Parameter Identification for Nonlinear Abstract Cauchy Problems using Quasilinearization

TERRY L. HERDMAN [‡,⋄]

PEDRO MORIN [†,⋄]

and

RUBEN D. SPIES [†]

[‡] Interdisciplinary Center for Applied Mathematics
Virginia Polytechnic Institute and State University
Blacksburg VA - 24061 - 0531

[†] Instituto de Desarrollo Tecnológico para la Industria Química
Programa Especial de Matemática Aplicada
CONICET - Universidad Nacional del Litoral
Güemes 3450
3000 Santa Fe - Argentina

Abstract: An approach to quasilinearization for parameter identification in nonlinear abstract Cauchy problems in which the parameter appears in the nonlinear term, is presented. This new approach has two main advantages over the classical one: it is much more intuitive and the derivation of the algorithm is done without need of the sensitivity equations on which classical quasilinearization is based upon. Sufficient conditions for the convergence of the algorithm are derived in terms of the regularity of the solutions with respect to the parameters. A comparison with the standard approach is presented and an application example is shown in which the non-physical parameters in a mathematical model for shape memory alloys are estimated.

1. Introduction

Let Z and $\tilde{\mathcal{Q}}$ be two Banach spaces, A the infinitesimal generator of an analytic semigroup $T(t)$ on Z, D a subset of Z, $\mathcal{Q}$ a subset of $\tilde{\mathcal{Q}}$ and $F : \mathcal{Q} \times [0,T] \times D \to Z$. We shall consider the following nonlinear Cauchy problem in Z:

$$(P)_q \begin{cases} \dot{z}(t) = Az(t) + F(q,t,z(t)), & t \in (0,T) \\ \\ z(0) = z_0. \end{cases}$$

The spaces Z and $\tilde{\mathcal{Q}}$ will be referred to as the state-space and the parameter space, respectively, while $\mathcal{Q}$ will be called the admissible parameter set.

Let Y be a Hilbert space and $\mathcal{C}$ a bounded linear operator from Z into Y, $\mathcal{C} \in \mathcal{L}(Z,Y)$. We shall refer to $\mathcal{C}$ as the "observation operator". Let $\hat{z}_i \in Y$, be "observations" at times t_i, $i = 1, 2, \cdots, m$ of the process described by the IVP $(P)_q$. The *"parameter identification problem"* (ID problem in the sequel) associated to $(P)_q$ and the observations $\{z_i\}_{i=1}^m$ is:

(ID) : find $q \in \mathcal{Q}$ that minimizes the error criterion

$$J(q) \doteq \frac{1}{2} \sum_{i=1}^{m} \|\mathcal{C}z(t_i;q) - \hat{z}_i\|_Y^2 \tag{1.1}$$

where $z(t;q)$ denotes the unique solution of $(P)_q$ in the interval $[0,T]$. In the next section we will provide sufficient conditions for the existence and uniqueness of solutions.

There are two general approaches to ID problems. The first one, frequently used in linear problems, is the so called *indirect approach*. Here, the identification algorithm starts with a finite dimensional

[⋄]This research was supported in part by the Air Force Office of Scientific Research under grants F49620-93-1-0280 and F49620-96-1-0329.

[†]The work of the authors was supported in part by CONICET, Consejo Nacional de Investigaciones Científicas y Técnicas of Argentina and UNL, Universidad Nacional del Litoral through project CAI+D 94-0016-004-023.

approximation of the infinite dimensional problem, after which an optimization algorithm based on these approximations is implemented. The second approach, called the *direct approach*, consists of applying an optimization algorithm to the infinite dimensional problem $(P)_q$ and using finite dimensional approximations when needed to solve the resulting infinite dimensional subproblems. Depending on the problem being considered, one method may be more efficient than the other. Methods based on the indirect approach are usually easier to implement computationally, however in general, they require that the dynamic equations be solved a greater number of times than direct methods do. For this reason, in practical problems the use of indirect methods is mainly restricted to linear problems. Also, for indirect methods, no more than subsequential convergence can be obtained while "full" convergence can be proved for certain direct methods.

The convergence issue in ID problems is very important. Although direct methods usually generate much more efficient algorithms and quite often full convergence can be shown, they have the drawback that they require the solution of the system to be smooth with respect to the parameters. In many cases this smoothness does not exist or it may be difficult to prove.

Identification problems arise often in many physical, geological, chemical and biological systems. It is for that reason that a great amount of attention has been devoted to the study of identification methods for linear and nonlinear distributed parameter systems.

In particular, the quasilinearization approach to ID problems has been studied by several authors for different type of problems. Brewer, Burns and Cliff ([4]) have worked on many identification issues that arise in the study and application of quasilinearization methods for nonhomogeneous linear systems of the type $\dot{z}(t) = A(q)z(t) + u(t)$, where the dependence on the unknown parameter q comes through the linear operator $A(q)$. Later on, Hammer ([6]) applied these tools to nonlinear problems of the type $\dot{z}(t) = A(q)z(t)+f(t,z)$, where $f(t,z)$ is nonlinear in z but it does not depend on the unknown parameter q. Banks and Groome ([2]) considered a quasilinearization approach for ID problems arising in the study of general nonlinear problems of the type $\dot{z}(t) = g(t, z(t), q)$, but their work is valid in finite dimensional state spaces only, i.e., $z(t) \in \mathbb{R}^n$ and it does not extend to the infinite dimensional context. ID problems for systems of the type $(P)_q$ have never been studied previously.

The organization of this article is as follows. In Section 2 the quasilinearization algorithm for parameter identification in an abstract context is derived. In Section 3 sufficient conditions for the convergence the algorithm are given. In Section 4 a comparison is made between the approach presented here and the standard approach to quasilinearization. In Section 5 an application is presented in which the parameters that define the free energy in a model for Shape Memory Alloys are identified.

2. Quasilinearization Algorithm

In this section we will introduce the algorithm, but first we need to recall some properties of analytic semigroups and make some assumptions on the nonlinear part of the equation.

Since A generates an analytic semigroup, $\omega \doteq \sup\{\mathrm{Re}(\lambda) : \lambda \in \sigma(A)\}$ is finite and for any complex λ with $\mathrm{Re}(\lambda) > \omega$, the fractional powers $(\lambda I - A)^\delta$ of $\lambda I - A$ are closed, linear and invertible operators in Z for $\delta \in [0,1]$ (see [8]). From now on, λ will be fixed and $\mathrm{Re}(\lambda) > \omega$, Z_δ shall denote the space $D\left((\lambda I - A)^\delta\right)$ imbedded with the norm of the graph of $(\lambda I - A)^\delta$. Due to the fact that $\mathrm{Re}(\lambda) > \omega$, one has $\lambda \in \rho(A)$ and this norm is equivalent to the norm $\|z\|_\delta \doteq \left\|(\lambda I - A)^\delta z\right\|_Z$.

Consider the following standing hypothesis.

(H1). *There exists $\delta \in (0,1)$ such that $Z_\delta \subset D$ and $F : \mathcal{Q} \times [0,T] \times Z_\delta \to Z$ is locally Lipschitz continuous in t and z, i.e., for any $q \in \mathcal{Q}$ and any bounded subset U of $[0,T] \times Z_\delta$ there exists a constant $L = L(q,U)$ such that*

$$\|F(q,t_1,z_1) - F(q,t_2,z_2)\|_Z \le L\left(|t_1 - t_2| + \|z_1 - z_2\|_\delta\right), \qquad \forall (t_i, z_i) \in U$$

where the constant L can be chosen independent of q on any compact subset $\mathcal{Q}_C$ of $\mathcal{Q}$.

The following theorem follows immediately from Theorem 6.3.1 in [8].

Theorem 1. *Let $q \in \mathcal{Q}$ and $z_0 \in Z_\delta$. If F satisfies (H1), then there exists $t_1 = t_1(q, z_0) > 0$ such that $(P)_q$ has a unique classical solution on $[0, t_1)$. i.e., there exists a function $z(\cdot) \in C^0([0,t_1) : Z_\delta) \cap C^1((0,t_1) : Z)$ such that*

$$\begin{cases} \dot{z}(t) = Az(t) + F(q,t,z(t)), & t \in (0,t_1) \\ z(0) = z_0. \end{cases}$$

The function $z(t)$ satisfies the integral equation

$$z(t) = T(t)z_0 + \int_0^t T(t-s)F(q,s,z(s))\,ds, \qquad \forall t \in [0,t_1).$$

Also, $t_1(q,z_0) > 0$ can be chosen independent of q on compact subsets of $\mathcal{Q}$.

