Volume 5, Number 1, 2005 ISSN 1562-8353

Nonlinear Dynamics and Systems Theory

An International Journal of Research and Surveys

Academic Periodical Press

Nonlinear Dynamics and Systems Theory

An International Journal of Research and Surveys

NONLINEAR DYNAMICS AND SYSTEMS THEORY
An International Journal of Research and Surveys
Published since 2001

Volume 5 Number 1 2005

CONTENTS

Founded by A.A.Martynyuk in 2001.
Registered in Ukraine Number: KB №5267 / 04.07.2001.

NONLINEAR DYNAMICS AND SYSTEMS THEORY
An International Journal of Research and Surveys

Nonlinear Dynamics and Systems Theory (ISSN 1562-8353 (Print), ISSN 1813-7385 (Online)) is an international journal published under the auspices of the S.P.Timoshenko Institute of Mechanics of National Academy of Sciences of Ukraine and the Laboratory for Industrial and Aplied Mathematics (LIAM) at York University (Toronto, Canada). It is aimed at publishing high quality original scientific papers and surveys in area of nonlinear dynamics and systems theory and technical reports on solving practical problems. The scope of the journal is very broad covering:

SCOPE OF THE JOURNAL

Analysis of uncertain systems • Bifurcations and instability in dynamical behaviors • Celestial mechanics, variable mass processes, rockets • Control of chaotic systems • Controllability, observability, and structural properties • Deterministic and random vibrations • Differential games • Dynamical systems on manifolds • Dynamics of systems of particles • Hamilton and Lagrange equations • Hysteresis • Identification and adaptive control of stochastic systems • Modelling of real phenomena by ODE, FDE and PDE • Nonlinear boundary problems • Nonlinear control systems, guided systems • Nonlinear dynamics in biological systems • Nonlinear fluid dynamics • Nonlinear oscillations and waves • Nonlinear stability in continuum mechanics • Non-smooth dynamical systems with impacts or discontinuities • Numerical methods and simulation • Optimal control and applications • Qualitative analysis of systems with aftereffect • Robustness, sensitivity and disturbance rejection • Soft computing: artificial intelligence, neural networks, fuzzy logic, genetic algorithms, etc. • Stability of discrete systems • Stability of impulsive systems • Stability of large-scale power systems • Stability of linear and nonlinear control systems • Stochastic approximation and optimization • Symmetries and conservation laws

PUBLICATION AND SUBSCRIPTION INFORMATION

The Nonlinear Dynamics and Systems Theory is published four times per year in 2005.

Base list subscription price per volume: US$149.00. This price is available only to individuals whose library subscribes to the journal OR who warrant that the Journal is for their own use and provide a home address for mailing. Separate rates apply to academic and corporate/ government institutions. Our charge includes postage, packing, handling and airmail delivery of all issues. Mail order and inquires to: Department of Processes Stability, S.P.Timoshenko Institute of Mechanics NAS of Ukraine, Nesterov str.,3, 03057, Kiev-57, MSP 680, Ukraine, Tel: ++38-044-456-6140, Fax: ++38-044-456-1939, E-mail: anchern@stability.kiev.ua, http:\\www.sciencearea.com.ua

ABSTRACTING INFORMATION

Nonlinear Dynamics and Systems Theory is currently abstracted/indexed by Zentralblatt MATH and Mathematical Reviews.

Nonlinear Dynamics and Systems Theory, **5**(1) (2005) 1–8

Topological Sequence Entropy and Chaos of Star Maps*

J.S. Cánovas

Departamento de Matemática Aplicada y Estadística,
Universidad Politécnica de Cartagena,
Paseo de Alfonso XIII, 30203 Cartagena (Murcia), Spain

Received: October 02, 2003; Revised: May 25, 2004

Abstract: Let $\mathbb{X}_n = \{z \in \mathbb{C}\colon z^n \in [0,1]\}$, $n \in \mathbb{N}$, and let $f\colon \mathbb{X}_n \to \mathbb{X}_n$ be a continuous map such that $f(0) = 0$. In this paper we prove that f is chaotic in the sense of Li–Yorke iff there is a strictly increasing sequence of positive integers A such that the topological sequence entropy of f relative to A is positive.

Keywords: *Star maps; Li–Yorke chaos; topological sequence entropy.*

Mathematics Subject Classification (2000): 37B40, 37E25.

1 Introduction

Let (X, d) be a compact metric space and let $C(X)$ denote the set of continuous maps $f\colon X \to X$. For any $f \in C(X)$, the pair (X, f) is called a *discrete (semi)dynamical system*. Given $x \in X$, the sequence $(f^i(x))_{i=0}^{\infty}$ is the *trajectory* of x (also *orbit* of x). Recall that a point $x \in X$ is *periodic* if $f^i(x) = x$ for some $i \in \mathbb{N}$. Denote by $\mathrm{Per}\,(f)$ the set of periodic points of f. The map f is said to be *chaotic in the sense of Li–Yorke* (or simply *chaotic*) if there is an uncountable set $S \subset X \backslash \mathrm{Per}\,(f)$ such that for any $x, y \in S$, $x \neq y$, and any $p \in \mathrm{Per}\,(f)$ it holds

$$\liminf_{n\to\infty} d(f^n(x), f^n(y)) = 0, \tag{1}$$

$$\limsup_{n\to\infty} d(f^n(x), f^n(y)) > 0, \tag{2}$$

$$\limsup_{n\to\infty} d(f^n(x), f^n(p)) > 0. \tag{3}$$

*This paper has been partially supported by the grant PI–8/00807/FS/01 from Fundación Séneca (Comunidad Autónoma de Murcia).

The set S is called a *scrambled* set of f (see [18] or [21]).

The notion of chaos in the sense of Li–Yorke has been studied in the case of $X = [0,1]$ and $X = S^1$ (see e.g. [10, 14, 16, 17, 22] or [21]). In this setting, topological sequence entropy plays a special role to characterize chaotic maps. Given a strictly increasing sequence of positive integers A, denote by $h_A(f)$ the *topological sequence entropy* of f with respect to A (see the definition in the next section). Then

Theorem 1 *Let $f \in C([0,1]) \cup C(S^1)$. Then:*

(1) *f is chaotic iff there is a strictly increasing sequence of positive integers A such that $h_A(f) > 0$;*

(2) *for any sequence A there is a chaotic map $f_A \in C(I)$ (resp. $f_A \in C(S^1)$) such that $h_A(f_A) = 0$.*

Statement (1) was proved by Franzová and Smítal [14] for interval maps and by Hric [17] for circle maps. Statement (2) was also proved by Hric [16, 17]).

Theorem 1 (1) is false in general in the case of two dimensional maps (see [13, [20]). So, the following question remains open: is Theorem 1 true for continuous maps defined on finite graphs?

In this paper we give a partial answer to this question. More precisely, we consider the n-star $\mathbb{X}_n = \{z \in \mathbb{C} \colon z^n \in [0,1]\}$, $n \in \mathbb{N}$. Dynamical systems generated by continuous maps on the n-star have been studied in the literature (see [1, 3−6, 8]). Moreover, the construction of chaotic n-star maps holding (2) in Theorem 1 was made in [17]. Let $C_0(\mathbb{X}_n)$ be the set of continuous maps $f \in C(\mathbb{X}_n)$ such that $f(0) = 0$. The aim of this paper is to prove the following result which extends Theorem 1 (1) to the space $C_0(\mathbb{X}_n)$.

Theorem 2 *Let $f \in C_0(\mathbb{X})$. Then f is chaotic iff there is an strictly increasing sequence of positive integers A such that $h_A(f) > 0$.*

This paper is organized as follows. Next section is devoted to introduce basic notation and definitions. In Section 3 we prove useful technical results which are used in the last section, where the main result is proved.

2 Basic Notation

First we introduce the notion of topological sequence entropy for continuous maps defined on compact metric spaces. Let (X,d) be a compact metric space and let $f \in C(X)$. Consider a strictly increasing sequence of positive integers $A = (a_i)_{i=1}^{\infty}$ and let $Y \subseteq X$ and $\varepsilon > 0$. We say that a subset $E \subset Y$ is (A,ε,m,Y,f)-separated if for any $x,y \in E$, $x \neq y$, there is an $i \in \{1,\dots,m\}$ such that $d(f^{a_i}(x), f^{a_i}(y)) > \varepsilon$. Denote by $s_m(A,\varepsilon,Y,f)$ the cardinality of any maximal (A,ε,m,Y,f)-separated set. Define

$$s(A,\varepsilon,Y,f) = \limsup_{m\to\infty} \frac{1}{m} \log s_m(A,\varepsilon,Y,f). \tag{4}$$

Let

$$h_A(f,Y) = \lim_{\varepsilon\to 0} s(A,\varepsilon,Y,f). \tag{5}$$

The *topological sequence entropy* of f respect to the sequence A is

$$h_A(f) = h_A(f,X). \tag{6}$$

When $A = (i)_{i=0}^{\infty}$, we receive the usual definition of topological entropy (see [2, Chapter 4]).

For $x \in X$, let $\omega(x, f)$ denote the set of limit points of the sequence $(f^i(x))_{i=0}^{\infty}$. $\omega(x, f)$ is called the *omega limit set* of f at x. Let $\omega(f) = \bigcup_{x\in X} \omega(x, f)$ be the *omega limit set* of f.

Now, we introduce some definitions on n-star maps. The point $0 \in \mathbb{X}_n$ is called the *branching point* of $\mathbb{X}_n$. The connected components of $\mathbb{X}_n \backslash \{0\}$ are called *branches* of $\mathbb{X}$, denoted by $B_1, \dots, B_n$. For $Y \subset \mathbb{X}_n$, $\overline{Y}$ denotes the closure of Y. $|x|$ denotes the module of $x \in \mathbb{X}$. For a fixed $i \in \{1, \dots, n\}$ and $x, y \in \overline{B}_i$, we write $x < y$ (resp. $x \le y$) to denote $|x| < |y|$ (resp. $|x| \le |y|$). For $x, y \in \overline{B}_i$, $x \le y$, by an *interval* we understand the set $[x, y] = \{z \in \overline{B}_i \colon x \le z \le y\}$, $(x, y]$, $[x, y)$ and (x, y) will be understand in the obvious way. Then, for $1 \le i \le n$, the closure $\overline{B}_i = [0, z_i]$, with $z_i^n = 1$. Now, define a metric on $\mathbb{X}_n$ as follows. For any $x, y \in \mathbb{X}_n$, let

$$\rho(x, y) = \begin{cases} |x - y| & \text{if } x \text{ and } y \text{ lie in the same branch;} \\ |x| + |y| & \text{if } x \text{ and } y \text{ do not lie in the same branch.} \end{cases}$$

For any $x \in \mathbb{X}_n$ and $\varepsilon > 0$, let $B(x, \varepsilon) = \{y \in \mathbb{X}_n \colon \rho(x, y) < \varepsilon\}$.

Finally, we recall the notion of horseshoe (see [19]). Let $k \in \mathbb{N}$. We say that f has a k-*horseshoe* if there is a closed interval J contained in one branch of $\mathbb{X}_n$ and there are k closed subintervals $J_i \subset J$, $1 \le i \le k$, with pairwise disjoint interiors such that $J \subseteq f(J_i)$ for $1 \le i \le k$.

3 Preliminary Results

This section is devoted to state some results which help us to prove the main theorem. We use basically two ideas in the proof. The first one is based on the following proposition.

Proposition 3 *Let (X, d) be a compact metric space and let $f \in C(X)$. Then, for any $k \in \mathbb{N}$ the following statements hold:*

(1) *f^k is chaotic iff f is chaotic;*

(2) *for any strictly increasing sequence A there is a strictly increasing sequence B such that $h_B(f^k) \ge h_A(f)$.*

Proof (1) is a well-known fact which is due to the uniform continuity of f. (2) was proved in [17].

