{T}}(t_j; q)[M(t_j; q) - M(t_j; q^*)](q - q^*)
$$

$$
- D(q)^{-1} \sum_{j=1}^{m} M^{\mathrm{T}}(t_j; q)[\gamma(t_j; q) - \gamma(t_j; q^*) - M(t_j; q)(q - q^*)]
$$

$$
- D(q)^{-1} \sum_{j=1}^{m} [M^{\mathrm{T}}(t_j; q) - M^{\mathrm{T}}(t_j; q^*)][\gamma(t_j; q^*) - \overline{y}_j].
$$

$$(5.2)$$

Note that $D(q)^{-1}$ exists and is bounded in a neighborhood of q^* since $D(q^*)^{-1}$ exists by assumption and $D(q)^{-1}$ is continuous at q^* by Lemma 5.1.

Let $\varepsilon > 0$ be given. Using Lemma 5.1 it is easy to see that there exist constants $\beta_1, \delta_1 > 0$ such that the first term in Eq. (5.2) is bounded by $\varepsilon\beta_1|q - q^*|$ for $|q - q^*| \le \delta_1$. Furthermore, since $M(t_j; q^*)$ is the Fréchet derivative of $\gamma(t_j; q)$ at q^*, one can show that there exist constants $\beta_2, \delta_2 > 0$ such that the second term of Eq. (5.2) is bounded by $\varepsilon\beta_2|q - q^*|$ for $|q - q^*| \le \delta_2$. Combining these estimates with Eq. (5.2) yields

$$
|f(q) - f(q^*)|
$$
$$
\le \varepsilon\beta|q - q^*| + |D(p)^{-1}| \sum_{j=1}^{m} |M^{\mathrm{T}}(t_j; q) - M^{\mathrm{T}}(t_j; q^*)|\,|\gamma(t_j; q^*) - \overline{y}_j| \tag{5.3}
$$

for $|q - q^*| \le \delta = \min(\delta_1, \delta_2)$ and $\beta = \beta_1 + \beta_2$. Since $J(q^*) = 0$, the last term in Eq. (5.3) is zero. This estimate yields the desired result.

The following theorem does not require an exact fit to data, but does place some technical restrictions on the behaviour of M near q^*. Note that if Lemmas 3.3 and 3.4 hold then there exists $\overline{\delta} > 0$ such that for $0 < \delta < \overline{\delta}$ there exists a constant $K(\delta)$ such that

$$
\sum_{j=1}^{m} |M^{\mathrm{T}}(t_j; q) - M^{\mathrm{T}}(t_j; q^*)| \le K(\delta)|q - q^*|
$$

for $|q - q^*| \le \delta$. Let $K^* = \limsup_{\delta \downarrow 0} K(\delta)$ and define

$$
\lambda^* = K^*|D(q^*)^{-1}| \max |\gamma(t_j; q^*) - \overline{y}_j|. \tag{5.4}
$$

THEOREM 5.2. Suppose the hypotheses of Lemmas 3.3 and 3.4 are satisfied. Moreover, assume $[D(q^*)]^{-1}$ exists and $f(q^*) = q^*$. Let λ^* be defined by Eq. (5.4) and assume $\lambda^* < 1$. Then there exists $\delta^* > 0$ such that

$$
|f(q) - f(q^*)| \le \lambda^*|q - q^*|
$$

for $|q - q^*| \le \delta^*$. In particular, $q_k \to q^*$ as $k \to \infty$ whenever $|q_0 - q^*| \le \delta^*$.

Proof. This estimate is a direct consequence of Eq. (5.3).

6. Numerical examples. In this section we consider several examples in which the above algorithm was used to solve parameter estimation problems in delay-differential

equations. In these examples, the emphasis is on delay identification, since in the abstract setting, this represents an unbounded perturbation of the generator as noted in Sec. 4.