Let us denote by $z(t;q)$ the solution $z(t)$ of $(P)_q$.

Consider now the parameter estimation problem (ID). In order to obtain the algorithm, we assume from now on, that for each fixed $t \in [0,t_1)$ the mapping $q \to z(t;q)$ is Fréchet differentiable. Sufficient conditions on F that guarantee this assumption can be found in [5]. Assume for the time being that there exists a unique minimizer $q^* \in \mathcal{Q}$ of $J(q)$. The following algorithm is proposed.

Step 1: Given an estimate q^k of q^*, approximate $z(t;q)$ by its first order Taylor expansion about q^k, i.e., let $z^{k+1}(t;q) \doteq z(t;q^k) + z_q(t;q^k)\left(q-q^k\right)$ where $z_q(t;q)$ denotes the Fréchet derivative of $z(t;q)$ with respect to q.

Step 2: Define the modified error criterion by

$$\begin{aligned} J^k(q) &\doteq \frac{1}{2}\sum_{i=1}^m \left\|\mathcal{C}z^{k+1}(t_i;q) - \hat{z}_i\right\|_Y^2 \\ &= \frac{1}{2}\sum_{i=1}^m \left\|\mathcal{C}\left[z(t_i;q^k) + z_q(t_i;q^k)\left(q-q^k\right)\right] - \hat{z}_i\right\|_Y^2. \end{aligned}$$

Step 3: Next, define q^{k+1} to be a minimizer of the modified error criterion $J^k(q)$. In order to find q^{k+1}, differentiate $J^k(q)$, set the result equal to zero and solve for q. Finally, call this solution q^{k+1}, replace k with $k+1$ and repeat Step 1.

Observe that, unless $z_q(t_i;q^k) = 0$, for all $i = 1,2,\cdots,m$ the functional $J^k(q)$ is strictly convex and therefore, there exists only one solution of $D_q(J^k(q)) = 0$ and this solution is a minimizer. Also, the condition $D_q(J^k(q)) = 0$ is satisfied if and only if

$$\sum_{i=1}^m \left\langle \mathcal{C}\left[z_q(t_i;q^k)h\right], \mathcal{C}\left[z_q(t_i;q^k)(q-q^k)\right]\right\rangle_Y = -\sum_{i=1}^m \left\langle \mathcal{C}\left[z_q(t_i;q^k)h\right], \mathcal{C}z(y_i;q^k) - \hat{z}_i\right\rangle_Y$$

for every $h \in \tilde{\mathcal{Q}}$.

Assume for the moment that $\tilde{\mathcal{Q}}$ is finite dimensional and $\{g_j : j = 1,2,\cdots,s\}$ is an orthonormal basis of $\tilde{\mathcal{Q}}$. Then, the equation above is equivalent to

$$\sum_{i=1}^m \left\langle \mathcal{C}\left[z_q(t_i;q^k)g_j\right], \mathcal{C}\left[z_q(t_i;q^k)(q-q^k)\right]\right\rangle_Y = -\sum_{i=1}^m \left\langle \mathcal{C}\left[z_q(t_i;q^k)g_j\right], \mathcal{C}z(y_i;q^k) - \hat{z}_i\right\rangle_Y, \tag{2.1}$$

for $j = 1,2,\cdots,s$.

Since $\{g_j\}$ is an orthonormal basis, $q \in \tilde{\mathcal{Q}}$ iff there exists a unique $\alpha = (\alpha_1,\alpha_2,\cdots,\alpha_s) \in \mathbb{R}^s$ such that $q = \sum_{j=1}^s \alpha_j g_j$ and $\|q\|_{\tilde{\mathcal{Q}}} = |\alpha|$. Therefore the parameter identification problem can be reformulated in terms of the coefficients of q as follows. Define $\mathbb{R}^s(\mathcal{Q}) = \{\alpha \in \mathbb{R}^s : q_\alpha = \sum_{j=1}^s \alpha_j g_j \in \mathcal{Q}\}$. Given $\alpha^k \in \mathbb{R}^s(\mathcal{Q})$ $(q^k \in \mathcal{Q})$ determine $\alpha^{k+1} \in \mathbb{R}^s(\mathcal{Q})$ by solving equations 2.1 for q.

More precisely, for each α, let q_α denote the expression $\sum_{j=1}^s \alpha_j g_j$, and

$$D(\alpha)\gamma = \sum_{i=1}^m M(t_i;q_\alpha)^*\left[M(t_i;q_\alpha)\gamma\right], \qquad \gamma \in \mathbb{R}^s,$$

where for each $q \in \tilde{\mathcal{Q}}$, $t \in [0,T]$, $M(t;q) : \mathbb{R}^s \to Y$ is defined by

$$M(t;q)\alpha = \left[\,\mathcal{C}z_q(t;q)g_1 \quad \mathcal{C}z_q(t;q)g_2 \quad \cdots \quad \mathcal{C}z_q(t;q)g_s\,\right]\begin{bmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \alpha_s\end{bmatrix}$$

and $M(t;q)^*: Y \to \mathbb{R}^s$ denotes the adjoint operator of $M(t;q)$.

With this notation α^{k+1} can be computed as

$$\begin{aligned}\alpha^{k+1} &= \alpha^k - \left[D(\alpha^k)\right]^{-1} \sum_{i=1}^{m} M(t_i; q_{\alpha^k})^* \left[\mathcal{C} z(t_i; q_{\alpha^k}) - \hat{z}_i\right] \\ &\doteq E\left(\alpha^k\right)\end{aligned}$$

whenever $\left[D(\alpha^k)\right]^{-1}$ exists.

3. Convergence of the quasilinearization algorithm

In this section we shall deal with the convergence of the algorithm which was introduced in the previous section. Two results will be presented giving sufficient conditions for the algorithm to converge. The following two preliminary lemmas will be needed.

Lemma 2. *Let $t \in [0,T]$ be fixed and $M(t;q)$ be defined as above. If the mapping $q \to z(t;q)$ from $\mathcal{Q} \to Z_\delta$ has a locally Lipschitz continuous Fréchet derivative, then the mapping $\alpha \to M(t;q_\alpha)$ is continuous from $\mathbb{R}^s(\mathcal{Q}) \to \mathcal{L}(\mathbb{R}^s, Y)$. Moreover, for any $\alpha \in \mathbb{R}^s(\mathcal{Q})$, there exist positive constants η_α and $\mathcal{L}_\alpha$ depending on t such that*

$$\| M\left(t; q_\alpha\right) - M\left(t; q_{\tilde{\alpha}}\right) \| \leq \mathcal{L}_\alpha \left| \alpha - \tilde{\alpha} \right|, \qquad \forall \tilde{\alpha} \in B\left(\alpha, \eta_\alpha\right).$$

The same result holds for the mapping $\alpha \to M(t;q_\alpha)^*$.