We begin with the n-star case. For any $x \in \mathbb{X}_n$, let $s(x) = (s_i)_{i=0}^{\infty} \in \{0, 1, \dots, n\}^{\mathbb{N}}$ be defined by $s_i = j$ iff $f^i(x) \in B_j$ for some $j \in \{1, \dots, n\}$, and $s_i = 0$ iff $f^i(x) = 0$. We say that $s(x)$ is *eventually constant* if there is $k \in \mathbb{N}$ such that $s_i = s_k$ for all $i \ge k$. We say that $f \in C_0(\mathbb{X}_n)$ has *property* $\mathcal{P}$ if the condition $x \in \mathrm{Per}\,(f)$ implies that $s(x)$ is a constant sequence. The following remark is immediate but useful in what follows. We will use it without citation.

Remark 1 If for some $k \in \mathbb{N}$ we have $f^k(x) = 0$ then $f^l(x) = 0$ for each $l \ge k$ and, hence, $\omega(x, f) = \{0\}$ and $s(x)$ is eventually constant.

Property $\mathcal{P}$ is the other key which allows us to prove the main result. As we will see later, any map from $C_0(\mathbb{X}_n)$ with zero topological entropy has an iterate which holds property $\mathcal{P}$. This fact jointly with Proposition 3 are the main ideas for proving our result. Notice that maps from $C_0(\mathbb{X}_n)$ having property $\mathcal{P}$ have every periodic orbit contained in one branch, which is useful for proving next three lemmas proved previously in the proof of Lemmas 5 and 6 from [11].

Lemma 4 *Assume $f \in C_0(\mathbb{X}_n)$ has property $\mathcal{P}$. Let $x \in B_i$, $1 \le i \le n$, be such that $f(x) \notin B_i$. Then for any $k \in \mathbb{N}$ such that $f^k(x) \in \overline{B}_i$ it follows that $f^k(x) < x$.*

Lemma 5 *Assume $f \in C_0(\mathbb{X}_n)$ has property $\mathcal{P}$. Let $x \in \mathbb{X}_n$ be such that $s(x)$ is not eventually constant. Then $\lim_{k\to\infty} f^k(x) = 0$, that is, $\omega(x, f) = \{0\}$.*

For any $Y \subset \mathbb{X}_n$, let $\tau_Y : \mathbb{X}_n \to Y$ be the natural retraction from $\mathbb{X}_n$ to Y. For $f \in C_0(\mathbb{X}_n)$, let $f_Y \in C(Y)$ be defined by $f_Y = \tau_Y \circ f|_Y$, where $f|_Y$ means the restriction of f to Y. For each $i \in \{1, \dots, n\}$ define $f_i = f_{\overline{B}_i} \in C(\overline{B}_i)$. Notice that f_i is conjugated to a map $g \in C([0, 1])$ holding $g(0) = 0$.

Lemma 6 *Assume $f \in C_0(\mathbb{X}_n)$ has property $\mathcal{P}$. Then for all $j \in \{1, \dots, n\}$, $\omega(f) \cap \overline{B}_j = \omega(f_j)$. In particular, $\omega(f) = \bigcup_{j=1}^n \omega(f_j)$ and hence $\omega(f)$ is compact.*

Let $x, y \in \mathbb{X}_n$. We write $x \prec y$ to mean that either $x < y$ or $x \in B_i$ and $y \in B_j$ for some $i, j \in \{1, \dots, n\}$, $i \ne j$. Given $S, T \subset \mathbb{X}$, we say that $S \prec T$ if $s \prec t$ for all $s \in S$ and $t \in T$.

Lemma 7 *Assume $f \in C_0(\mathbb{X}_n)$ has property $\mathcal{P}$ and $h(f) = 0$. Let $x \in B_i$ for some $1 \le i \le n$ and let $a \in B_i$ be such that $x \prec a$ and $f(1) = 0$. Then $f^i(x) \prec a$ for all $i \in \mathbb{N}$.*

Proof Assume the contrary and let $j \in \mathbb{N}$ be such that $a \prec f^j(x)$. Then $a \le f^j(x)$, $f^j(1) = f^j(0) = 0$ and $0 < x < a$. Therefore $[0, a] \subseteq f^j([0, x])$ and $[0, a] \subseteq f^j([x, a])$, that is, f^j has a 2-horseshoe. By [19, Theorem A], $h(f^j) > 0$. By [2, Chapter 4], $h(f^j) = h(f)j$. Then $h(f) > 0$, which leads us to a contradiction.

Let $x \in \mathbb{X}_n$ and $0 < \varepsilon < \min\{|x|, 1 - |x|\}$. Denote by $x_{-\varepsilon}$ and x_ε the elements such that $x_{-\varepsilon} < x < x_\varepsilon$ and $|x - x_\varepsilon| = |x - x_{-\varepsilon}| = \varepsilon$.

Lemma 8 *Assume $f \in C_0(\mathbb{X}_n)$ has property $\mathcal{P}$ and $h(f) = 0$. Let $J \subset \mathbb{X}$ be an open interval such that $J \cap \omega(f) = \varnothing$. Then, for any $y \in J$ there is an interval J_y, $y \in J_y$, containing at most two points of each orbit.*

Proof For $y \in J \subset B_j$, $j \in \{1, \dots, n\}$, we distinguish three cases: $f(y) \notin \overline{B}_j$, $f(y) \in B_j$ and $f(y) = 0$.

First, assume that $f(y) \notin \overline{B}_j$. Let $(a, b) \subset B_j$ be such that $y \in (a, b)$, $f(1) = 0$ and $f(a, b) \cap B_j = \varnothing$. If $f^i(a, b) \prec (a, b)$ for all $i \in \mathbb{N}$, then the proof concludes. So, let $m \in \mathbb{N}$ be the first integer such that $f^m(a, b) \cap (a, b) \ne \varnothing$. Assume that if any positive integer i is big enough, then it is held $f^m(y_{-\varepsilon_i}, y_{\varepsilon_i}) \cap (y_{-\varepsilon_i}, y_{\varepsilon_i}) \ne \varnothing$ with $\varepsilon_i = 1/i$. Hence $\cap_i (f^m(y_{-\varepsilon_i}, y_{\varepsilon_i}) \cap (y_{-\varepsilon_i}, y_{\varepsilon_i})) = \{y\}$. Since f^m is continuous, we would have $f^m(y) = y$, which leads us to a contradiction. So there is $i \in \mathbb{N}$ such that $f^m(y_{-\varepsilon_i}, y_{\varepsilon_i}) \prec (y_{-\varepsilon_i}, y_{\varepsilon_i})$ (cf. Lemma 4) and $(y_{-\varepsilon_i}, y_{\varepsilon_i}) \subset (a, b)$. Now, we distinguish two cases. If $a = 0$ then by Lemma 4 $f^k(y_{-\varepsilon_i}, y_{\varepsilon_i}) \prec (y_{-\varepsilon_i}, y_{\varepsilon_i})$ for all $k \ge m$ and the proof concludes. If $a \ne 0$, then applying Lemmas 4 and 7, $f^k(y_{-\varepsilon_i}, y_{\varepsilon_i}) \prec (y_{-\varepsilon_i}, y_{\varepsilon_i})$ for all $k \ge m$, which finishes the proof.

Now, assume that $f(y) \in B_j$. Let $(a, b) \subset B_j$ be such that $y \in (a, b)$ and $f(a, b) \subset B_j$. Assume that any open subinterval J containing y contains at least three points of some orbit, that is, there is an $x \in \mathbb{X}$ and there are $n_1 < n_2 < n_3$ such that $f^{n_i}(x) \in J$, $1 \le i \le 3$. By Lemma 6 and [10, Proposition 11, Chapter 4], there is an interval J_y holding that for any $x \in \mathbb{X}_n$ with $f^{n_i}(x) \in J_y$, $1 \le i \le 3$, there is $k \in \mathbb{N}$, $n_1 < k < n_3$, such that $f^k(x) \notin B_j$. Then, $f^{k-1}(x) \in (c, d)$ with $f(c) = 0$ and $f(c, d) \cap B_j = \varnothing$.

By Lemma 7, $(c,d) \prec J_y$. Then, by Lemma 4, for all integer $m > k$ it holds that $f^m(x) \prec J_y$, a contradiction.

Finally, assume that $f(y) = 0$. Since $f(0) = 0$ and f is uniformly continuous, there are real numbers $\varepsilon_n > \cdots > \varepsilon_1 > 0$ such that

$$f(B(0,\varepsilon_j)) \subset B(0,\varepsilon_{j+1}) \quad \text{for} \quad j = 1, 2, \ldots, n-1. \tag{7}$$

Since $f(y) = 0$, there is $\delta > 0$ such that $f(y_{-\delta}, y_\delta) \subset B(0,\varepsilon_1)$. On the other hand, let $K = \max\{|f_j(z)|\colon z \in [0,y]\}$ and let $z_0 \in [0,y]$ be such that $|f(z_0)| = |f_j(z_0)| = K$. Clearly ε_n and δ can be chosen such that

$$(y_{-\delta}, y_\delta) \bigcap [0, f_j(z_0)] = \varnothing \quad \text{(cf. Lemma 7)} \tag{8}$$

and $(y_{-\delta}, y_\delta) \cap B(0,\varepsilon_n) = \varnothing$. Now, let $x \in (y_{-\delta}, y_\delta)$ and notice that $f(x) \in B(0,\varepsilon_1)$. If $f_j^i(x) = f^i(x)$ for all $i \in \mathbb{N}$ then, as $(y_{-\delta}, y_\delta) \cap [0, f_j(z_0)] = \varnothing$, we conclude that $f^i(x) \notin (y_{-\delta}, y_\delta)$ for all $i \in \mathbb{N}$ and we finish. So, let m be the first integer such that $f^m(x) \notin B_j$. If $m > 1$, and $k > m$ holds $f^k(x) \in \overline{B}_j$, then by Lemma 4, $f^k(x) < f(x)$ and hence $f^k(x) \in B(0,\varepsilon_1)$. This, jointly with (8) gives us $\{f^i(x)\colon i \in \mathbb{N}\} \cap (y_{-\delta}, y_\delta) = \varnothing$. To finish the proof, assume $m = 1$ and let k be the smallest integer such that $f^k(x) \in \overline{B}_j$ (if such k does not exist we finish). Let $l < n$ be the number of branches in which the set $\{f^i(x)\colon 1 \le i \le k\}$ lies. Notice that if an element $z \in B(0,\varepsilon_s) \cap \overline{B}_r$, $1 \le r, s \le n$, $f^i(z) \in \overline{B}_r$ for some $i \in \mathbb{N}$ and $f^{i+1}(z) \notin \overline{B}_r$, then by Lemmas 4 and 7, $f^i(z) \in B(0,\varepsilon_s)$. Then, by (7), $f^k(x) \in B(0,\varepsilon_l)$ and by Lemma 4 and (8) we conclude that $\{f^i(x)\colon i \in \mathbb{N}\} \cap (y_{-\delta}, y_\delta) = \varnothing$, which ends the proof.

The argument of the proof of the following result is very similar to the analogous result for interval continuous maps.

Corollary 9 *Assume $f \in C_0(\mathbb{X}_n)$ has property $\mathcal{P}$ and $h(f) = 0$. For any open set $U \supset \omega(f)$ there is a positive integer $q = q(U)$ such that at most q points of any trajectory lie outside U.*

Proof The set $\mathbb{X}_n \backslash U$ is compact. By Lemma 8, for any $y \in \mathbb{X}_n \backslash U$, there is an open interval J_y (relative to $\mathbb{X}_n \backslash U$) containing at most two points of any orbit. Since $\mathbb{X}_n \backslash U$ is a compact set we can obtain a finite number of such intervals covering $\mathbb{X}_n \backslash U$, which ends the proof.