With the exception of Example 6.8, the various unknown parameters are estimated using data generated from closed-form expressions for the solution found by the "method of steps." The algorithm is implemented by an averaging scheme [2] that approximates the state equation and the associated sensitivity equations by a system of ordinary differential equations. This system is solved by a fourth-order Runge-Kutta routine.

In the one-delay examples the averaging scheme is implemented with the delay interval $[-r, 0]$ divided into sixteen equal segments, except that Example 6.8 uses sixty-four equal segments. In the two-delay examples the intervals $[-r2, -r1]$ and $[-r1, 0]$ are divided into sixteen equal segments. All computations were done on a VAX 11/750 minicomputer or a SUN Microsystem at the Institute for Computer Applications in Science and Engineering (ICASE).

EXAMPLE 6.1. This example illustrates the rapid convergence of the method for a single unknown parameter—the delay in the following equation—with an initial guess that is an order of magnitude greater than the "true value" of $r = 1.0$. The equation and the results of the iteration are given below.

$$\begin{cases} \dot{x}(t) = -2x(t) + 3x(t - r), & t \geq 0, \\ x(t) = t + 1, & t \leq 0. \end{cases}$$

iterate	r	error
0	10.000	34.056
1	1.299	0.955
2	0.946	0.175
3	0.989	0.115
4	0.987	0.115

The convergence of the states to ten data points on the interval $[0, 2]$ is illustrated in Figure 1.

EXAMPLE 6.2. The data is the same as for Example 6.1; however, in this case, the algorithm is asked to estimate the coefficients as well as the delay. The equation shows an insensitivity to the individual coefficients, which leads to the inaccuracy in the converged estimates. In fact, because of errors introduced by the averaging scheme for computing the state, the estimated values fit the data better than the "true values" used to compute the data by the method of steps. The "true values" are $a = -2$, $b = 3$, and $r = 1$. The equation and the results of the iteration are

$$\begin{cases} \dot{x}(t) = ax(t) + bx(t - r), & t \geq 0, \\ x(t) = t + 1, & t \leq 0. \end{cases}$$