Proof. Let $t \in [0,T]$ be fixed. By hypothesis, for all $\alpha \in \mathbb{R}^s(\mathcal{Q})$ there exist $\eta_\alpha > 0$, L_α such that $\|z_q(t;q_\alpha) - z_q(t;q_{\tilde{\alpha}})\|_{\mathcal{L}(\tilde{\mathcal{Q}},Z)} < L_\alpha \|q_\alpha - q_{\tilde{\alpha}}\|$ for every $\tilde{\alpha} \in B(\alpha, \eta_\alpha)$. It follows that

$$\begin{aligned}\| M\left(t; q_\alpha\right) - M\left(t; q_{\tilde{\alpha}}\right) \| &= \sup_{\gamma \in \mathbb{R}^s,\, |\gamma|=1} \| \left[M\left(t; q_\alpha\right) - M\left(t; q_{\tilde{\alpha}}\right)\right] \gamma \|_Y \\ &= \sup_{\gamma \in \mathbb{R}^s,\, |\gamma|=1} \| \mathcal{C} z_q(t; q_\alpha) q_\gamma - \mathcal{C} z_q(t; q_{\tilde{\alpha}}) q_\gamma \|_Y \\ &\leq \|\mathcal{C}\|_{\mathcal{L}(Z,Y)} \sup_{\gamma \in \mathbb{R}^s,\, |\gamma|=1} \left\{ \| z_q(t; q_\alpha) - z_q(t; q_{\tilde{\alpha}}) \|_{\mathcal{L}(\tilde{\mathcal{Q}},Z)} \|q_\gamma\|_{\tilde{\mathcal{Q}}} \right\} \\ &\leq \|\mathcal{C}\|_{\mathcal{L}(Z,Y)} L_{q_\alpha} \|q_\alpha - q_{\tilde{\alpha}}\|_{\tilde{\mathcal{Q}}} \\ &= \|\mathcal{C}\|_{\mathcal{L}(Z,Y)} L_{q_\alpha} |\alpha - \tilde{\alpha}| \doteq \mathcal{L}_\alpha |\alpha - \tilde{\alpha}|\end{aligned}$$

for every $\tilde{\alpha} \in B(\alpha, \eta_\alpha)$. ■

Lemma 3. *Under the same hypothesis of Lemma 2, the mapping $\alpha \to D(\alpha)$ is locally Lipschitz continuous from $\mathbb{R}^s(\mathcal{Q}) \to \mathcal{L}(\mathbb{R}^s, \mathbb{R}^s)$.*

Proof. The result easily follows from Lemma 2. In fact, observe that

$$\begin{aligned}\left[D(\alpha) - D(\tilde{\alpha})\right] \gamma &= \sum_{i=1}^{m} M(t_i; q_\alpha)^* M(t_i; q_\alpha) \gamma - \sum_{i=1}^{m} M(t_i; q_{\tilde{\alpha}})^* M(t_i; q_{\tilde{\alpha}}) \gamma \\ &= \sum_{i=1}^{m} M(t_i; q_\alpha)^* \left[M(t_i; q_\alpha) - M(t_i; q_{\tilde{\alpha}})\right] \gamma \\ &\quad + \sum_{i=1}^{m} \left[M(t_i; q_\alpha)^* - M(t_i; q_{\tilde{\alpha}})^*\right] M(t_i; q_{\tilde{\alpha}}) \gamma.\end{aligned}$$

■

Before stating our main results concerning the convergence of the quasilinearization algorithm (QA), we will need to introduce the concept of *point of attraction.* We give its definition below as well as a sufficient condition for an iteration mapping on a Banach space to have a point of attraction.

Definition 4. *Let U be an open subset of a Banach space X and $E : U \subset X \to X$. We say that x^* is a point of attraction of the iteration $x^{k+1} = E(x^k)$ if there exists an open neighborhood S of x^* such that $S \subset U$ and for any $x^0 \in S$, the iterates $x^k \in U$, for all $k \geq 1$ and $x^k \to x^*$ as $k \to \infty$.*

Lemma 5. (Contraction mapping theorem). *Let U be an open subset of a Banach space X, $E : U \subset X \to X$, $x^* \in U$ and suppose there is a ball $B = B(x^*, \eta) \subset U$ and $\alpha \in (0,1)$ such that*

$$\|E(x) - x^*\| \leq \alpha \|x - x^*\|, \qquad \forall x \in B.$$

Then x^* is a point of attraction of the iteration $x^{k+1} = E(x^k)$.

Proof. Whenever $x^0 \in B$, we have that $\left\|x^1 - x^*\right\| = \left\|E(x^0) - x^*\right\| \leq \alpha \left\|x^0 - x^*\right\|$, from which $x^1 \in B$. By induction $\left\|x^{k+1} - x^*\right\| = \left\|E(x^k) - x^*\right\| \leq \alpha \left\|x^k - x^*\right\| \leq \alpha^{k+1} \left\|x^0 - x^*\right\|$ and $\alpha^{k+1} \left\|x^0 - x^*\right\| \to 0$ as $k \to \infty$. ∎

Theorem 6. (Local convergence of the QA under exact fit-to-data assumption). *Assume the hypothesis of Lemma 2 holds. Assume also that there exist an open set $U \subset \mathbb{R}^s(\mathcal{Q})$ and $\alpha^* \in U$ such that $[D(\alpha^*)]^{-1}$ exists and $J(q_{\alpha^*}) = 0$. Let E be the iteration mapping defined at the end of the previous section. Then, for every $\epsilon > 0$, there exists a constant $\delta > 0$ so that $|\alpha - \alpha^*| < \delta$ implies*

$$|E(\alpha) - \alpha^*| \leq K|\alpha - \alpha^*|^2 + \epsilon|\alpha - \alpha^*|$$

where K is a constant depending only on α^* (not on ϵ). In particular, α^* is a point of attraction of the iteration $\alpha^{k+1} = E(\alpha^k)$.

Proof. By definition

$$E(\alpha) = \alpha - [D(\alpha)]^{-1} \left\{ \sum_{i=1}^{m} M(t_i; q_\alpha)^* \left(\mathcal{C}z(t_i; q_\alpha) - \hat{z}_i\right) \right\}$$

whenever $[D(\alpha)]^{-1}$ exists. Hence we have that

$$\begin{aligned}
E(\alpha) - \alpha^* &= \alpha - [D(\alpha)]^{-1} \left\{ \sum_{i=1}^{m} M(t_i; q_\alpha)^* \left(\mathcal{C}z(t_i; q_\alpha) - \hat{z}_i\right) \right\} - \alpha^* \\
&= [D(\alpha)]^{-1} \left\{ D(\alpha)\left(\alpha - \alpha^*\right) - \sum_{i=1}^{m} M(t_i; q_\alpha)^* \left(\mathcal{C}z(t_i; q_\alpha) - \hat{z}_i\right) \right\} \\
&= [D(\alpha)]^{-1} \left\{ \sum_{i=1}^{m} M(t_i; q_\alpha)^* \left[M(t_i; q_\alpha)\left(\alpha - \alpha^*\right) - \mathcal{C}z(t_i; q_\alpha) + \hat{z}_i\right] \right\} \\
&= [D(\alpha)]^{-1} \left\{ \sum_{i=1}^{m} M(t_i; q_\alpha)^* \left[M(t_i; q_\alpha) - M(t_i; q_{\alpha^*})\right]\left(\alpha - \alpha^*\right) \right\} \\
&\quad - [D(\alpha)]^{-1} \left\{ \sum_{i=1}^{m} M(t_i; q_\alpha)^* \left[\mathcal{C}z(t_i; q_\alpha) - \mathcal{C}z(t_i; q_{\alpha^*}) - M(t_i; q_{\alpha^*})\left(\alpha - \alpha^*\right)\right] \right\} \\
&\quad - [D(\alpha)]^{-1} \left\{ \sum_{i=1}^{m} M(t_i; q_\alpha)^* \left[\mathcal{C}z(t_i; q_{\alpha^*}) - \hat{z}_i\right] \right\}.
\end{aligned}$$

Since $J\left(q_{\alpha^*}\right) = 0$, the third term on the right hand side equals zero. Also, since $[D(\alpha^*)]^{-1}$ exists, by continuity there exist positive constants δ_1 and D so that for $|\alpha - \alpha^*| < \delta_1$, we have that $\left|[D(\alpha)]^{-1}\right| \leq D$. From Lemma 2 there exists M such that $\|M(t_i; q_\alpha)^*\| \leq M$ for $i = 1, 2, \cdots, m$, whenever $|\alpha - \alpha^*| < \delta_1$.