Proposition 10 *Assume $f \in C_0(\mathbb{X}_n)$ has property $\mathcal{P}$. Then f is chaotic iff $f_i \in C(\overline{B}_i)$ is chaotic for some $i \in \{1, \ldots, n\}$.*

Proof First, assume that f_i is chaotic for some $i \in \{1, 2, \ldots, n\}$, and let $S \subset \overline{B}_i$ be a scrambled set. Notice that if $x \in S$ then $f_i^j(x) \neq 0$ for all $j \in \mathbb{N}$ (in other case $\omega(x,f) = \{0\}$ and $x \notin S$). Hence the trajectory of any $x \in S$ is contained in B_i which implies that the trajectories of x under f_i and f are the same. Then S is also a scrambled set for f.

Now, assume that f is chaotic and let $S \subset \mathbb{X}_n$ be an uncountable scrambled set of f. Let $x \in S$. By Lemma 5, the sequence $s(x)$ must be eventually constant, because in other case the orbit of x would be attracted by the fixed point 0. Let r be such that $s_j = s_r$ for all $j \ge r$, but $s_{r-1} \neq s_r$. On the other hand, let $y \in S$. Since $\liminf_{j\to\infty} d(f^j(x), f^j(y)) = 0$, we have that the trajectory of y is eventually contained in B_{s_r}. Let $[0,a] = \bigcap_{j \ge 0} (f_{s_r})^j(\overline{B}_{s_r})$. Since f_{s_r} is an interval map, by [9, Lemma 3.5], if $\{(f_{s_r})^j(f^r(x))\colon j\in\mathbb{N}\}\cap[0,a]$ is empty, then the trajectory of $f^r(x)$ will be attracted by a periodic orbit and then x cannot belong to any scrambled set. So, there is $j_x\geq r$ such that $f^{j_x}(x)\in[0,a]$. Since $f_{s_r}|_{[0,a]}$ is surjective, there is $x_0\in[0,a]$ such that $f^{j_x}(x_0)=f^{j_x}(x)$. Similarly, there are $y_0\in[0,a]$ and $j_y\geq r$ such that $f^{j_y}(y_0)=f^{j_y}(y)$. Let $S_0=\{y_0\in B_{s_e}\colon y\in S\}\subset[0,a]$. Then it is straightforward to see that S_0 is a scrambled set for f_{s_r} and therefore f_{s_r} is chaotic.

In order to finish the preparatory work to prove our main result, we prove the following lemma, which is an extension of a similar lemma from [14].

Lemma 11 *Assume $f\in C_0(\mathbb{X}_n)$ has property $\mathcal{P}$ and $h(f)=0$. Suppose f is non-chaotic. Then, for any $\varepsilon>0$ there are points $x_1,\ldots,x_k\in\omega(f)$, and a set $U\supset\omega(f)$, relatively open in $\mathbb{X}_n$, with the following property: if*

$$f^j(x)\in U \quad \text{for} \quad 0\leq j\leq r,$$

then there is some i such that for any j with $0\leq j\leq r$

$$d(f^j(x),f^j(x_i))<\varepsilon.$$

Proof Let f be non-chaotic. By Proposition 10, f_i is non-chaotic for $i=1,\ldots,n$. Then, by [12, Theorem 2.3], for $i=1,\ldots,n$ it holds that $f_i|_{\omega(f_i)}$ are Lyapunov stable (it has equicontinuous powers), and any point $y\in\omega(f_i)$ is almost periodic (for any neighborhood G of y there is an integer $m>0$ such that $f^{m\cdot j}(y)\in G$ for any $j\geq 0$). By Lemma 6 it is easy to see that

$$f|_{\omega(f)} \quad \text{is Lyapunov stable} \tag{9}$$

and

$$\text{every point in } \omega(f) \text{ is almost periodic.} \tag{10}$$

Then, using (9) and (10) and following the proof of the lemma from [14] we obtain the result.

4 Proof of Theorem 2

First, consider the case $h(f)=0$. Following the proof of Theorem 1.5 from [5] we see that f^N, $N=n!(n-1)!\ldots 2!1!$, holds property $\mathcal{P}$. Additionally, by [2, Chapter 4], $h(f^N)=Nh(f)=0$. So, by Proposition 3 we may assume without loss of generality that f has property $\mathcal{P}$.

First, assume f is non-chaotic and let A be a strictly increasing sequence of positive integers. Then, applying Lemma 11 and Corollary 9 and proceeding as in the first part of the proof of the main result of [14], we obtain that $h_A(f)=0$ for all A. Now assume that f is chaotic. By Proposition 10, f_i is chaotic for some $i\in\{1,\ldots,n\}$. Following [14], there is a interval $J\subset B_i$, with $f_i^{2^j}(J)=f^{2^j}(J)=J$ for some $j\in\mathbb{N}$ and such that $f_i^k(J)\subset B_i$ for $1\leq k\leq 2^j$. Additionally, $f_i^{2^j}|_J$ is chaotic and hence $h_A(f_i^{2^j}|_J)>0$. Then

$$h_{2^j\cdot A}(f)\geq h_{2^j\cdot A}(f,J)=h_{2^j\cdot A}(f_i,J)=h_A(f_i^{2^j}|_J)=h_A(f_i^{2^j},J)>0,$$

where $2^j \cdot A = (2^j \cdot a_i)_{i=1}^{\infty}$.

Finally, assume $h(f) > 0$. By [19], there is an $l \in \mathbb{N}$ such that f^l has a k-horseshoe. Since $h(f^l) = lh(f) > 0$, by Proposition 3 we may assume that $l = 1$. So, there is an interval J and k subintervals $J_1, \dots, J_k$ with pairwise disjoint interiors and such that $J \subseteq f(J_i)$. There is an invariant compact subset Y included in at most two branches such that $f|_Y$ is semiconjugate to a shift map defined on $\Sigma = \{(x_j)_{j=1}^{\infty} : \ x_j \in \{0,1\}\}$ (see e.g. [10, Chapter 2]). Then, it is straightforward to check that f is chaotic, and the proof concludes.

Corollary 12 *Let $f \in C_0(\mathbb{X}_n)$ be such that $0 \in Per(f)$. Then f is chaotic iff there is an increasing sequence of positive integers A such that $h_A(f) > 0$.*

Proof Just apply Proposition 3 and Theorem 2.

Acknowledgments

I wish to thank to professor L'ubomir Snoha for suggesting me the problem.

References

[1] Alsedá, Ll., Llibre, J. and Misiurewicz, M. Periodic orbits of maps of *Y*. *Trans. Amer. Math. Soc.* **313** (1989) 475–538.

[2] Alsedá, Ll., Llibre, J. and Misiurewicz, M. *Combinatorial Dynamics And Entropy in Dimension One.* World Scientific Publishing, Singapore, 1993.

[3] Alsedá,Ll. and Moreno, J.M. Lower bounds of the topological entropy of maps of Y. *J. Math. Anal. Appl.* **155** (1991) 513–530.

[4] Alsedá, Ll. and Moreno, J.M. Linear orderings and the full periodicity kernel for the n-star. *J. Math. Anal. Appl.* **180** (1993) 599–616.

[5] Alsedá, Ll. and Ye, X. Division for star maps with the branching point fixed. *Acta Math. Univ. Comenian.* **62** (1993) 237–248.

[6] Alsedá, Ll. and Ye, X. Minimal sets of maps of *Y*. *J. Math. Anal. Appl.* **187** (1994) 324–338.

[7] Aulbach, B. and Kieninger, B. On three definitions of chaos. *Nonlin. Dynamics and Systems Theory* **1** (2001) 23–38.

[8] Baldwin, S. An extension of Sarkovskii's Theorem to the n-od. *Ergod. Th. Dyman. Sys.* **11** (1991) 249–271.

[9] Balibrea, F., Cánovas Peña, J.S. and Jiménez López, V. Commutativity and non-commutativity of the topological sequence entropy. *Ann. Inst. Fourier (Grenoble)* **49** (1999) 1693–1709.

[10] Block, L.S. and Coppel, W.A. *Dynamics in One Dimension.* Lecture Notes in Math., Springer–Verlag, Berlin, 1992.

[11] Cánovas, J.S. Distributional chaos on tree maps: the star case. *Comment. Math. Univ. Carolinae* **42** (2001) 583–590.

[12] Fedorenko, V.V., Sharkovskii, A.N. and Smítal, J. Characterizations of weakly chaotic maps of the interval. *Proc. Amer. Math. Soc.* **110** (1990) 141–148.

[13] Forti, F., Paganoni, L. and Smítal, J. Strange triangular maps of the square. *Bull. Austral. Math. Soc.* **51** (1995) 395–415.

[14] Franzová, N. and Smítal, J. Positive sequence topological entropy characterizes chaotic maps. *Proc. Amer. Math. Soc.* **112** (1991) 1083–1086.

[15] Goodman, T.N.T. Topological sequence entropy. *Proc. London Math. Soc.* **29** (1974) 331–350.

[16] Hric, R. Topological sequence entropy for maps of the interval. *Proc. Amer. Math. Soc.* **127** (1999) 2045–2052.

[17] Hric, R. Topological sequence entropy for maps of the circle. *Comment. Math. Univ. Carolin.* **41** (2000) 53–59.

[18] Li, T.Y. and Yorke, J.A. Period three implies chaos. *Amer. Math. Monthly* **82** (1975) 985–992.

[19] Llibre, J. and Misiurewicz, M. Horseshoes, entropy and periods for graph maps. *Topology* **32** (1993) 649–664.

[20] Paganoni, L. and Santambrogio, P. Chaos and sequence topological entropy for triangular maps. *Grazer Math. Ber.* **339** (1999) 279–290.

[21] Sharkovsky, A.N., Kolyada, S.F., Sivak, A.G. and Fedorenko, V.V. *Dynamics of One-Dimensional Maps.* Kluwer Academic Publishers, Dordrecht, 1997.

[22] Schweizer, B. and Smítal, J. Measures of chaos and a spectral decomposition of dynamical systems on the interval. *Trans. Amer. Math. Soc.* **344** (1994) 737–754.

[23] Smítal, J. Chaotic functions with zero topological entropy. *Trans. Amer. Math. Soc.* **297** (1986) 269–282.

Nonlinear Dynamics and Systems Theory, **5**(1) (2005) 9–38

Global Stability Properties for a Class of Dissipative Phenomena via One or Several Liapunov Functionals

A. D'Anna and G. Fiore

Dipartimento di Matematica e Applicazioni, Facoltà di Ingegneria, Università di Napoli, V. Claudio 21, 80125 Napoli, Italy

Received: July 07, 2003; Revised: June 22, 2004

Abstract: We find some new results regarding the existence, uniqueness, boundedness, stability and attractivity of the solutions of a class of initial-boundary-value problems characterized by a quasi-linear third order equation which may have non-autonomous forcing terms. The class includes equations arising in superconductor theory, quantum mechanics and in the theory of viscoelastic materials.

Keywords: *Nonlinear higher order PDE; stability; boundedness; boundary value problems.*

Mathematics Subject Classification (2000): 35B35, 35G30.