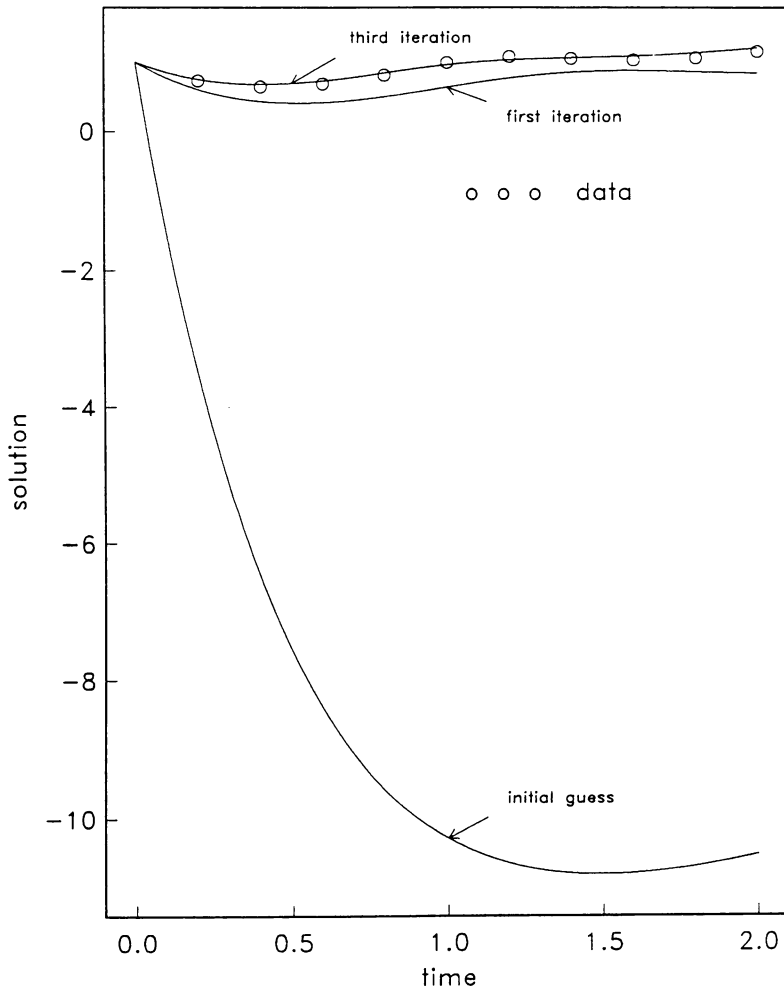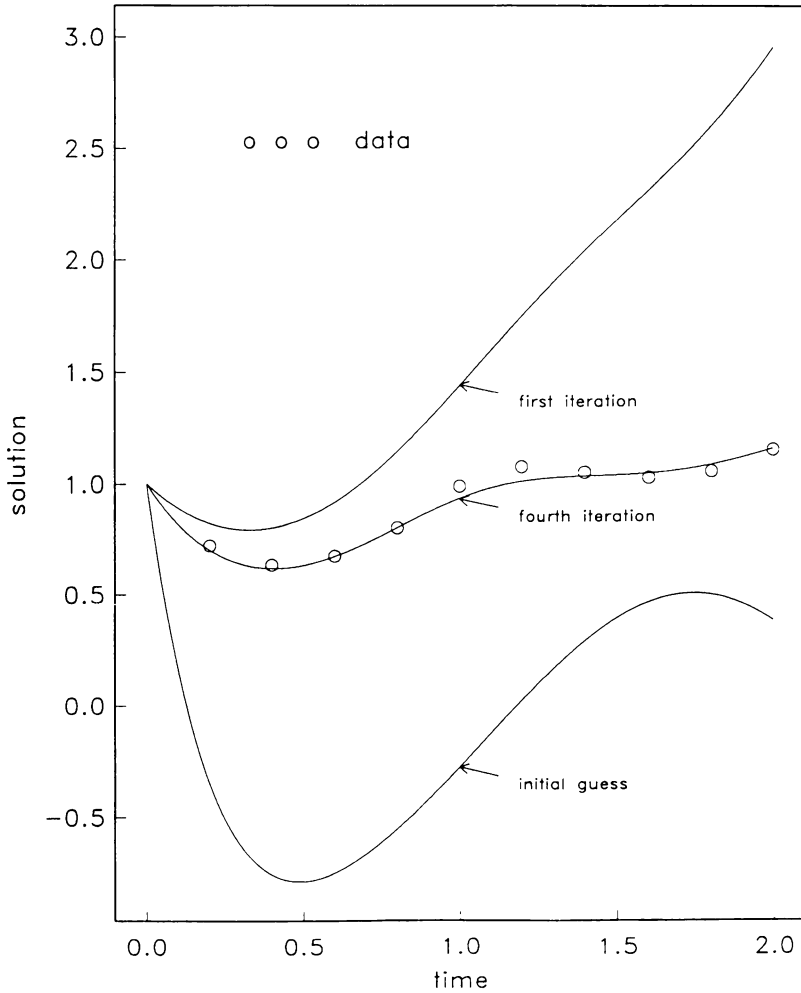| iterate	a	b	r	error
0	−4.000	7.000	2.000	3.379
1	−0.815	3.537	1.184	2.968
2	−1.596	3.342	1.122	0.775
3	−2.403	3.713	1.002	0.188
4	−2.250	3.361	1.015	0.094
5	−2.352	3.483	1.006	0.093

FIG. 1

FIG. 2

The convergence of the states is illustrated in Figure 2.