Consequently,

$$\begin{aligned}
|E(\alpha) - \alpha^*| &\leq DM \sum_{i=1}^{m} \|[M\left(t_i; q_\alpha\right) - M\left(t_i; q_{\alpha^*}\right)]\left(\alpha - \alpha^*\right)\| \\
&\quad + DM \sum_{i=1}^{m} \|\mathcal{C}z(t_i; q_\alpha) - \mathcal{C}z(t_i; q_{\alpha^*}) - M(t_i; q_{\alpha^*})\left(\alpha - \alpha^*\right)\| \\
&\doteq A + B.
\end{aligned}$$

By Lemma 2, if $|\alpha - \alpha^*| < \eta_{\alpha^*}$, then $A \leq DMm\mathcal{L}_{\alpha^*}|\alpha - \alpha^*|^2$. Also, since

$$M(t_i; q_{\alpha^*})(\alpha - \alpha^*) = \mathcal{C}z_q(t_i; q_{\alpha^*})(q_\alpha - q_{\alpha^*}),$$

from the definition of the Fréchet derivative $z_q(t;q)$, for every $\epsilon > 0$, there exists $\delta_2 = \delta_2(\epsilon, \alpha^*) > 0$ such that $|\alpha - \alpha^*| < \delta_2$ implies

$$\|\mathcal{C}z(t_i; q_\alpha) - \mathcal{C}z(t_i; q_{\alpha^*}) - M(t_i; q_{\alpha^*})(\alpha - \alpha^*)\| \leq \epsilon \|q_\alpha - q_{\alpha^*}\| = \epsilon |\alpha - \alpha^*|,$$

$i = 1, 2, \cdots, m$.

Summarizing, we note that

$$|E(\alpha) - \alpha^*| \leq DMm\left[\mathcal{L}_{\alpha^*}|\alpha - \alpha^*|^2 + \epsilon|\alpha - \alpha^*|\right]$$

for any α such that $|\alpha - \alpha^*| < \delta^* \doteq \min\{\delta_1, \delta_2, \eta_{\alpha^*}\}$. By Lemma 5, α^* is a point of attraction of the iteration $\alpha^{k+1} = E(\alpha^k)$. ■

It is important to note that in Theorem 6 we have assumed an exact fit-to-data at the minimizer α^*. In practice, when working with real parameter identification problems, this is not a realistic assumption due to possible observation, measuring and modelling errors. In the next theorem we weaken this exact fit-to-data assumption.

Theorem 7. (Local convergence of the QA with noisy data). *Assume the hypothesis of Lemma 2 holds. Assume also that there exist an open set $U \subset \mathbb{R}^s(\mathcal{Q})$ and $\alpha^* \in U$ such that $D(\alpha^*)$ is nonsingular and $\alpha^* = E(\alpha^*)$ (fixed point). Let $D \doteq \sup\left\{|D(\alpha)|^{-1} : |\alpha - \alpha^*| \leq \delta_1\right\}$ as in Theorem 6 and $\mathcal{L}$ the smallest constant satisfying*

$$\|M(t_i; q_\alpha)^* - M(t_i; q_{\alpha^*})^*\| \leq \mathcal{L}|\alpha - \alpha^*|, \qquad \forall |\alpha - \alpha^*| < \delta_1, \qquad i = 1, 2, \cdots, m,$$

and suppose

$$\sum_{i=1}^m \|\mathcal{C}z(t_i; q_{\alpha^*}) - \hat{z}_i\| < \frac{1}{D\mathcal{L}}.$$

Then α^* is a point of attraction of the iteration $\alpha^{k+1} = E\left(\alpha^k\right)$.

Proof. Following the same steps as in the proof of Theorem 8, we find that

$$\begin{aligned} |E(\alpha) - \alpha^*| \leq{}& DMm\left[\mathcal{L}|\alpha - \alpha^*|^2 + \epsilon|\alpha - \alpha^*|\right] \\ &+ \left\|[D(\alpha)]^{-1}\sum_{i=1}^m M(t_i; q_\alpha)^*\left[\mathcal{C}z(t_i; q_{\alpha^*}) - \hat{z}_i\right]\right\|. \end{aligned} \tag{3.1}$$

But,

$$\sum_{i=1}^m M(t_i; q_{\alpha^*})^*\left[\mathcal{C}z(t_i; q_{\alpha^*}) - \hat{z}_i\right] = 0, \tag{3.2}$$

since, by assumption, $\alpha^* = E(\alpha^*)$. Combining (3.1) and (3.2) we obtain

$$\begin{aligned} |E(\alpha) - \alpha^*| &\leq DMm\left[\mathcal{L}|\alpha - \alpha^*|^2 + \epsilon|\alpha - \alpha^*|\right] \\ &\quad + D\left\|\sum_{i=1}^m \left[M(t_i; q_\alpha)^* - M(t_i; q_{\alpha^*})^*\right]\left[\mathcal{C}z(t_i; q_{\alpha^*}) - \hat{z}_i\right]\right\| \\ &\leq DMm\left[\mathcal{L}|\alpha - \alpha^*|^2 + \epsilon|\alpha - \alpha^*|\right] \\ &\quad + D\mathcal{L}|\alpha - \alpha^*|\sum_{i=1}^m \|\mathcal{C}z(t_i; q_{\alpha^*}) - \hat{z}_i\| \\ &= DMm\left[\mathcal{L}|\alpha - \alpha^*|^2 + \epsilon|\alpha - \alpha^*|\right] + \gamma|\alpha - \alpha^*| \end{aligned}$$

where $\gamma < 1$ by hypothesis. This concludes the proof. ■

4. A comparison with the standard approach to quasilinearization

In the standard literature ([4], [6]), the quasilinearization algorithm is introduced in a rather different manner than the one shown in the previous sections. For the sake of completennes, we briefly present here this standard, although less intuitive approach. In spite of the fact that at a first glance, the methods look completely dissimilar, we shall show that they both lead to the same iterative process.

Assume for the time being that the nonlinear term $F(q,t,z)$ is Fréchet differentiable with respect to q and z. Given an estimate $q^k \in \mathcal{Q}$ of the minimizer $q^* \in \mathcal{Q}$, we define $z^k(t) = z(t;q^k)$ and linearize problem $(P)_q$ about $\left(q^k, z^k(t)\right)$. This procedure yields the following IVP $(P)_q^k$

$$(P)_q^k \begin{cases} \dot{z}(t) = & Az^k(t) + F(q^k;t,z^k(t)) \\ & +F_q(q^k;t,z^k(t))(q-q^k) \\ & +A(z(t)-z^k(t)) + F_z(q^k;t,z^k(t))(z(t)-z^k(t)) \\ & \\ z(0) = & z_0. \end{cases}$$

Next, we define $z^{k+1}(t;q)$ to be the solution of $(P)_q^k$ and choose q^{k+1} to be a minimizer of the modified error criterion

$$J^k(q) = \frac{1}{2}\sum_{i=1}^m \left\|\mathcal{C}z^{k+1}(t_i;q) - \hat{z}_i\right\|^2.$$

Observing $(P)_q^k$, we see that $v(t) = z^{k+1}(t;q) - z^k(t)$ is a solution of the IVP

$$\begin{cases} \dot{v}(t) = Av(t) + F_q(q^k;t,z(t;q^k))(q-q^k) + F_z(q^k;t,z(t;q^k))v(t), \\ v(0) = 0. \end{cases} \tag{4.1}$$

This system is known as the "sensitivity equations" associated to the ID problem. In [5], Theorem 2.4, it is proved that $v(t)$ is the Fréchet q-derivative of $z(t;q)$ evaluated at q^k and applied to $(q-q^k)$, i.e.,

$$z^{k+1}(t;q) = z(t;q^k) + z_q(t;q^k)(q-q^k).$$

Hence,

$$J^k(q) = \frac{1}{2}\sum_{i=1}^m \left\|\mathcal{C}\left[z(t_i;q^k) + z_q(t_i;q^k)(q-q^k)\right] - \hat{z}_i\right\|_Y^2,$$

which is the same error criterion obtained in Section 2.

As we can see, the classical quasilinearization approach is based upon the *linearization of the initial value problem* around the solution corresponding to the guess parameter and its derivation requires previous knowledge of the sensitivity equations (4.1). On the other hand the method introduced in Section 2 is based simply upon the *linearization of the solution* of the IVP $(P)_q$ around the guess parameter and for the derivation of the algorithm the sensitivity equations are not necessary. We emphasize however that in the computational immplementation both methods make use of the derivatives of the solutions with respect to the unknown parameters and, therefore, of equations (4.1).