1 Introduction

In this paper we deal with questions regarding the existence, uniqueness, boundedness, stability and attractivity of solutions u of the following class of initial-boundary-value problems:

$$Lu = f(x, t, u, u_x, u_{xx}, u_t), \quad 0 < x < 1, \quad 0 < t < T, \tag{1.1}$$

where $L = -\varepsilon\partial_{xxt} - c^2\partial_{xx} + \partial_{tt}$, f is a continuous function of its arguments, c and ε are positive constants, and

$$u(x, 0) = u_0(x), \quad u_t(x, 0) = u_1(x), \quad 0 < x < 1, \tag{1.2}$$

$$u(0, t) = h_1(t), \quad u(1, t) = h_2(t), \quad 0 < t < T, \tag{1.3}$$

where $T \leq +\infty$, $h_1, h_2 \in C^2([0,T[)$, $u_0, u_1 \in C^2([0,1])$ are assigned and fulfill the consistency condition

$$\begin{aligned} h_1(0) &= u_0(0), \qquad \frac{dh_1(0)}{dt} = u_1(0), \\ h_2(0) &= u_0(1), \qquad \frac{dh_2(0)}{dt} = u_1(1). \end{aligned} \tag{1.4}$$

Solutions u of such problems describe a number of physically remarkable continuous phenomena occurring on a finite space interval. In the operator L the D'Alembertian $-c^2\partial_{xx} + \partial_{tt}$ induces wave propagation, $-\varepsilon\partial_{xxt}$ dissipation. The term on the right-hand side of (1.1) may contain forcing terms, nonlinear (local) couplings of u to itself, further dissipative terms. For instance, when $f = -b\sin u - au_t + F(x,t)$, where a, b are positive constants, we deal with the perturbed Sine-Gordon equation, which can be used e.g. to describe the classical Josephson effect with driving force F in the theory of superconductors [6, 11]. F is a forcing term, $-au_t$ is a dissipative one and $-b\sin u$ is a nonlinear coupling. On the other hand it is well known [12] that equation (1.1) describes the evolution of the displacement $u(x,t)$ of the section of a rod from its rest position x in a Voigt material when an external force f is applied; in this case $c^2 = E/\rho$, $\varepsilon = 1/(\rho\mu)$, where ρ is the (constant) linear density of the rod at rest, and E, μ are respectively the elastic and viscous constants of the rod, which enter the stress-strain relation $\sigma = E\nu + \partial_t\nu/\mu$, where σ is the stress, ν is the strain. As we shall see in the sequel, even considering only one of these examples, e.g. the perturbed Sine-Gordon equation $f = -b\sin u - au_t$, it is important to keep room for a more general f because the latter will naturally appear when asking whether a particular solution u^* of the problem is stable or attractive, or when reducing the original problem to one with trivial boundary conditions.

Several papers [2−5, 7−9, 13] have already been devoted to the analysis of the operator L and more specifically to the investigation of the boundedness, stability and attractivity of the solutions of the above problem. Here we improve previous results, by weakening the assumptions on f, and find some new ones. In Section 2 we improve the existence and uniqueness Theorem 2.1 proved in [2], in that we require f to satisfy only *locally* Lipschitz condition. In Section 3.2 we improve the boundedness and stability Theorem 3.1 of the same reference, in that we require only a suitable *time average* of the quadratic norm of f to be bounded. While doing so we prove two lemmas concerning boundedness and attractivity of the zero solution for a class of first order ordinary differential equations in one unknown; the second lemma is a generalization of a lemma due to Hale [10]. In Sections 4 and 5 we respectively improve the exponential asymptotic stability Theorem 3.3 of [2] and the uniform asymptotic stability Theorem 2 of [5], valid for some special f, by removing the boundedness assumption on the latter. The trick we use is to associate to each neighbourhood of the origin with radius σ (the 'error') a Liapunov functional depending on a parameter γ adapted to σ, instead of fixing γ once and for all.

2 Existence and Uniqueness of the Solution

To discuss the existence and uniqueness of the above problem it is convenient to formulate it as an equivalent integro-differential equation so as to apply the fixed-point theorem.

As in [2], we start from the identity

$$\begin{aligned} \partial_\xi(c^2 u w_\xi - c^2 u_\xi w + \varepsilon u_\xi w_\tau - \varepsilon u w_{\xi\tau}) + \partial_\tau(u_\tau w - u w_\tau - \varepsilon u_{\xi\xi} w) \\ = f w - u(\varepsilon w_{\xi\xi\tau} - c^2 w_{\xi\xi} + w_{\tau\tau}), \end{aligned} \tag{2.1}$$

that follows from (1.1) for any smooth function $w(\xi,\tau)$, assuming $u(\xi,\tau)$ is a smooth solution of (1.1). We choose w as a function depending also on x, t and fulfilling the equation $Lw = 0$, more precisely

$$w(x,\xi,\, t-\tau) = \theta(x-\xi,\, t-\tau) - \theta(x+\xi,\, t-\tau), \tag{2.2}$$

with

$$\theta(x,t) = K(|x|,t) + \sum_{m=1}^{\infty} \left[K(|x+2m|,\, t) + K(|x-2m|,\, t)\right]. \tag{2.3}$$

The function K represents the fundamental solution of the linear equation $LK = 0$. It has been determined and studied in [3], and reads

$$K(|x|,t) = \int_0^t \frac{e^{-c^2\tau/\varepsilon}}{\sqrt{\pi\varepsilon\tau}}\, d\tau \int_0^\infty \frac{x^2(z+1)}{4\varepsilon\tau}\, e^{-x^2(z+1)^2/4\varepsilon\tau} I_0\left(\frac{c}{\varepsilon}\, 2|x|\sqrt{z}\right) dz, \tag{2.4}$$

where I_0 is the modified Bessel function of order zero. Since $\theta(-x,t) = \theta(x,t)$ and $\theta(x+2m,\, t) = \theta(x,t)$, $m \in N$, it is sufficient to restrict our attention to the domain $0 \le x < 2$, and note that θ is continuous together with its partial derivatives and satisfies the equation $L\theta = 0$. Moreover, from the analysis of K developed in [3], we can deduce that θ is a positive function that has properties similar to ones of the analogous function θ used for the heat operator, see [1].

As for the data we shall assume that:

$$\begin{gathered} f(x,t,n,p,q,r) \quad \text{is defined and continuous on the set} \\ \{(x,t,n,p,q,r)\colon\ 0 \le x \le 1,\ \ 0 \le t \le T,\ \ -\infty < n,p,q,r < \infty,\ \ T > 0\}, \end{gathered} \tag{2.5}$$

it locally satisfies a Lipschitz condition, namely for any bounded set $\Omega \subset [0,T] \times \mathbb{R}^4$ there exists a constant μ_Ω such that for any $(t,n_1,p_1,q_1,r_1), (t,n_2,p_2,q_2,r_2) \in \Omega$ and $x \in [0,1]$ (2.6)

$$\begin{aligned} &|f(x,t,n_1,p_1,q_1,r_1) - f(x,t,n_2,p_2,q_2,r_2)| \\ &\le \mu_\Omega\{|n_1-n_2| + |p_1-p_2| + |q_1-q_2| + |r_1-r_2|\}, \end{aligned}$$

$$u_0,\ u_0',\ u_0'',\ u_1 \quad \text{continuous on} \quad 0 \le x \le 1, \tag{2.7}$$

$$h_i,\ \frac{dh_i}{dt},\ \ i = 1,2, \quad \text{continuous on} \quad 0 \le t \le T, \tag{2.8}$$

$$h_1(0) = u_0(0), \quad h_2(0) = u_0(1), \quad \frac{dh_1(0)}{dt} = u_1(0), \quad \frac{dh_2(0)}{dt} = u_1(1). \tag{2.9}$$

Given a solution u of (1.1) − (1.3), by integrating (2.1) on $\{(\xi,\tau):\ 0<\xi<1,\ \delta<\tau<t-\delta\},\ \delta>0$, and letting $\delta\to 0$, we find that it satisfies the following integral equation

$$u(x,t)=\int_0^1 w_t(x,\xi,t)u_0(\xi)\,d\xi+\int_0^1 w(x,\xi,t)[u_1(\xi)-\varepsilon u_0''(\xi)]\,d\xi \tag{2.10}$$
$$-2\int_0^t h_1(\tau)(c^2+\varepsilon\partial_t)\theta_x(x,\,t-\tau)\,d\tau+2\int_0^t h_2(\tau)(c^2+\varepsilon\partial_t)\theta_x(1-x,\,t-\tau)\,d\tau$$
$$+\int_0^t d\tau\int_0^1 w(x,\xi,\,t-\tau)f(\xi,\tau,u(\xi,\tau),u_\xi(\xi,\tau),u_\tau(\xi,\tau),u_{\xi\xi}(\xi,\tau))\,d\xi.$$

Conversely, one can immediately verify that under the assumptions (2.5) − (2.9) a solution u of (2.10) satisfies (1.1) using the fact that $L\theta=0$ and $Lw=0$. We refer the reader to [2] for the slightly longer proof that the initial conditions (1.2) and the boundary conditions are satisfied.

If $f=f(x,t)$, (2.10) gives the unique explicit solution of (1.1) − (1.3). On the contrary, if f depends on u (2.10) is an integro-differential equation. We shall now discuss the existence and uniqueness of its solutions.

For any $c,d\in[0,T],\ c\le d$, we shall denote

$$\mathcal{B}_{[c,d]}:=\{u(x,t):\ u,u_x,u_t,u_{xx}\in C([0,1]\times[c,d])\}.$$

For any $a\in[0,T],\ v\in\mathcal{B}_{[0,a]}$ and $t\in[a,T]$ we define a mapping of $\mathcal{B}_{[a,T]}$ into itself by

$$\mathcal{T}_v u(x,t):=\omega_v(x,t)+\int_a^t d\tau\int_0^1 w(x,\xi,\,t-\tau)f(\xi,\tau,u(\xi,\tau),u_\xi(\xi,\tau),u_\tau(\xi,\tau),u_{\xi\xi}(\xi,\tau))\,d\xi \tag{2.11}$$

where

$$\omega_v(x,t)=\int_0^1 w_t(x,\xi,t)u_0(\xi)\,d\xi+\int_0^1 w(x,\xi,t)[u_1(\xi)-\varepsilon u_0^{''}(\xi)]\,d\xi$$
$$-2\int_0^t h_1(\tau)(c^2+\varepsilon\partial_t)\theta_x(x,\,t-\tau)\,d\tau+2\int_0^t h_2(\tau)(c^2+\varepsilon\partial_t)\theta_x(1-x,\,t-\tau)\,d\tau$$
$$+\int_0^a d\tau\int_0^1 w(x,\xi,\,t-\tau)f(\xi,\tau,v(\xi,\tau),v_\xi(\xi,\tau),v_\tau(\xi,\tau),v_{\xi\xi}(\xi,\tau))\,d\xi.$$

We fix a $\rho>0$ and for any $t\in[a,T]$ we consider the domain

$$S_{v,t}:=\{u\in\mathcal{B}_{[a,T]}:\ \forall\,x\in[0,1]\ \ |u(x,t)-\omega_v(x,t)|<\rho,\ \ |u_x(x,t)-\omega_{vx}(x,t)|<\rho,$$
$$|u_{xx}(x,t)-\omega_{vxx}(x,t)|<\rho,\ \ |u_t(x,t)-\omega_{vt}(x,t)|<\rho\}$$

and define

$$M = M(a,T,v,\rho) := \sup_{\substack{\tau\in[a,T]\\ \xi\in[0,1]}} \sup_{u\in S_{v,\tau}} |f(\xi,\tau,u(\xi,\tau),u_\xi(\xi,\tau),u_\tau(\xi,\tau),u_{\xi\xi}(\xi,\tau))|,$$

$$\begin{gathered} b-a = \min\left\{T-a,\ \frac{\rho}{M},\ \frac{c\rho}{M},\ \frac{\varepsilon\rho}{M},\ \sqrt{\frac{2\rho}{M}}\right\}, \\ R_{a,b,v} := \{u\in\mathcal{B}_{[a,b]}\colon\ \forall\,(x,t)\in[0,1]\times[a,b] \quad |u(x,t)-\omega_v(x,t)|\le\rho, \\ |u_x(x,t)-\omega_{vx}(x,t)|\le\rho,\quad |u_{xx}(x,t)-\omega_{vxx}(x,t)|\le\rho, \\ |u_t(x,t)-\omega_{vt}(x,t)|\le\rho\}. \end{gathered} \tag{2.12}$$