EXAMPLE 6.3. This case illustrates the effect of a forcing function on the state equation. The nonhomogeneous delay-differential equation

$$\begin{cases} \dot{x}(t) = ax(t) + bx(t - r) + u(t), & t \geq 0, \\ x(t) = t + 1, & t \leq 0, \end{cases}$$

where

$$u(t) = \begin{cases} 0, & t < 0.1, \\ 1, & t \geq 0.1, \end{cases}$$

is solved in closed form by the method of steps with parameter values $a = -2$, $b = 3$, $r = 1$ as in Example 6.2. The results of the parameter estimation algorithm are

iterate	a	b	r	error
0	−4.000	7.000	2.000	4.0527
1	1.022	3.165	1.140	39.2657
2	−2.637	23.652	1.168	24.9577
3	−5.979	28.631	1.141	11.6964
4	−8.034	23.250	1.118	3.5425
5	−5.167	5.417	1.028	2.0471
6	−1.239	4.195	1.008	4.8981
7	−2.861	6.222	1.005	1.8930
8	−2.485	3.795	0.998	0.0819
9	−2.115	3.201	1.013	0.0724
10	−2.247	3.380	0.998	0.0691

The results are similar to those of Example 6.3, except that the solution has become somewhat more sensitive to the coefficients.

EXAMPLE 6.4. This example indicates the ability of the algorithm to estimate two unknown delays. The algorithm converges rapidly from a relatively poor initial guess. The "true values" are $r_1 = 1.0$ and $r_2 = 2.0$. The equation and the results of the parameter estimation algorithm are given below and the convergence of the states to ten data points on the interval $[0, 3]$ is illustrated in Figure 3 (see p. 16).

$$\begin{cases} \dot{x}(t) = -x(t) + x(t - r_1) - x(t - r_2), & t \geq 0, \\ x(t) = t + 1, & t \leq 0. \end{cases}$$