5. An application example - Numerical results

In this section we consider an example in which the quasilinearization algorithm is used to solve a parameter estimation problem in the following system of nonlinear partial differential equations:

$$\rho u_{tt} - \beta\rho u_{xxt} + \gamma u_{xxxx} = f(x,t) + \left(2\alpha_2(\theta-\theta_1)u_x - 4\alpha_4 u_x^3 + 6\alpha_6 u_x^5\right)_x, \quad x \in (0,1),\ 0 \le t \le T \tag{5.1a}$$

$$C_v\theta_t - k\theta_{xx} = g(x,t) + 2\alpha_2\theta u_x u_{xt} + \beta\rho u_{xt}^2, \quad x \in (0,1),\ 0 \le t \le T \tag{5.1b}$$

$$u(x,0) = u_0(x),\ u_t(x,0) = v_0(x),\ \theta(x,0) = \theta_0(x), \quad x \in (0,1) \tag{5.1c}$$

$$u(0,t) = u(1,t) = u_{xx}(0,t) = u_{xx}(1,t) = 0, \quad 0 \le t \le T \tag{5.1d}$$

$$\theta_x(0,t) = \theta_x(1,t) = 0, \quad 0 \le t \le T. \tag{5.1e}$$

These equations arise from the conservation laws of linear momentum and energy in a one-dimensional shape memory body. The functions u and θ represent displacement and absolute temperature, respectively. Subindex "x" and "t" denote partial derivatives and ρ, C_v, k, β, γ, α_2, α_4, α_6, θ_1 are positive constants depending on the material being considered. The functions $f(x,t)$ and $g(x,t)$ denote distributed forces and distributed heat sources. For a detailed explanation of the model and the meaning of the parameters involved we refer the reader to [9] and the references therein.

We are interested in using experimental data to estimate the parameters α_2, α_4, α_6 and θ_1. We note here that these are non-physical parameters and therefore they cannot be estimated from laboratory experiments.

Next, we shall formulate the IBVP (5.1) as an abstract nonlinear Cauchy Problem in an appropriate Banach Space. In particular we define the admissible parameter set as $\mathcal{Q} \doteq \left\{q = (\alpha_2, \alpha_4, \alpha_6, \theta_1) \,|q \in \mathbb{R}^4_+\right\}$, the state space Z as the Hilbert space $H^1_0(0,1) \cap H^2(0,1) \times L^2(0,1) \times L^2(0,1)$ with the inner product

$$\left\langle \begin{pmatrix} u \\ v \\ \theta \end{pmatrix}, \begin{pmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{\theta} \end{pmatrix} \right\rangle \doteq \gamma \int_0^1 u''(x)\overline{\tilde{u}''(x)}\,dx + \rho \int_0^1 v(x)\overline{\tilde{v}(x)}\,dx + \frac{C_v}{k}\int_0^1 \theta(x)\overline{\tilde{\theta}(x)}\,dx.$$

The operator A on Z is defined by

$$D(A) = \left\{ \begin{pmatrix} u \\ v \\ \theta \end{pmatrix} \in Z \;\middle|\; \begin{array}{l} u \in H^4(0,1),\ u(0) = u(1) = u''(0) = u''(1) = 0 \\ v \in H^1_0(0,1) \cap H^2(0,1) \\ \theta \in H^2(0,1), \theta'(0) = \theta'(1) = 0 \end{array} \right\}$$

and for $z = \begin{pmatrix} u \\ v \\ \theta \end{pmatrix} \in D(A)$,

$$A\begin{pmatrix} u \\ v \\ \theta \end{pmatrix} \doteq \begin{pmatrix} 0 & I & 0 \\ -\frac{\gamma}{\rho}D^4 & \beta D^2 & 0 \\ 0 & 0 & \frac{k}{C_v}D^2 \end{pmatrix} \begin{pmatrix} u \\ v \\ \theta \end{pmatrix},$$

where $D^n \doteq \frac{\partial^n}{\partial x^n}$.

We also define $z_0(x) = \begin{pmatrix} u_0(x) \\ v_0(x) \\ \theta_0(x) \end{pmatrix}$ and $F(q,t,z) : \mathcal{Q} \times [0,T] \times D \to Z$ by

$$F(q,t,z) = \begin{pmatrix} 0 \\ f_2(q,t,z) \\ f_3(q,t,z) \end{pmatrix},$$

where

$$\begin{aligned} \rho f_2(q,t,z)(x) &= f(x,t) + \left(2\alpha_2(\theta - \theta_1)u_x - 4\alpha_4 u_x^3 + 6\alpha_6 u_x^5\right)_x, \\ C_v f_3(q,t,z)(x) &= g(x,t) + 2\alpha_2 \theta u_x v_x + \beta\rho v_x^2 \end{aligned}$$

and $D = H^1_0(0,1) \cap H^2(0,1) \times H^1(0,1) \times H^1(0,1)$.

With the above notation, the IBVP (5.1a-e) is equivalent to the following abstract Cauchy problem in the Hilbert space Z:

$$(\mathcal{P}) \begin{cases} \frac{d}{dt}z(t) = Az(t) + F(q,t,z), & 0 \le t \le T \\ z(0) = z_0. \end{cases} \tag{5.2}$$

We assume the following standing hypothesis.

(H2). *For each fixed $t \geq 0$, the functions $f(x,t)$, $g(x,t)$ are in $L^2(0,1)$ and there exist nonnegative functions $K_f(x)$, $K_g(x) \in L^2(0,1)$ such that*

$$|f(x,t_1) - f(x,t_2)| \le K_f(x)|t_1 - t_2|, \quad |g(x,t_1) - g(x,t_2)| \le K_g(x)|t_1 - t_2|$$

for all $x \in (0,1)$, $t_1, t_2 \in [0,T]$.

The following results can be easily derived from theorems 3.7 and 3.11 in [10] with only slight modifications in order to take into account for the different boundary conditions being considered here. Since the modifications needed are trivial and not important for the goals pursued by this article, we do not give details here.

Theorem 11. *The operator A defined above generates an analytic semigroup $T(t)$ in Z and if* **(H2)** *holds, then the mapping F as defined above satisfies* **(H1)** *for any $\delta \in (\frac{3}{4}, 1)$.*

The following theorem shows that the operator A and the function F satisfy certain regularity conditions, which, in view of Theorems 2.4 and 3.1 in [5], ensure the existence and Lipschitz continuity of the Fréchet derivative of the mapping $q \to z(t;q)$. This result, together with Theorems 6 and 7, will lead to the local convergence of the quasilinearization algorithm to the optimal parameter.

Theorem 12. *Let Z, A and $F(q,t,z)$ be as defined above and assume* **(H2)** *holds. Then the mapping $(q, z(\cdot)) \to F(q, \cdot, z(\cdot))$ from $\mathcal{Q} \times L^\infty(0, T : Z_\delta)$ into $L^\infty(0, T : Z)$ is Fréchet differentiable in both variables. Also, the mappings $(q, z(\cdot)) \to F_q(q, \cdot, z(\cdot))$ and $(q, z(\cdot)) \to F_z(q, \cdot, z(\cdot))$ are locally Lipschitz continuous from $\mathcal{Q} \times L^\infty(0, T : Z_\delta)$ into $L^\infty(0, T : \mathcal{L}(\tilde{\mathcal{Q}}, Z))$ and from $\mathcal{Q} \times L^\infty(0, T : Z_\delta)$ into $L^\infty(0, T : \mathcal{L}(Z_\delta, Z))$, respectively.*