Note that by its definition M is finite because f is continuous and evaluated on a compact subset of R^6. We denote by $\mu=\mu(a,b,v,\rho)$ the constant μ_Ω of (2.6) corresponding to the choice

$$\begin{gathered}\Omega = \{(t,n,p,q,r)\colon\ \text{with}\ t\in[a,b],\ \text{and such that}\ \exists\,x\in[0,1],\ \exists\,u\in R_{a,b,v} \\ \text{such that}\ n=u(x,t),\ \ p=u_x(x,t),\ \ q=u_{xx}(x,t),\ \ r=u_t(x,t)\},\end{gathered}$$

we choose a positive constant λ

$$\lambda = \lambda(a,b,v,\rho) > \max\left\{1,\ \mu\left(2+\frac{1}{c}+\frac{1+2c^2}{\varepsilon}\right)\right\}$$

and we introduce a norm

$$\begin{aligned}\|u\|_{a,b} &:= \sup_{[0,1]\times[a,b]} |\mathrm{e}^{-\lambda t}u(x,t)| + \sup_{[0,1]\times[a,b]} |\mathrm{e}^{-\lambda t}u_x(x,t)| \\ &+ \sup_{[0,1]\times[a,b]} |\mathrm{e}^{-\lambda t}u_t(x,t)| + \sup_{[0,1]\times[a,b]} |\mathrm{e}^{-\lambda t}u_{xx}(x,t)|.\end{aligned} \tag{2.14}$$

We now show that $\mathcal{T}_v$ is a map of $R_{a,b,v}$ into itself, more precisely a contraction (w.r.t the above norm). From (2.11) we get for any $(x,t)\in[0,1]\times[a,b]$

$$|\mathcal{T}_v u(x,t)-\omega_v(x,t)| \le M(a,T,v,\rho)\int_a^t d\tau \int_0^1 |w(x,\xi,t-\tau)|\,d\xi,$$

and, because of the inequality [3]

$$\int_0^1 |w(x,\xi,t-\tau)|\,d\xi = \int_0^1 |\theta(x-\xi,\,t-\tau)-\theta(x+\xi,\,t-\tau)|\,d\xi \le t-\tau, \tag{2.15}$$

and the definition of b we find

$$|\mathcal{T}_v u(x,t)-\omega_v(x,t)| \le M(a,T,v,\rho)\,\frac{(b-a)^2}{2} \le \rho. \tag{2.16}$$

Similarly, one can prove that

$$|\mathcal{T}_v u_x(x,t) - \omega_{vx}(x,t)| \le \rho, \tag{2.17}$$
$$|\mathcal{T}_v u_{xx}(x,t) - \omega_{vxx}(x,t)| \le \rho, \tag{2.18}$$
$$|\mathcal{T}_v u_t(x,t) - \omega_{vt}(x,t)| \le \rho, \tag{2.19}$$

making use of the basic properties of K proved in [3], which lead to the following estimates:

$$\begin{aligned}
&\int_0^1 |w_x(x,\xi,t-\tau)|\, d\xi \le 1/c,\\
&\int_0^1 |w_t(x,\xi,t-\tau)|\, d\xi \le 1,\\
&\int_0^1 |w_{xx}(x,\xi,t-\tau)|\, d\xi \le \frac{1}{\epsilon}\,[1+2c^2(t-\tau)].
\end{aligned} \tag{2.20}$$

The first two inequalities were already given in [2], together with

$$\int_0^1 |(\partial_t - \partial_x^2) w(x,\xi,t-\tau)|\, d\xi \le 1. \tag{2.21}$$

The third was used but not explicitly written, and easily follows from the latter inequality, the equation $L\theta = 0$, and the relation $\theta(x,0) = 0$. In fact, from $L\theta = 0$ it immediately follows that

$$\theta_t - \theta_{xx} = \partial_t\Big[\theta + \frac{\epsilon}{c^2}\theta_{xx} - \frac{1}{c^2}\,\theta_t\Big],$$

and therefore

$$w_t(x,\xi,t-\tau) - w_{xx}(x,\xi,t-\tau) = \partial_t\Big[w(x,\xi,t-\tau) + \frac{\epsilon}{c^2}\, w_{xx}(x,\xi,t-\tau) - \frac{1}{c^2}\, w_t(x,\xi,t-\tau)\Big].$$

Integrating over ξ and using (2.21) we find $|\partial_t A(x,t-\tau)| \le 1$, where

$$A(x,\, t-\tau) := \int_0^1 \Big[w(x,\xi,\, t-\tau) + \frac{\epsilon}{c^2}\, w_{xx}(x,\xi,\, t-\tau) - \frac{1}{c^2}\, w_t(x,\xi,\, t-\tau)\Big] d\xi.$$

As $\theta(x,0) = 0$, then $A(x,0) = 0$. By the comparison principle we therefore find

$$\tau - t \le A(x,\, t-\tau) = \int_0^1 w\, d\xi + \int_0^1 \frac{\epsilon}{c^2}\, w_{xx}\, d\xi - \int_0^1 \frac{1}{c^2}\, w_t\, d\xi \le t-\tau,$$

implying

$$\left|\int_0^1 \frac{\epsilon}{c^2}\, w_{xx}\right| \le (t-\tau) + \left|\int_0^1 w\, d\xi\right| + \left|\int_0^1 \frac{1}{c^2}\, w_t\, d\xi\right|;$$

using (2.15) and $(2.20)_2$ to bound the integrals on the right hand-side we find $(2.20)_3$.

From the above results we conclude that $\mathcal{T}_v u(x,t) \in R_{a,b,v}$ as claimed.

From (2.11), (2.15)) we get for $t \in [a,b]$

$$\begin{aligned} |\mathcal{T}_v u_1(x,t) - \mathcal{T}_v u_2(x,t)|\mathrm{e}^{-\lambda t} &\le \mu \|u_1 - u_2\|_{a,b} \int_a^t \mathrm{e}^{-\lambda(t-\tau)}\,d\tau \int_0^1 |w(x,\xi,t-\tau)|\,d\xi \\ &\le \mu \|u_1 - u_2\|_{a,b} \int_a^t \mathrm{e}^{-\lambda(t-\tau)}(t-\tau)\,d\tau \le \frac{\mu}{\lambda^2}\|u_1 - u_2\|_{a,b}. \end{aligned} \tag{2.22}$$

From (2.11), (2.20) one can get analogous results for the partial derivatives ∂_x, ∂_t, ∂_x^2 of (2.11):

$$\begin{aligned} |\mathcal{T}_v u_{1x}(x,t) - \mathcal{T}_v u_{2x}(x,t)|\,\mathrm{e}^{-\lambda t} &\le \frac{\mu}{\lambda c}\,\|u_1 - u_2\|_{a,b}, \\ |\mathcal{T}_v u_{1t}(x,t) - \mathcal{T}_v u_{2t}(x,t)|\,\mathrm{e}^{-\lambda t} &\le \frac{\mu}{\lambda}\,\|u_1 - u_2\|_{a,b}, \\ |\mathcal{T}_v u_{1xx}(x,t) - \mathcal{T}_v u_{2xx}(x,t)|\,\mathrm{e}^{-\lambda t} &\le \frac{\mu}{\lambda\varepsilon}\left(1 + \frac{2c^2}{\lambda}\right)\|u_1 - u_2\|_{a,b}. \end{aligned} \tag{2.23}$$

Thus, we obtain

$$\|\mathcal{T}_v u_1(x,t) - \mathcal{T}_v u_2(x,t)\|_{a,b} \le \frac{\mu}{\lambda}\left[\frac{1}{\lambda} + \frac{1}{c} + 1 + \frac{1}{\varepsilon} + \frac{2c^2}{\varepsilon\lambda}\right]\|u_1 - u_2\|_{a,b}, \tag{2.24}$$

with $\mu \equiv \mu(a,b,v,\rho)$, $\lambda \equiv \lambda(a,b,v,\rho)$. Under assumption (2.13), inequality (2.24) shows that $\mathcal{T}_v$ is a contraction of $R_{a,b,v}$ into itself. Thus we are in the conditions to apply the fixed point theorem, and we find that there exists a unique solution in $R_{a,b,v}$ of the problem $\mathcal{T}_v u = u$ in the time interval $[a,b]$.

We now apply the above result iteratively. We start by choosing $a = 0 = a_0$, $v = 0$; the last integral disappears from (2.12). From the definition of b we determine the corresponding $b = a_1$ and by the fixed point theorem a unique solution $u^{(1)}(x,t)$ of the problem (1.1)–(1.4) in the time interval $[0,a_1]$. Next we choose $a = a_1$, $v = u^{(1)}$; from (2.12) we determine the corresponding $b = a_2$ and by the fixed point theorem a unique solution of the problem $\mathcal{T}_{u^{(1)}} u = u$ in the time interval $[a_1,a_2]$. This is also a smooth continuation of $u^{(1)}$, therefore we have found a unique solution $u^{(2)}(x,t)$ of the problem (1.1)–(1.4) in the time interval $[0,a_2]$, and so on. We conclude by stating the following

Theorem *Under hypotheses (2.5)–(2.9), the quasilinear problem (1.1)–(1.3) has a unique smooth solution in the time interval* $[0,a_\infty]$, *where*

$$a_\infty := \lim_{k\to+\infty} a_k \le T.$$

3 Eventual Boundedness and Asymptotic Stability

3.1 Preliminaries

By the rescaling $t \to t/c$, $\varepsilon \to c\varepsilon$ and of $f \to c^2 f$ we can factor c out of (1.1), so that it completely disappears from the problem, without loosing generality. In the sequel we shall assume we have done this. Moreover, without loss of generality we can also consider $h_1(t) = h_2(t) \equiv 0$ in (1.3), as any problem (1.1)−(1.4) is equivalent to another one of the same kind with trivial boundary conditions and a different f. In fact, setting for any $t \in J = [0, \infty[$

$$v(x,t) := u(x,t) + p(x,t), \qquad p(x,t) := (1-x)h_1(t) + xh_2(t)$$

we immediately find that $v(0,t) = v(1,t) \equiv 0$, that the initial condition for v, v_t are completely determined and that v fulfills the equation

$$-\varepsilon v_{xxt} + v_{tt} - v_{xx} = \tilde{f}(x,t,v,v_x,v_{xx},v_t),$$

where

$$\tilde{f}(x,t,v,v_x,v_{xx},v_t) := f(x,\, t,\, v-p,\, v_x - h_2 + h_1,\, v_{xx},\, v_t - p_t) - p_{tt}.$$

The difference $u = \tilde{u} - u^*$ between a generic solution $\tilde{u}$ and a given one u^* of the problem (1.1)−(1.4) is also a solution of a new problem of the same kind, which we denote by problem $\mathcal{P}$, but with $h_1(t) \equiv h_2(t) \equiv 0$, namely

$$\begin{gathered} -\varepsilon u_{xxt} + u_{tt} - u_{xx} = f(x,t,u,u_x,u_{xx},u_t), \quad x \in]0,1[, \quad t > t_0 \in J, \\ u(0,t) = 0, \quad u(1,t) = 0, \quad t \in J, \end{gathered} \tag{3.1}$$

with the initial conditions

$$u(x,0) = u_0(x), \quad u_t(x,0) = u_1(x), \quad x \in]0,1[, \tag{3.2}$$

fulfilling the consistency conditions

$$u_0(0) = u_1(0) = u_0(1) = u_1(1) = 0. \tag{3.3}$$

Here

$$\begin{aligned} f(x,t,u,u_x,u_{xx},u_t) = {} & f(x,t,\, u+u^*,\, u_x+u^*_x,\, u_{xx}+u^*_{xx},\, u_t+u^*_{xx}) \\ & - f(x,t,u^*,u^*_x,u^*_{xx},u^*_t) \end{aligned}$$

and $u_0(x) := \tilde{u}_0(x) - u^*_0(x)$, $u_1(x) = \tilde{u}_1(x) - u^*_1(x)$. The two solutions $\tilde{u}$, u^* are 'close' to each other iff u is a 'small' solution of the latter problem, and coincide iff u is the zero solution.