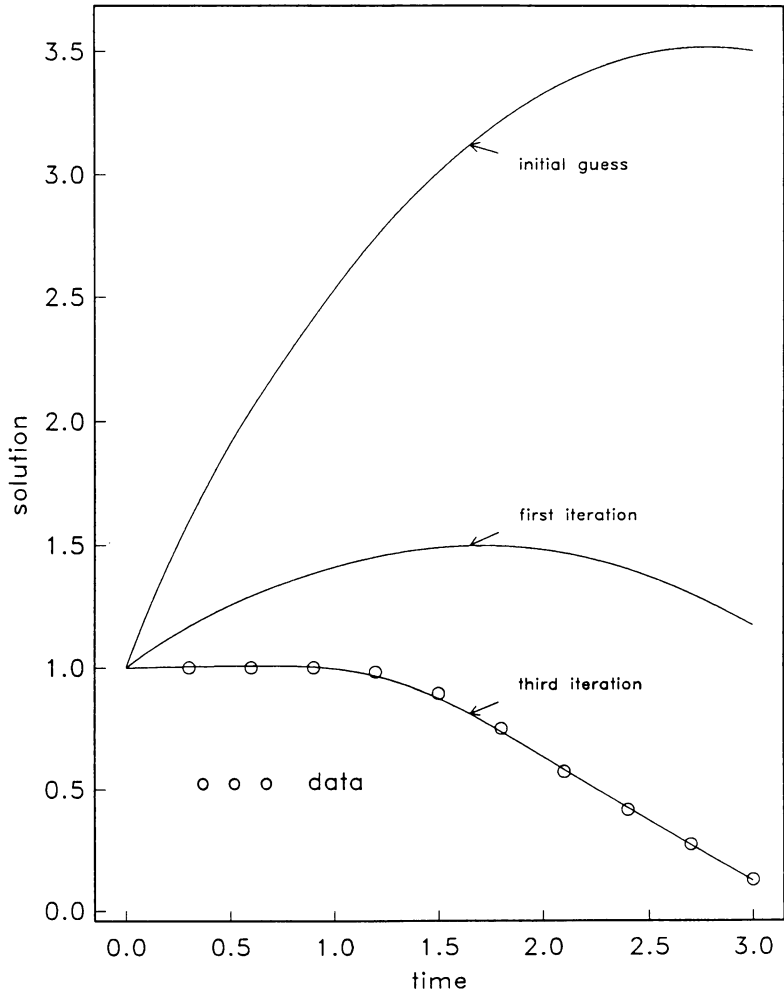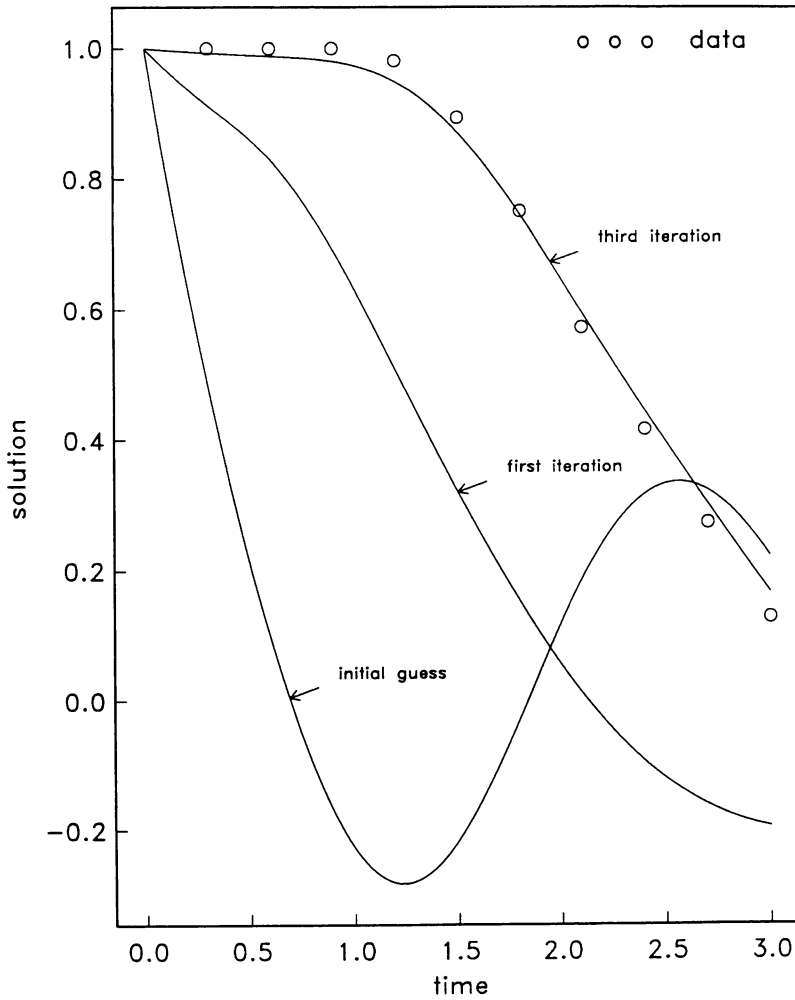| iterate	r_1	r_2	error
0	0.600	4.000	7.500
1	1.569	3.216	2.295
2	1.146	2.100	0.100
3	0.977	1.998	0.034
4	0.978	2.003	0.032

FIG. 3

FIG. 4

EXAMPLE 6.5. The equation and data for this example are the same as in Example 6.4. In this case the initial guess reverses the order of the "true" delay values. The results of this iteration are given below and convergence of the states on the interval [0, 3] is illustrated in Figure 4.

iterate	r_1	r_2	error
0	2.000	1.000	2.460
1	0.483	1.151	1.379
2	1.561	2.014	0.788
3	1.100	2.072	0.077
4	0.980	2.002	0.033

EXAMPLE 6.6. In this case the algorithm is asked to estimate parameters in a delay model of a system with no delay. Ten data points on the interval $[0, 2]$ are computed from the exponential solution of

$$\begin{cases} \dot{x}(t) = -2x(t), \\ x(0) = 1, \end{cases}$$

and the algorithm is asked to estimate unknown parameters in the system

$$\begin{cases} \dot{x}(t) = ax(t) + bx(t - r), & t \geq 0, \\ x(t) = t + 1, & t \leq 0. \end{cases}$$