Proof. This result follows immediately observing that $f_2(q,t,z)$ and $f_3(q,t,z)$, as previously defined, are Fréchet differentiable with respect to q and z. Moreover, these derivatives can be computed explicitely and are given by:

$$D_z f_2(q, t, \begin{pmatrix} u \\ v \\ \theta \end{pmatrix}) \begin{pmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{\theta} \end{pmatrix} = f_{2,u}\tilde{u} + f_{2,v}\tilde{v} + f_{2,\theta}\tilde{\theta},$$

$$D_z f_3(q, t, \begin{pmatrix} u \\ v \\ \theta \end{pmatrix}) \begin{pmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{\theta} \end{pmatrix} = f_{3,u}\tilde{u} + f_{3,v}\tilde{v} + f_{3,\theta}\tilde{\theta},$$

$$D_q f_2(q, t, \begin{pmatrix} u \\ v \\ \theta \end{pmatrix}) = \frac{1}{\rho}\left[2\theta' u' + 2(\theta - \theta_1)u'' \,,\, -12\,(u')^2\, u'' \,,\, 30\,(u')^4\, u'' \,,\, -2\alpha_2 u''\right],$$

$$D_q f_3(q, t, \begin{pmatrix} u \\ v \\ \theta \end{pmatrix}) = \frac{1}{C_v}\left[2\theta u' v' \,,\, 0 \,,\, 0 \,,\, 0\right],$$

where the linear operators $f_{i,u}$, $f_{i,v}$ and $f_{i,\theta}$, $i = 2, 3$ are given by

$$\begin{aligned}
f_{2,u} &= \frac{1}{\rho}\Big\{2\alpha_2\theta' D + 2\alpha_2(\theta - \theta_1)D^2 - 24\alpha_4 u' u'' D - 12\alpha_4\,(u')^2\, D^2 \\
&\qquad + 120\alpha_6\,(u')^3\, u'' D + 30\alpha_6\,(u')^4\, D^2\Big\}, \\
f_{2,v} &= 0 \\
f_{2,\theta} &= \frac{1}{\rho}\left\{2\alpha_2 u' D + 2\alpha_2 u''\right\}, \\
f_{3,u} &= \frac{1}{C_v}\left\{2\alpha_2\theta v' D\right\} \\
f_{3,v} &= \frac{1}{C_v}\left\{2\alpha_2\theta u' D + 2\beta\rho v' D\right\}, \\
f_{3,\theta} &= \frac{1}{C_v}\left\{2\alpha_2 u' v'\right\}.
\end{aligned}$$

∎

In all the examples that follow we make use of the parameter values reported by F. Falk in [FALK] for the alloy $Au_{23}Cu_{30}Zn_{47}$. These values are: $\alpha_2 = 24$ J cm^{-3} K^{-1}, $\alpha_4 = 1.5 \times 10^5$ J cm^{-3} , $\alpha_6 = 7.5 \times 10^6$ J cm^{-3} K^{-1}, $\theta_1 = 208$ K, $C_v = 2.9$ J cm^{-3} K^{-1}, $k = 1.9$ w cm^{-1} K^{-1}, $\rho = 11.1$ g cm^3, $\beta = 1$ and $\gamma = 10^{-12}$ J cm^{-1}. We want to estimate $q^* = (\alpha_2, \alpha_4, \alpha_6, \theta_1) = (24, 1.5 \times 10^5, 7.5 \times 10^6, 208)$.

Example 1: *Exact data*.

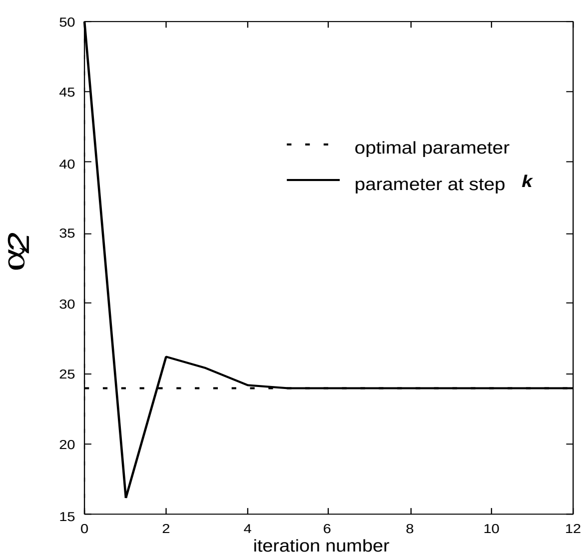
For this example we take $u_0 \equiv 0$, $v_0 \equiv 0$, $\theta_0 \equiv 200$ K, $g(x,t) \equiv 0$,

$$f(x,t) = \begin{cases} 1 \times 10^5, & \text{if } 0.4 \le x \le 0.6, \\ 0, & \text{otherwise} \end{cases}$$

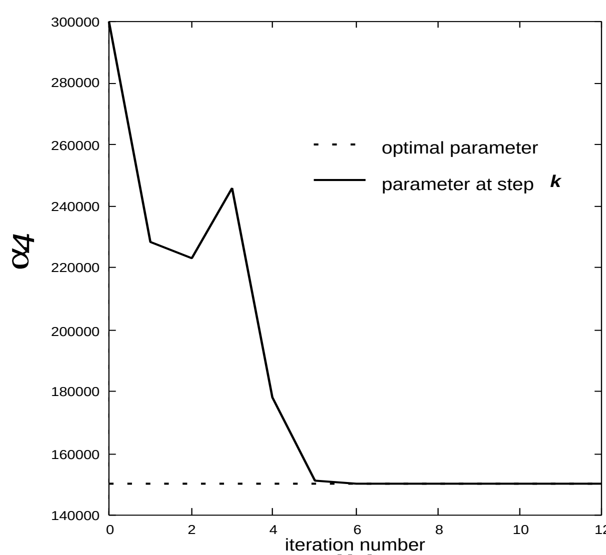
and $T = 0.01$. First, we obtain $u(t,x,q^*)$ and $\theta(t,x,q^*)$ by numerically solving the problem. For this purpose we make use of the spectral method proposed in [7]. The observations are then taken to be $\hat{z}_i = \left\{ \begin{pmatrix} u(x_j,t_i;q^*) \\ \theta(x_j,t_i;q^*) \end{pmatrix} \right\}_{j=1}^{9}$, where $t_i = 0.001\,i$, $i = 1,2,\cdots,10$. We start with an initial estimate $q^0 = (50,\, 3\times 10^5,\, 15\times 10^6,\, 420)$, approximately equal to twice q^*. The results of the iterations produced by the quasilinearization algorithm are shown in Table 1 and Figures 1.a-d. Figure 2b shows a comparison between $u(x,T;q^*)$ and $u(x,T;q^k)$ while in Figure 2b $\theta(x,T;q^*)$ and $\theta(x,T;q^k)$ are drawn for different values of k.

k	α_2	α_4	α_6	θ_1	$J(q^k)$
0	50.0000	300000	1.50000e+07	420.000	1994.6900
1	16.1807	228111	1.40769e+07	459.904	611.1950
2	26.1790	222964	8.71784e+06	33.096	280.8220
3	25.3531	246241	8.83171e+06	126.468	15.3156
4	24.2770	178223	7.87660e+06	181.091	7.1313
5	24.0166	151184	7.51451e+06	206.550	0.6210
6	24.0012	150073	7.50096e+06	207.927	0.0122
7	24.0001	150006	7.50008e+06	207.994	0.0030
8	24.0001	150002	7.50003e+06	207.998	0.0029
9	24.0000	150002	7.50002e+06	207.998	0.0029
10	24.0000	150002	7.50002e+06	207.998	0.0029
11	24.0000	150002	7.50002e+06	207.998	0.0029
12	24.0000	150002	7.50002e+06	207.998	0.0029

Table 1: Values of the parameters and of the error criterion at different iteration steps.



(a)



(b)

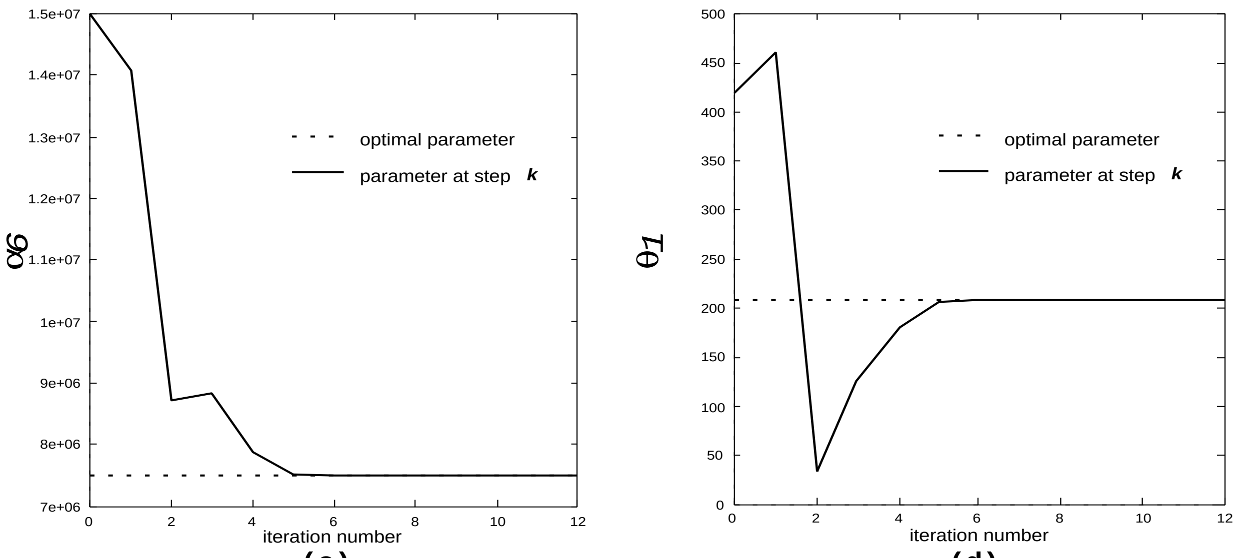


Figure 1: Evolution of the iterations for Example 1.