We introduce the distance between the origin O and a nonzero element $\big(u(\cdot,t), u_t(\cdot,t)\big) \in \Gamma := \big(C_0([0,1]) \cap C^2([0,1])\big) \times C_0([0,1])$ as the functional $d(u,u_t)$, where for any $(\varphi,\psi) \in \Gamma$ we define

$$d^2(\varphi,\psi) = \int_0^1 (\varphi^2 + \varphi_x^2 + \varphi_{xx}^2 + \psi^2)\, dx. \tag{3.4}$$

The notions of (eventual) boundedness, stability, attractivity, etc. are formulated using this distance. Imposing the condition that φ, ψ vanish in $0,1$ one easily derives that $|\varphi(x)|, |\varphi_x(x)| \leq d(\varphi,\psi)$ for any x; therefore a convergence in the norm d implies also a uniform pointwise convergence of φ, φ_x.

Definition 3.1 The *solutions* of (3.1)–(3.3) are *eventually uniformly bounded* if for any $\alpha > 0$ there exist an $s(\alpha) \geq 0$ and a $\beta(\alpha) > 0$ such that if $t_0 \geq s(\alpha)$, $d(u_0, u_1) \leq \alpha$, then $d(u(t), u_t(t)) < \beta(\alpha)$ for all $t \geq t_0$. If $s(\alpha) = 0$ the solutions of (3.1) are uniformly bounded.

Definition 3.2 The *origin* O of Γ is *eventually quasi-uniform-asymptotically stable in the large* for the solutions of (3.1) if for any $\rho, \alpha > 0$ there exist $s(\alpha) \geq 0$, and $\widehat{T}(\rho, \alpha) > 0$ such that if $d(u_0, u_1) \leq \alpha$, $t_0 \geq s(\alpha)$ then $d(u, u_t) < \rho$ for any $t \geq t_0 + \widehat{T}$. If $s(\alpha) = 0$, $u(x,t) \equiv 0$ is said to be quasi-uniform-asymptotically stable in the large for the solutions of (3.1).

Suppose now that problem $\mathcal{P}$ admits the solution $u(x,t) \equiv 0$.

Definition 3.3 The *solution* $u(x,t) \equiv 0$ is *uniform-asymptotical stable in the large* if it is uniformly stable and quasi-uniform-asymptotically stable in the large, and the solutions of problem $\mathcal{P}$ are uniformly bounded.

Definition 3.4 The *solution* $u(x,t) \equiv 0$ of the problem $\mathcal{P}$ is *exponential-asymptotically stable in the large* if for any $\alpha > 0$ there are two positive constants $D(\alpha)$, $C(\alpha)$ such that if $d(u_0, u_1) \leq \alpha$, then

$$d(u(t), u_t(t)) \leq D(\alpha) \exp\left[-C(\alpha)(t - t_0)\right] d(u_0, u_1), \quad \forall\, t \geq t_0. \tag{3.5}$$

To prove our theorems we shall use the Liapunov direct method. We introduce the Liapunov functional

$$V(\varphi, \psi) = \frac{1}{2} \int_0^1 \{(\varepsilon\varphi_{xx} - \psi)^2 + \gamma\psi^2 + (1+\gamma)\varphi_x^2\}\, dx, \tag{3.6}$$

where γ is an arbitrary positive constant. Using the inequality $|2\varepsilon\varphi_{xx}\psi| \leq \varepsilon(\varphi_{xx}^2 + \psi^2)$ we find

$$V \leq \frac{1}{2} \int_0^1 \{\varepsilon^2\varphi_{xx}^2 + \psi^2 + \varepsilon\varphi_{xx}^2 + \varepsilon\psi^2 + \gamma\psi^2 + (1+\gamma)\varphi_x^2\}\, dx.$$

Setting

$$c_2^2 = \max\{\varepsilon(1+\varepsilon)/2,\ (1+\varepsilon+\gamma)/2\}, \tag{3.7}$$

we thus derive

$$V(\varphi, \psi) \leq c_2^2 d^2(\varphi, \psi). \tag{3.8}$$

Moreover, it is known that [13]

$$\varphi(0) = 0, \quad \varphi(1) = 0 \quad \Longrightarrow \quad \begin{cases} \int_0^1 \varphi_x^2(x)\, dx \geq \pi^2 \int_0^1 \varphi^2(x)\, dx \\ \int_0^1 \varphi_{xx}^2(x)\, dx \geq \pi^2 \int_0^1 \varphi_x^2(x)\, dx \end{cases} \tag{3.9}$$

(these inequalities can be easily proved by Fourier analysis of φ). In view of the bounds we shall consider below we introduce the notation

$$\omega_1 := \frac{\pi^4}{1+\pi^4} \approx 0.99, \quad \omega_2 := \frac{\pi^4}{1+\pi^2+\pi^4} \approx 0.90, \quad \omega_3 := \frac{\pi^2}{1+\pi^2} \approx 0.91. \tag{3.10}$$

Using (3.9) and an argument employed in [2], we get

$$V(\varphi,\psi) \geq c_1^2 d^2(\varphi,\psi). \tag{3.11}$$

where

$$c_1^2 = \min\left\{\frac{\varepsilon^2}{8}\,\omega_1,\ \frac{1}{2}\left(\gamma-\frac{1}{2}\right)\right\}, \qquad (\gamma > 1/2). \tag{3.12}$$

Therefore, from (3.8) and (3.11) we find

$$\frac{V}{c_2^2} \leq d^2 \leq \frac{V}{c_1^2}. \tag{3.13}$$

On the other hand, choosing $\gamma = 1$ in (3.6) and reasoning as it has been done in [2] it turns out that

$$\begin{aligned}
\frac{dV}{dt} &= \int_0^1 \left\{ -\frac{\varepsilon}{2}\,u_{xx}^2 - \varepsilon u_{xt}^2 + \frac{\varepsilon}{2}\,u_t^2 - \frac{\varepsilon}{2}(u_{xx}+f)^2 - \frac{\varepsilon}{2}(u_t - 2f/\varepsilon)^2 + Af^2 \right\} dx \\
&\leq -\int_0^1 \left\{ \frac{\varepsilon}{2}\omega_2(u^2+u_x^2+u_{xx}^2) + \varepsilon\left(\pi^2-\frac{1}{2}\right)u_t^2 + Af^2 \right\} dx \\
&\leq -c_3^2 d^2(u,u_t) + \int_0^1 Af^2 dx
\end{aligned} \tag{3.14}$$

where we have set

$$A := \frac{\varepsilon}{2} + \frac{2}{\varepsilon}, \qquad c_3^2 := \frac{\omega_2}{2}\,\varepsilon, \tag{3.15}$$

and we have used (3.9). In the sequel we shall set also $p := c_3^2/c_2^2$.

3.2 Eventual boundedness and asymptotic stability

We assume that

$$A\int_0^1 f^2\,dx \leq g(t)c_1^2 d^2 + \tilde{g}_1(t,d^2) + \tilde{g}_2(t,d^2), \tag{3.16}$$

where f is the function appearing in (3.1), and $g(t)$, $\tilde{g}_i(t,\eta)$ $(i = 1,2$ and $t \in J,\ \eta > 0)$ denote suitable nonnegative continuous functions fulfilling the following conditions:

(1) there exists a constant $\sigma > 0$ such that for any $t \geq t_0 \geq 0$

$$\int_{t_0}^t g(\tau)\,d\tau - p(t-t_0) \leq \sigma; \tag{3.17}$$

(2) there exist constants $\chi \in [0,1]$, $\kappa \in [0,1]$ and $q \geq 0$ (with $q < p$ if $\chi = 1$) and $M > 0$ such that

$$\left| \frac{\int_0^t g(\tau)d\tau}{1+t^\chi} - q \right| < \frac{M}{1+t^\kappa}; \tag{3.18}$$

(3) for any $\eta > 0$

$$\lim_{t\to+\infty} \tilde{g}_1(t,\eta)e^{\xi(t^\chi - t^\kappa)} = 0,$$
$$\int_0^\infty \tilde{g}_2(\tau,\eta)e^{\xi(\tau^\chi-\tau^\kappa)}\,d\tau = \sigma_2(\eta) < +\infty, \tag{3.19}$$

where ξ is some positive constant if $\chi > \kappa$, while $\xi = 0$ if $\chi \le \kappa$.

Without loss of generality we can assume that $\tilde{g}_i(t,\eta)$ are non-decreasing in η; if originally this is not the case, we just need to replace $\tilde{g}_i(t,\eta)$ by $\max\limits_{0\le u\le\eta} \tilde{g}_i(t,u)$ to fulfill this condition.

From (3.14), using (3.4), (3.16), (3.13) we now find

$$\begin{aligned} \frac{dV(u,u_t)}{dt} &\le -c_3 d^2(u,u_t) + g(t)c_1^2 d^2 + \tilde{g}_1(t,d^2) + \tilde{g}_2(t,d^2) \\ &\le -p\,V + g(t)V + g_1(t,V) + g_2(t,V), \end{aligned} \tag{3.20}$$

where we have set

$$g_i(t,\eta) = \tilde{g}_i\left(t, \frac{\eta}{c_1^2}\right). \tag{3.21}$$

By the “Comparison Principle” (see e.g. [14]) V is bounded from above

$$V(t) \le y(t), \tag{3.22}$$

by the solution $y(t)$ of the Cauchy problem

$$\frac{dy}{dt} = -p\,y + g(t)y + g_1(t,y) + g_2(t,y), \qquad y(t_0) = y_0 \equiv V(t_0) \ge 0. \tag{3.23}$$

We therefore study the latter, proving first a theorem of eventual uniform boundedness.

Lemma 1 *Assume $g(t)$, $\tilde{g}_i(t,\eta)$ $(i = 1,2$ and $t \in J$, $\eta > 0)$ are continuous nonnegative functions fulfilling the conditions (3.17)–(3.19). Then $\forall\,\tilde{\alpha} > 0$ there exist $\tilde{s}(\tilde{\alpha}) \ge 0$, $\tilde{\beta}(\tilde{\alpha}) > 0$ such that if $|y_0| \le \tilde{\alpha}$, $t_0 \ge \tilde{s}(\tilde{\alpha})$, the modulus of the solution $y(t;t_0,y_0)$ of (3.23) is bounded by $\tilde{\beta}$:*

$$|y(t;t_0,y_0)| < \tilde{\beta}, \quad t \ge t_0 \ge \tilde{s}(\tilde{\alpha}); \tag{3.24}$$

if in particular $y_0 \in [0,\tilde{\alpha}]$, then

$$0 \le y(t;t_0,y_0) < \tilde{\beta}, \quad t \ge t_0 \ge \tilde{s}(\tilde{\alpha}). \tag{3.25}$$

Proof Problem (3.23) is equivalent to the integral equation

$$\begin{aligned} y(t) &= y_0\, e^{-p(t-t_0)+\int_{t_0}^{t} g(\tau)d\tau} \\ &\quad + e^{-pt+\int_0^t g(\tau)d\tau} \int_{t_0}^{t} \left[g_1\left(\tau,y(\tau)\right) + g_2\left(\tau,y(\tau)\right)\right] e^{p\tau-\int_0^\tau g(z)dz}\, d\tau. \end{aligned} \tag{3.26}$$

Take $\tilde{\beta}(\tilde{\alpha}) := \tilde{\alpha}(e^{\sigma} + \frac{e^{2M}}{m} + e^{2M})$, where

$$m = \begin{cases} \dfrac{p}{2} & \text{if } \chi < 1 \\ \dfrac{p-q}{2} & \text{if } \chi = 1. \end{cases} \tag{3.27}$$