The first four iterations are

iterate	a	b	r	error
0	−3.000	3.000	2.000	1.2577
1	−3.060	−0.637	1.947	0.2551
2	−1.687	0.235	1.981	0.1144
3	−1.967	0.025	1.985	0.0110
4	−2.000	0.000	1.986	0.0001

On the fifth iteration the algorithm aborted when it was asked to invert a nearly singular matrix. This reflects the fact that at the true parameter values the state is completely insensitive to the delay.

EXAMPLE 6.7. This case is the same as the previous example except that the data is taken from the closed form solution of the nonhomogeneous undelayed equation

$$\begin{cases} \dot{x}(t) = -2x(t) + u(t), \\ x(0) = 1, \end{cases}$$

where u is the same step function as in Example 6.3. The results are similar to those of the previous example.

iterate	a	b	r	error
0	−3.000	3.000	2.000	1.3135
1	−2.848	0.099	1.804	0.5121
2	−1.841	0.138	2.401	0.0811
3	−1.971	0.003	2.508	0.0197

EXAMPLE 6.8. In this example we consider the second-order equation

$$\begin{cases} \dfrac{d^2x}{dt^2}(t) + \omega^2 x(t) + a_0 \dfrac{dx}{dt}(t-r) + a_1 x(t-r) = u(t), & t \geq 0, \\ x(t) = 1, & t < 0, \end{cases}$$

where $u(t)$ is the step function of Example 6.3. This equation models a harmonic oscillator with retarded damping and restoring forces. In [13] a quasilinearization algorithm is used to estimate coefficients in this equation. The methods of this paper allow the delay r to be added to the set of unknown parameters. For this example the averaging method was used to compute "data" values for the parameter estimation algorithm with "true" values of $\omega = 6$, $a_0 = 2.5$, $a_1 = 9$, and $r = 1$. The results of the iterative algorithm are given below and the convergence of the states (displacement and velocity) on the interval $[0, 2]$ is illustrated in Figures 5 and 6 (see pp. 20 and 21).