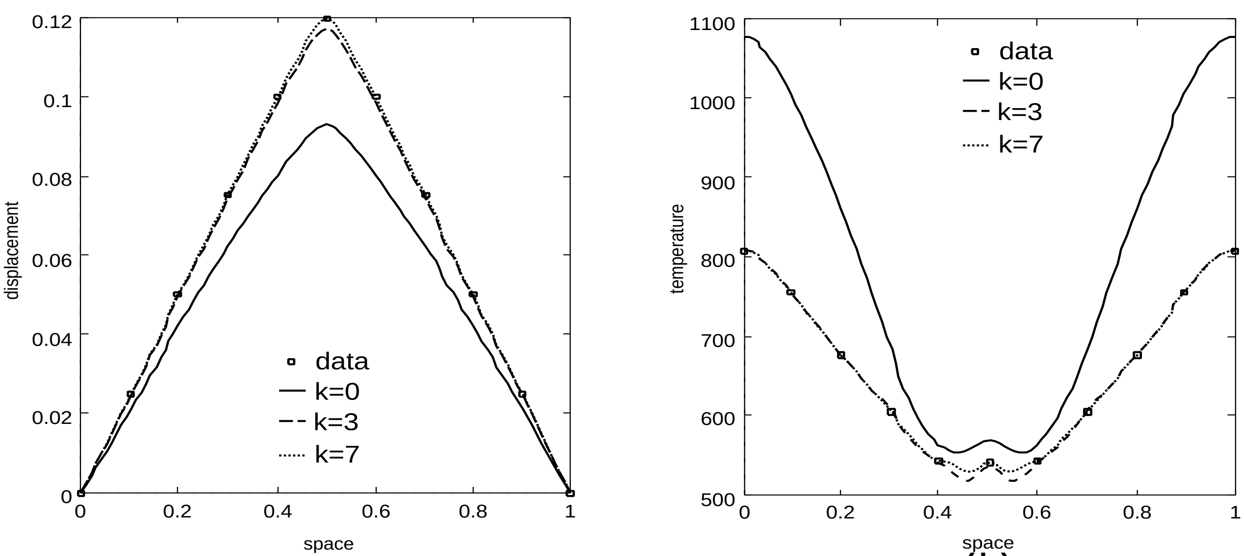


Figure 2: Displacement (a) and Temperature (b) at $T = 0.01$ for $q = q^k$, $k = 0, 3, 7$.

Example 2: *Noisy data*.

This example is analogous to Example 1, except that now we add random noise to the observation data in order to simulate measuring errors. More precisely, the observations are taken to be $\hat{z}_i = \left\{ \begin{pmatrix} u(x_j, t_i; q^*) + r_{i,j} \\ \theta(x_j, t_i; q^*) + \tilde{r}_{i,j} \end{pmatrix} \right\}_{j=1}^{9}$, where $r_{i,j}$ and $\tilde{r}_{i,j}$ are random numbers uniformly distributed in $(-0.05\overline{u}, 0.05\overline{u})$ and $\left(-0.05\,\overline{\theta}, 0.05\,\overline{\theta}\right)$, respectively, with $\overline{u} = \dfrac{1}{90}\displaystyle\sum_{i=1}^{10}\sum_{j=1}^{9} |u(x_i, t_i; q^*)|$ and $\overline{\theta} = \dfrac{1}{90}\displaystyle\sum_{i=1}^{10}\sum_{j=1}^{9} \theta(x_i, t_i; q^*)$. The initial estimate is again $q^0 = (50,\, 3 \times 10^5,\, 15 \times 10^6,\, 420)$. The results of the iterations are shown in Table 2, and Figure 3. Figure 4a shows a comparison between $u(x, T; q^*)$ and $u(x, T; q^k)$ while in Figure 4b $\theta(x, T; q^*)$ and $\theta(x, T; q^k)$ are drawn for different values of k.

k	α_2	α_4	α_6	θ_1	$J(q^k)$
0	50.0000	300000	1.50000e+07	420.000	1987.240
1	16.5263	251533	1.43413e+07	450.975	604.570
2	26.7351	173032	7.92651e+06	77.3584	261.591
3	25.1282	223785	8.54573e+06	148.386	111.619
4	24.2875	176007	7.84280e+06	189.479	111.030
5	24.4436	183683	7.95702e+06	183.663	110.985
6	24.4070	180771	7.91592e+06	186.193	110.977
7	24.4184	181677	7.92857e+06	185.411	110.979
8	24.4151	181408	7.92483e+06	185.645	110.978
9	24.4161	181487	7.92593e+06	185.576	110.979
10	24.4158	181464	7.92560e+06	185.596	110.978
11	24.4159	181471	7.92570e+06	185.590	110.978
12	24.4159	181469	7.92567e+06	185.592	110.978
13	24.4159	181469	7.92568e+06	185.592	110.978

Table 2: Values of the parameters and of the error criterion at different iteration steps.