Because of (3.17), if $|y_0| \le \tilde{\alpha}$, for any $t \ge t_0$ one finds

$$|y_0| e^{-p(t-t_0)+\int\limits_{t_0}^{t} g(\tau)d\tau} \le \tilde{\alpha} e^{\sigma}. \tag{3.28}$$

Moreover, because of (3.18), we obtain

$$q(1+t^{\chi}) - M\,\frac{1+t^{\chi}}{1+t^{\kappa}} < \int\limits_0^t g(z)\,dz < q(1+t^{\chi}) + M\,\frac{1+t^{\chi}}{1+t^{\kappa}}. \tag{3.29}$$

Let

$$\vartheta := \begin{cases} 0 & \text{if } \chi \le \kappa, \\ \min\left\{1,\ \dfrac{\xi}{2M}\right\} & \text{if } 1 > \chi > \kappa \\ \min\left\{1,\ \dfrac{p-q}{2M},\ \dfrac{\xi}{2M}\right\} & \text{if } 1 = \chi > \kappa, \end{cases}$$

$$t_{\vartheta} := \begin{cases} 0 & \text{if } \vartheta = 0, \\ \left(\dfrac{1-\vartheta}{\vartheta}\right)^{1/\kappa} & \text{if } \vartheta > 0; \end{cases}$$

considering separately the cases $\chi \le \kappa$, $\chi > \kappa$ and recalling the definition of ξ, we find

$$\frac{1+t^{\chi}}{1+t^{\kappa}} = 1 + \frac{t^{\chi}-t^{\kappa}}{1+t^{\kappa}} \le 1 + \vartheta(t^{\chi}-t^{\kappa})$$

for any $t \ge t_{\vartheta}$. Then from (3.29)

$$q(1+t^{\chi}) - M[1+\vartheta(t^{\chi}-t^{\kappa})] < \int\limits_0^t g(z)\,dz < q(1+t^{\chi}) + M[1+\vartheta(t^{\chi}-t^{\kappa})] \tag{3.30}$$

for any $t \ge t_{\vartheta}$. Consequently, for $i = 1,2$ and $|y| \le \tilde{\beta}$

$$\begin{aligned} & e^{-pt+\int\limits_0^t g(\tau)d\tau} \int\limits_{t_0}^{t} g_i(\tau,y) e^{p\tau - \int\limits_0^{\tau} g(z)dz} d\tau \\ & < e^{-pt+q(1+t^{\chi})+M[1+\vartheta(t^{\chi}-t^{\kappa})]} \int\limits_{t_0}^{t} g_i(t,\tilde{\beta}) e^{p\tau - q(1+\tau^{\chi})+M[1+\vartheta(\tau^{\chi}-\tau^{\kappa})]}\, d\tau \\ & = e^{qt^{\chi}+M\vartheta(t^{\chi}-t^{\kappa})-pt} e^{2M} \int\limits_{t_0}^{t} g_i(t,\tilde{\beta}) e^{p\tau - q\tau^{\chi}+M\vartheta(\tau^{\chi}-\tau^{\kappa})} d\tau, \end{aligned} \tag{3.31}$$

where we have used also the fact that $g_i(t,\eta)$ are non-decreasing functions of η.

Now consider the function

$$h(\tau) := p\tau - q\tau^{\chi} - M\vartheta(\tau^{\chi} - \tau^{\kappa}) \tag{3.32}$$

and its derivative $h'(\tau) = p - q\chi\tau^{\chi-1} - M\vartheta(\chi\tau^{\chi-1} - \kappa\tau^{\kappa-1})$. We now show that, for any $\chi \in [0,1]$,

$$h'(\tau) \ge h'(\tilde t) = m \quad \text{if} \quad \tau \ge \tilde t := \left[\frac{\chi(2q+\xi)}{p}\right]^{\frac{1}{1-\chi}} \tag{3.33}$$

with the m defined in (3.27) (this implies that for $\tau \ge \tilde t$ the function $h(\tau)$ is increasing). In fact, if $\vartheta > 0$, then it is either $0 \le \kappa < \chi < 1$, implying

$$h'(\tau) > p - (q + M\vartheta)\chi\tau^{\chi-1} \ge \frac{p}{2} = m$$

for any $\tau \ge \tilde t$, or $0 \le \kappa < \chi = 1$, implying (because of the inequality $p - q > 0$ and the definition of ϑ)

$$h'(\tau) = p - q - M\vartheta + M\vartheta\kappa\tau^{\kappa-1} > p - q - M\vartheta \ge \frac{p-q}{2} = m$$

for any $\tau > 0$, in particular for $\tau \ge \tilde t$. If $\vartheta = 0$, then it is either $0 \le \chi \le \kappa \le 1$ with $\chi < 1$, implying

$$h'(\tau) > p - q\chi\tau^{\chi-1} \ge \frac{p}{2} = m$$

for any $\tau \ge \left[\dfrac{2q\chi}{p}\right]^{\frac{1}{1-\chi}} \equiv \tilde t$, or $\chi = \kappa = 1$, implying also $h'(\tau) = p - q > m$ (for any τ), as claimed.

On the other hand, because of (3.19) there exist $s_1(\tilde\alpha), s_2(\tilde\alpha) \ge 0$ (recall that $\tilde\beta = \tilde\beta(\tilde\alpha)$) such that

$$\begin{gathered} g_1(\tau,\tilde\beta)e^{\xi(\tau^{\chi}-\tau^{\kappa})} \le \tilde\alpha \quad \text{if} \quad \tau \ge t_0 \ge s_1(\tilde\alpha), \\ \int_{t_0}^{\infty} g_2(\tau,\tilde\beta)e^{\xi(\tau^{\chi}-\tau^{\kappa})}\,d\tau \le \tilde\alpha \quad \text{if} \quad t_0 \ge s_2(\tilde\alpha). \end{gathered} \tag{3.34}$$

Hence, for $t \ge t_0 \ge \max\{\tilde t, t_\vartheta, s_1(\tilde\alpha)\}$ we find that if $|\,y(\tau)| \le \tilde\beta$ for any $\tau \in [t_0, t[$ then

$$\begin{aligned} & e^{-pt+\int_0^t g(\tau)d\tau} \int_{t_0}^{t} g_1\big(\tau, y(\tau)\big) e^{p\tau - \int_0^\tau g(z)dz} d\tau \\ & < e^{-h(t)+2M} \int_{t_0}^{t} g_1\big(\tau,\tilde\beta\big) e^{h(\tau)+\xi(\tau^{\chi}-\tau^{\kappa})} d\tau \\ & \le e^{-h(t)+2M}\tilde\alpha \int_{t_0}^{t} \frac{h'(\tau)}{m}\, e^{h(\tau)} d\tau = \tilde\alpha\, \frac{e^{-h(t)+2M}}{m}\,(e^{h(t)} - e^{h(t_0)}) < \frac{e^{2M}}{m}\,\tilde\alpha \end{aligned} \tag{3.35}$$

where in the first line we have used (3.31) and the definition of ϑ, in the second (3.33) and $(3.34)_1$. Similarly, for $t \geq t_0 \geq \max\{s_2(\tilde{\alpha}), t_\vartheta, \tilde{t}\}$ we find that if $|y(\tau)| \leq \tilde{\beta}$ for any $\tau \in [t_0, t[$ then

$$\begin{aligned} & e^{-pt+\int\limits_0^t g(\tau)d\tau} \int\limits_{t_0}^{t} g_2\big(\tau, y(\tau)\big) e^{p\tau - \int\limits_0^\tau g(z)dz} d\tau \\ & < e^{-h(t)+2M} \int\limits_{t_0}^{t} g_2\big(\tau, \tilde{\beta}\big) e^{h(\tau)+\xi(\tau^\chi - \tau^\kappa)} d\tau \qquad (3.36) \\ & < e^{-h(t)+h(t)+2M} \int\limits_{t_0}^{\infty} g_2(\tau, \tilde{\beta}) e^{\xi(\tau^\chi - \tau^\kappa)} d\tau < \tilde{\alpha} e^{+2M}, \end{aligned}$$

(in the first inequality we have used (3.31) and again the definition of ϑ, in the second the monotonicities of h and g_2, in the third $(3.34)_2$). Summarizing, the inequalities (3.28), (3.35), (3.36) are fulfilled for $t \geq t_0 \geq \tilde{s}(\tilde{\alpha}) = \max\{\tilde{t}, t_\vartheta, s_1(\tilde{\alpha}), s_2(\tilde{\alpha}), \}$.

Now let us suppose *per absurdum* that there exists $t_1 > t_0 \geq \tilde{s}(\tilde{\alpha})$ such that

$$|y(t; t_0, y_0)| < \tilde{\beta} \quad \text{for} \quad t_0 \leq t < t_1, \qquad (3.37)$$

$$|y(t_1; t_0, y_0)| = \tilde{\beta}. \qquad (3.38)$$

Because of (3.37) the inequalities (3.35), (3.36) are fulfilled; together with equations (3.26), (3.28) for $t = t_1$ they imply

$$|y(t_1; t_0, y_0)| < \tilde{\beta},$$

against the assumption (3.38). Finally, from (3.26) and the nonnegativity of the functions g_i we find that $0 \leq y_0 < \tilde{\alpha}$ implies $y(t) > 0$ for any t, whence (3.25).

As a result of the previous lemma, for any $\tilde{\alpha} > 0$ the solution $y(t)$ of the Cauchy problem (3.23) remains eventually uniformly bounded by $\tilde{\beta}(\tilde{\alpha})$ if $0 \leq y_0 \leq \tilde{\alpha}$. By (3.22) and (3.13), the same applies with $V(t)$ and $d^2(u, u_t)$.

By the monotonicity of $g_i(t, \eta)$ in η and the comparison principle we find that $y(t)$ is also bounded

$$y(t) \leq z(t), \quad t \geq t_0 \qquad (3.39)$$

by the solution $z(t)$ of the Cauchy problem

$$\frac{dz}{dt} = -p\, z + g(t)z + g_1(t, \tilde{\beta}) + g_2(t, \tilde{\beta}), \qquad z(t_0) = z_0 \qquad (3.40)$$

(which differs from (3.23) in that the second argument of g_i is now a fixed constant $\tilde{\beta} > 0$), provided that $z_0 = y_0$, and $t_0 \geq \tilde{s}(\tilde{\alpha})$.

We therefore study the Cauchy problem (3.40), keeping in mind that for our final purposes we will choose $\tilde{\beta} = \tilde{\beta}(\tilde{\alpha})$, $t_0 = t_0(\tilde{\alpha}) \geq \tilde{s}(\tilde{\alpha})$.

Lemma 2 *Assume $g(t)$, $\tilde g_i(t,\eta)$ $(i=1,2$ and $t\in J,\ \eta>0)$ are continuous functions fulfilling the conditions (3.17) – (3.19). Then for any $\tilde\rho>0,\ t_0>0,\ \tilde\alpha>0$ there exists $\widehat T(\tilde\rho,\tilde\alpha,\tilde\beta,t_0)>0$ such that for $|z_0|\le\tilde\alpha\ \in[0,\tilde\alpha]$ the solution $z(t;t_0,z_0)$ of (3.40) is bounded as follows:*

$$|z(t;t_0,z_0)|<\tilde\rho,\quad \textit{if}\quad t\ge t_0+\widehat T. \tag{3.41}$$

If in particular $z_0\in[0,\tilde\alpha[$, then

$$0\le z(t;t_0,z_0)<\tilde\rho,\quad \textit{if}\quad t\ge t_0+\widehat T. \tag{3.42}$$

Proof The solution $z(t)=z(t;t_0,z_0)$ is of the form

$$\begin{aligned} z(t) &= z_0\, e^{-p(t-t_0)+\int\limits_{t_0}^{t} g(\tau)d\tau} \\ &\quad + e^{-pt+\int\limits_0^t g(\tau)d\tau}\int\limits_{t_0}^{t}\Big[g_1(\tau,\tilde\beta)+g_2(\tau,\tilde\beta)\Big]\, e^{p\tau-\int\limits_0^\tau g(\lambda)d\lambda}\,d\tau. \end{aligned} \tag{3.43}$$

We now consider each of the three terms on the right-hand side of (3.43) separately.