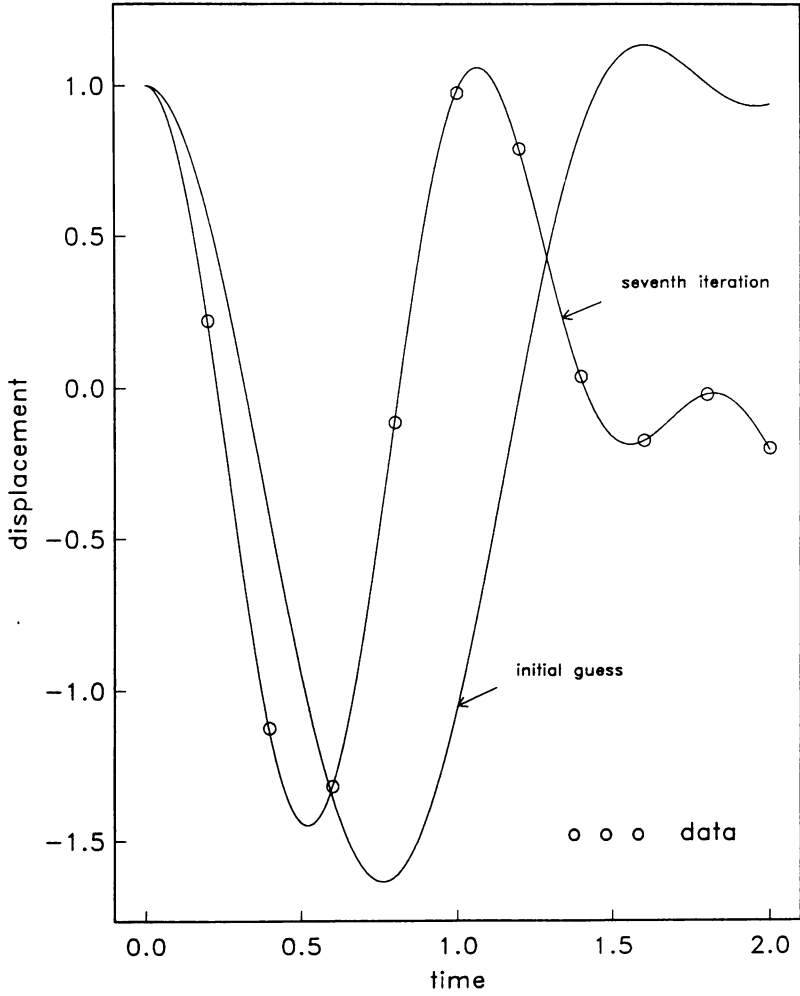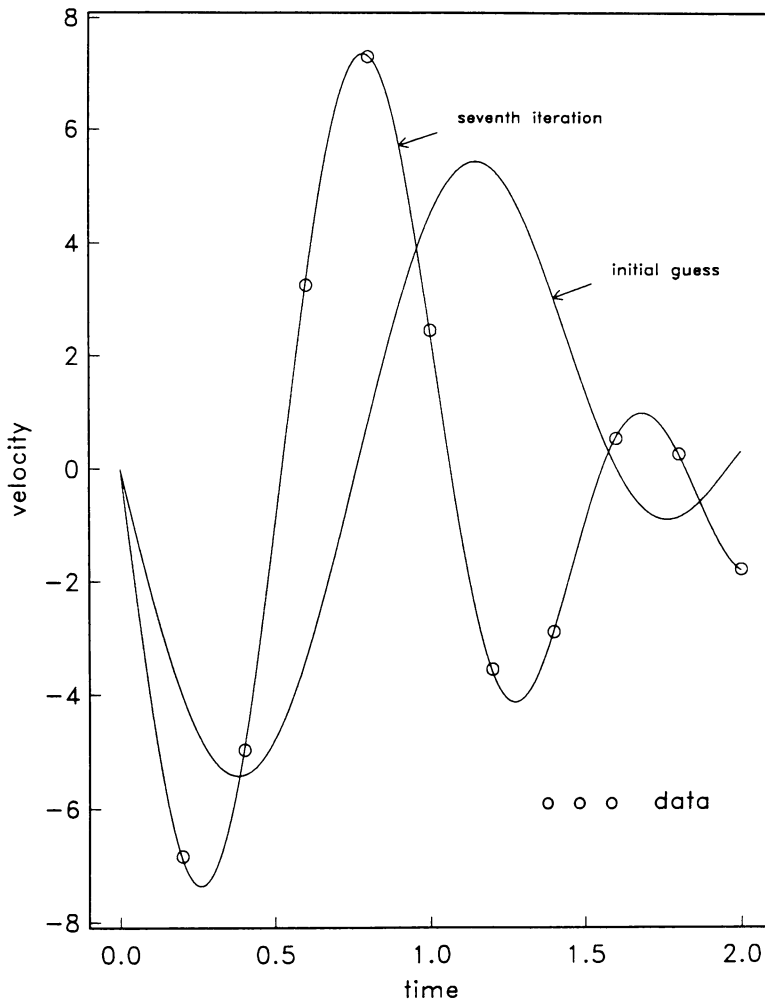| iterate	ω	a_0	a_1	r	error
0	4.100	4.600	6.300	1.500	15.212
1	5.073	6.025	−8.338	0.918	15.181
2	6.705	4.710	−0.682	1.524	12.389
3	6.188	−14.677	−4.838	1.102	31.950
4	5.902	12.347	8.396	1.068	25.234
5	5.964	2.994	8.980	1.061	2.186
6	5.995	2.416	9.016	1.004	0.344
7	6.000	2.503	8.999	1.000	0.007