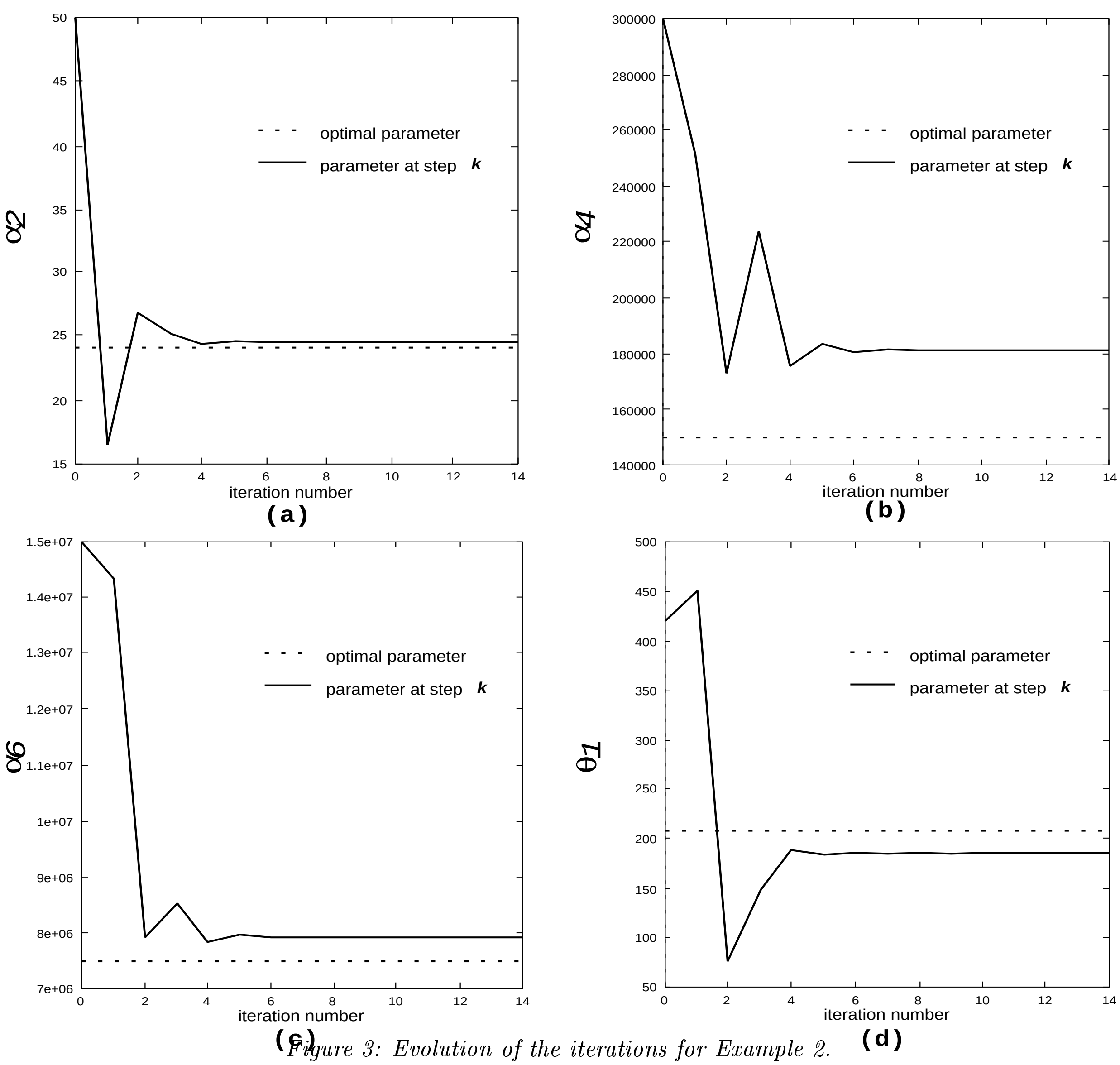


Figure 3: Evolution of the iterations for Example 2.

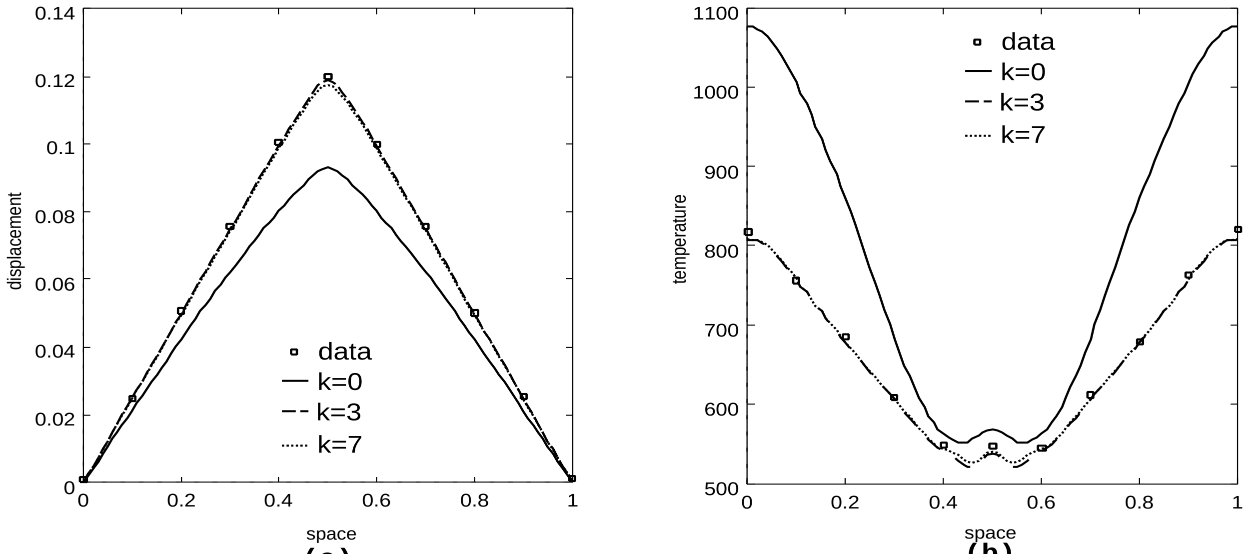


Figure 4: Displacement (a) and Temperature (b) at $T = 0.01$ for $q = q^k$, $k = 0, 3, 7$.

Example 3: *Comparison between direct and indirect methods*.

In this example simultaneously we solve the ID problem using an indirect method and the algorithm proposed in Section 2. The purpose is to illustrate the different convergence rates of the two approaches. We take $u_0 \equiv 0$, $v_0 \equiv 0$, $\theta_0 \equiv 200$, $f \equiv g \equiv 0$. The indirect method consits of approximating the solution of the dynamic equations using the algorithm proposed in [7] and applying afterwards the optimization algorithm of Hooke and Jeeves [3] to solve the resulting optimization problem. We obtain $\hat{z}_i$ as in Example 1 and start with the initial estimate $q^0 = \big(25,\, 2\times 10^5,\, 9\times 10^6,\, 220\big)$. The results of the iterations are shown in Table 3.

k	α_2		α_4		α_6		θ_1	
	D	I	D	I	D	I	D	I
0	25	25	200000	200000	9e+06	9e+06	215	215
12	24.004	24.1250	149991	176000	7500020	8865000	207.999	202.1
40	24.004	24.9375	149991	161500	7500020	7537500	207.999	202.1
100	24.004	25.7344	149991	154000	7500020	6907500	207.999	202.1
500	24.004	24.8140	149991	149738	7500020	7333770	207.999	206.564
1000	24.004	24.4651	149991	149967	7500020	7462530	207.999	207.355
2000	24.004	24.1638	149991	150040	7500020	7499950	207.999	207.801
3000	24.004	24.0651	149991	150011	7500020	7500490	207.999	207.924

Table 3: Comparison of the convergence speeds between a direct and an indirect method.

6. Conclusions

We have introduced a new approach for identifying the unknown parameter q in nonlinear abstract Cauchy problems of the type $\dot{z}(t) = Az(t) + F(q, t, z(t))$. This approach has two main advantages over classical methods. First of all it is much more intuitive since it is based upon linearization of the solution about an initial guess parameter rather than the linearization of the whole problem about a particular solution. Secondly, unlike in the classical setting, the derivation of the algorithm does not rely upon the sensitivity equations.

We have also derived sufficient conditions for the convergence of the algorithm in terms of the regularity of the solutions with respect to the unknown parameter.

Finally, an application was considered in which the nonphysical parameters that define the free energy potential in a mathematical model for shape memory alloys are estimated. Also, several numerical examples are presented and convergence speeds are compared with those of an indirect method.

References

[1] H. T. Banks, J. A. Burns, E. M. Cliff, *"Parameter Estimation and Identification for Systems with Delays"*, SIAM Journal on Control and Optimization, 19(1981), pp. 791-828.

[2] H. T. Banks, G. M. Groome, Jr., *"Convergence Theorems for Parameter Estimation by Quasilinearization"*, Journal of Mathematical Analysis and Aplications, 42(1973), pp. 91-109.

[3] M. S. Bazaraa, C. M. Shetty, *"Nonlinear Programming. Theory and Algorithms"*, John Wiley and Sons, 1979.

[4] D. W. Brewer, J. A. Burns, E. M. Cliff, *"Parameter Identification for an Abstract Cauchy Problem by Quasilinearization"*, Quarterly of Applied Mathematics, 51(1993), pp. 1-22.

[5] J. A. Burns, P. Morin, R. D. Spies, *"Differentiability with respect to a parameter of the solutions of nonlinear abstract Cauchy problems"*, ICAM Report 97-06-01, Interdisciplinary Center for Applied Mathematics, Virginia Polytechnic Institute and State University.

[6] P. W. Hammer, *"Parameter Identification in Parabolic Partial Differential Equations using Quasilinearization"*, Ph. D. Thesis, ICAM Report 90-07-01, Interdisciplinary Center for Applied Mathematics, Virginia Polytechnic Institute and State University.

[7] P. Morin, R. D. Spies, *"Convergent Spectral Approximations for the Thermomechanical Processes in Shape Memory Alloys"*, IMA Preprint # 1381, March 1996, Institute for Mathematics and its Applications, University of Minnesota.

[8] A. Pazy, *"Semigroups of Linear Operators and Applications to Partial Differential Equations"*, Corrected Second Printing, Springer-Verlag, 1983.

[9] R. D. Spies, *"A State-Space Approach to a One-Dimensional Mathematical Model for the Dynamics of Phase Transitions in Pseudoelastic Materials"*, Journal of Mathematical Analysis and Applications, 190, (1995), pp. 58-100.

[10] R. D. Spies, *"Results on a Mathematical Model of Thermomechanical Phase Transitions in Shape Memory Materials"*, Journal of Smart Materials and Structures, 3(1994), pp. 459-469.