FIG. 5

FIG. 6

REFERENCES

[1] H. T. Banks, *Identification of nonlinear delay systems using spline methods*, Internat. Conf. on Nonlinear Phenomena in the Mathematical Sciences, Univ. of Texas, Arlington, TX, 1980

[2] H. T. Banks and J. A. Burns, *Hereditary control problems: numerical methods based on averaging approximations*, SIAM J. Control Optim. **16**, 169–208 (1978)

[3] H. T. Banks, J. A. Burns, and E. M. Cliff, *A comparison of numerical methods for identification and optimization problems involving control systems with delays*, LCDS Report No. 79-7, Brown Univ., Providence, RI, 1979

[4] H. T. Banks, J. A. Burns, and E. M. Cliff, *Parameter estimation and identification for systems with delays*, SIAM J. Control Optim. **19**, 791–828 (1981)

[5] H. T. Banks, J. A. Burns, and E. M. Cliff, *Spline-based approximation methods for control and identification of hereditary systems*, Internat. Sympos. on Systems Optimization and Analysis (eds., A. Bensoussan and J. L. Lions), Lecture Notes in Control and Info. Sci., vol. 14, Springer, Heidelberg, 1979, pp. 314–320

[6] H. T. Banks and P. L. Daniel, *Estimation of delays and other parameters in nonlinear functional equations*, SIAM J. Control Optim. **21**, 893–915 (1983)

[7] H. T. Banks and G. M. Groome, Jr., *Convergence theorems for parameter estimation by quasilinearization*, J. Math. Anal. Appl. **42**, 91–109 (1973)

[8] H. T. Banks and F. Kappel, *Spline approximation for functional differential equations*, J. Differential Equations **34**, 496–522 (1979)

[9] H. T. Banks and K. Kunisch, *Estimation techniques for distributed parameter systems*, Birkhäuser, New York, 1989

[10] H. T. Banks and K. Kunisch, *Parameter estimation techniques for nonlinear distributed parameter systems*, Internat. Conf. on Nonlinear Phenomena in the Mathematical Sciences, Univ. of Texas, Arlington, TX, 1980

[11] D. W. Brewer, *The differentiability with respect to a parameter of the solution of a linear abstract Cauchy problem*, SIAM J. Math. Anal. **13**, 607–620 (1982)

[12] D. W. Brewer, *Quasi-Newton methods for parameter estimation in functional differential equations*, Proc. 27th IEEE Conf. on Decision and Control, Austin, TX, 1988, pp. 806–809

[13] J. A. Burns and E. M. Cliff, *An abstract quasi-linearization algorithm for estimating parameters in hereditary systems*, IEEE Trans. Automat. Control **25**, 126–129 (1980)

[14] J. A. Burns, T. L. Herdman, and H. Stech, *Linear functional differential equations as semi-groups on product spaces*, SIAM J. Math. Anal. **14**, 98–116 (1983)

[15] P. L. Butzer and H. Berens, *Semi-groups of operators and approximation*, Springer-Verlag, New York, 1967

[16] J. S. Gibson and L. G. Clark, *Sensitivity analysis of a class of evolution equations*, J. Math. Anal. Appl. **58**, 22–31 (1977)

[17] P. W. Hammer, *Parameter identification in parabolic partial differential equations using quasilinearization*, Ph.D. Thesis, ICAM Report 90-07-01, Virginia Polytechnic Institute and State Univ., Blacksburg, VA, 1990

[18] I. G. Rosen, *Discrete approximation methods for parameter identification in delay systems*, SIAM J. Control Optim. **22**, 95–120 (1984)

[19] E. Sachs, *Convergence rates of quasi-Newton algorithms for some nonsmooth optimization problems*, SIAM J. Control Optim. **23**, 401–418 (1985)