By equation (3.30) for $t\ge t_\vartheta$

$$-p(t-t_0)+\int\limits_{t_0}^{t} g(\tau)\,d\tau\le-(t-t_0)\left[p-q\frac{1+t^\chi}{t-t_0}-M\frac{1+\vartheta(t^\chi-t^\kappa)}{t-t_0}\right];$$

the right-hand side is negatively divergent for $t-t_0\to+\infty$, and so will be the left-hand side; this implies that there exists a $T_0(\tilde\rho,\tilde\alpha,t_0)\ge0$ such that

$$|z_0|\,e^{-p(t-t_0)+\int\limits_{t_0}^{t} g(\tau)d\tau}<\frac{\tilde\rho}{3},\quad t\ge t_0+T_0,\quad z_0\in[-\tilde\alpha,\tilde\alpha]. \tag{3.44}$$

As for the second term, given $\tilde\beta>0,\ \tilde\rho>0,$ because of $(3.19)_1$ there exist $T_1(\tilde\beta,\tilde\rho)\ge\max\{\tilde t,t_\vartheta\}$ and $\sigma_1(\tilde\beta)$ such that

$$\begin{aligned} g_1(\tau,\tilde\beta)&\le\sigma_1 &&\text{if}\quad \tau\ge0,\\ g_1(\tau,\tilde\beta)e^{\xi(\tau^\chi-\tau^\kappa)}&\le\frac16\,m\tilde\rho\,e^{-2M} &&\text{if}\quad \tau\ge T_1 \end{aligned} \tag{3.45}$$

($\tilde t, m$ have been defined respectively in (3.33), (3.27)). Since the function $h(t)$ defined in (3.32) is increasing as the first power of t for $t\ge\tilde t$, there exists $T_2(\tilde\beta,\tilde\rho)\ge T_1$ such that for $t\ge T_2$

$$\frac{\sigma_1}{p}\,e^{-h(t)+M+q+pT_1}<\frac{\tilde\rho}{6}. \tag{3.46}$$

Therefore, for $t \geq T_2$,

$$
\begin{aligned}
& e^{-pt+\int\limits_0^t g(\tau)d\tau} \int\limits_{t_0}^{t} g_1(\tau,\tilde{\beta}) e^{p\tau - \int\limits_0^\tau g(\lambda)d\lambda} d\tau \\
& < e^{-pt+q(1+t^\chi)+M[1+\vartheta(t^\chi - t^\kappa)]} \int\limits_0^t g_1(\tau,\tilde{\beta}) e^{p\tau - \int\limits_0^\tau g(\lambda)d\lambda} d\tau \\
& < e^{-h(t)+M+q} \int\limits_0^{T_1} g_1(\tau,\tilde{\beta}) e^{p\tau} d\tau + e^{-h(t)+2M} \int\limits_{T_1}^{t} g_1(\tau,\tilde{\beta}) e^{h(\tau)+\xi(\tau^\chi-\tau^\kappa)} d\tau \\
& < e^{-h(t)+M+q} \sigma_1 \int\limits_0^{T_1} e^{p\tau} d\tau + e^{-h(t)+2M} e^{-2M} \frac{m\tilde{\rho}}{6} \int\limits_{T_1}^{t} \frac{h'(\tau)}{m} e^{h(\tau)} d\tau \\
& < e^{-h(t)+M+q} \sigma_1 \frac{e^{pT_1}}{p} + \frac{\tilde{\rho}}{6} e^{-h(t)} (e^{h(t)} - e^{h(T_1)}) \\
& < \frac{\tilde{\rho}}{6}(1+1) = \frac{\tilde{\rho}}{3},
\end{aligned}
\tag{3.47}
$$

where in the first and in the second inequality we have used (3.30), the nonnegativity of g_1, the fact that $\xi(\tau^\chi - \tau^\kappa) \geq 0$ and the definition of $h(t)$, in the third (3.45) and (3.33), in the fourth we have performed integration over τ, and in the last we have used (3.46).

As for the third term on the right-hand side of (3.43), from $(3.19)_2$ it follows that there exists $T_3(\tilde{\beta},\tilde{\rho}) \geq \max\{\tilde{t}, t_\vartheta\}$ such that for $t \geq T_3$

$$
e^{2M} \int\limits_{T_3}^{t} g_2(\tau,\tilde{\beta}) e^{\xi(\tau^\chi-\tau^\kappa)} d\tau < \frac{\tilde{\rho}}{6}
\tag{3.48}
$$

and on the other hand that

$$
\int\limits_0^{T_3} g_2(\tau,\tilde{\beta}) e^{p\tau} d\tau < \sigma_2,
\tag{3.49}
$$

where σ_2 has been defined in (3.19). Moreover, there exists $T_4(\tilde{\beta},\tilde{\rho}) \geq T_3$ such that for $t \geq \tilde{t} + T_4$

$$
\sigma_2 e^{-h(t)+q+M} < \frac{\tilde{\rho}}{6}.
\tag{3.50}
$$

Therefore for $t \geq T_4$

$$
\begin{aligned}
& e^{-pt+\int\limits_0^t g(\tau)d\tau} \int\limits_{t_0}^{t} g_2(\tau,\tilde{\beta}) e^{p\tau - \int\limits_0^\tau g(z)dz} d\tau \\
& < e^{-pt+q(1+t^\chi)+M[\vartheta(t^\chi - t^\kappa)+1]} \int\limits_0^t g_2(\tau,\tilde{\beta}) e^{p\tau - \int\limits_0^\tau g(z)dz} d\tau \\
& < e^{-h(t)+q+M} \int\limits_0^{T_3} g_2(\tau,\tilde{\beta}) e^{p\tau} d\tau + e^{-h(t)+2M} \int\limits_{T_3}^{t} g_2(\tau,\tilde{\beta}) e^{\xi(\tau^\chi - \tau^\kappa)} e^{h(\tau)} d\tau \\
& < e^{-h(t)+q+M} \sigma_2 + e^{-h(t)+2M+h(t)} \int\limits_{T_3}^{t} g_2(\tau,\tilde{\beta}) e^{\xi(\tau^\chi - \tau^\kappa)} d\tau \\
& < \frac{\tilde{\rho}}{6} + \frac{\tilde{\rho}}{6} = \frac{\tilde{\rho}}{3},
\end{aligned}
\tag{3.51}
$$

where we have used the nonnegativity of g_2 and (3.30) in the first inequality, again (3.30), the fact that $\xi(\tau^\chi - \tau^\kappa) \geq 0$ and the nonnegativity of g in the second, (3.49) and the monotonicity of $h(\tau)$ for $\tau \geq T_3$ in the third, (3.48) and (3.50) in the last one.

Let $\widehat{T}(\tilde{\rho}, \tilde{\beta}, t_0) := \max\{T_0, T_2, T_4\}$. Collecting the results (3.44), (3.47), (3.51) we find that the solution $z(t)$ of (3.40) fulfills the condition

$$
z(t, t_0, z_0) < \frac{\tilde{\rho}}{3}[1+1+1] = \tilde{\rho}, \quad t \geq t_0 + \widehat{T}.
$$

Remark 1 This lemma is a generalization of Lemma 24.3 in [14], based in turn on an argument due to Hale [10].

Remark 2 If $\chi \leq \kappa$ then in the previous proof T_0, and therefore $\widehat{T}$, becomes independent of t_0. In fact, $\vartheta = 0$ and from (3.30) we find

$$
\int\limits_{t_0}^{t} g(z)\,dz = \int\limits_0^t g(z)\,dz - \int\limits_0^{t_0} g(z)\,dz < q(t^\chi - t_0^\chi) + 2M.
$$

By Lagrange's theorem there exists a $\tau \in]t_0, t[$ such that $t^\chi - t_0^\chi = \dfrac{\chi}{\tau^{1-\chi}}(t - t_0)$. Since $t_0 \geq \tilde{t} \equiv (2q\chi/p)^{1/(1-\chi)}$ we find $t^\chi - t_0^\chi < \dfrac{p}{2q}(t - t_0)$

$$
-p(t-t_0) + \int\limits_{t_0}^{t} g(z)dz < -p(t-t_0)\left[1 - \frac{1}{2}\right] + 2M = -\frac{p}{2}(t-t_0) + 2M.
$$

This implies that the left-hand side is negatively divergent for $t - t_0 \to +\infty$ uniformly in t_0, as anticipated. The argument is not applicable in the case $\chi > \kappa$.

We are now in the conditions to prove the following

Theorem 1 *Assume that the function f of (3.1) is bounded as in (3.16), where $g(t)$, $\tilde{g}_i(t,\eta)$ $(i=1,2$ and $t\in J,\ \eta>0)$ are continuous functions fulfilling the conditions (3.17) – (3.19). Then the solutions of the problem (3.1), (3.2) are eventually uniformly bounded. Moreover, the origin O is eventually quasi-uniform-asymptotically stable in the large with respect to the metric d.*

Proof Set $\tilde{\alpha} := \alpha^2 c_2^2$, and apply Lemma 1. Under the assumption $d(u_0,u_1)\le\alpha$, by (3.13) we find $y_0 = V(t_0)\le\tilde{\alpha}$, by (3.22) and the application of the lemma we find that $y(t)$ (and therefore $V(t)$) is bounded by $\tilde{\beta}(\alpha^2c_2^2)$, and again by (3.13) we find $d(t)\le\beta(\alpha) := \sqrt{\tilde{\beta}(\alpha^2c_2^2)/c_1^2}$ for $t\ge s(\alpha) := \tilde{s}(\alpha^2c_2^2)$, as claimed. Moreover, we can now apply the comparison principle (3.39) – (3.40) and Lemma 2: chosen $\rho>0$, we set $\tilde{\rho} := c_1^2\rho^2$. As a consequence of (3.39), (3.42), (3.13) we thus find that for $t_0\ge s(\alpha)$ and $t\ge\widehat{T}(c_1^2\rho, t_0(\alpha), c_2^2\alpha^2)\equiv T(\rho,\alpha)$

$$d^2(t)\le\frac{V(t)}{c_1^2}\le\frac{y(t)}{c_1^2}\le\frac{z(t,\tilde{\beta}(\alpha))}{c_1^2}<\frac{\tilde{\rho}}{c_1^2}=\rho^2.$$

Remark 3 This theorem is a generalization of Theorem 3.1 in reference [2]: the claims are the same, but the hypotheses on the function f are weakened. First, (3.16) is an upper bound condition only on the mean square value of f^2, rather than on its supremum (as in [2]). Second, this upper bound may depend on t in a more general way than in that reference. The hypotheses (3.17), (3.18), (3.19) considered here are fulfilled by the ones considered there with $g(t)\equiv\text{const}$ and $\chi=\kappa=1$. The former, but not the latter, are satisfied e.g. by the following family of examples.

Examples Let $f=b(t)\sin\varphi$, with a function $b(t)$ such that the integral $\int\limits_0^t b^2(\tau)\,d\tau$ grows as some power t^χ, where $\chi\le1$, and in the case $\chi=1$ is smaller than pt for sufficiently large t; then we can set $\hat{g}(t,\eta)\equiv b^2(t)$. For instance we could take b^2 a continuous function that vanishes everywhere except in intervals centered, say, at equally spaced points, where it takes maxima increasing with some power law $\sim t^\beta$, but keeps the integral bounded, e.g.

$$b^2(t)=b_0^2\begin{cases}4n^{\alpha+\beta}(t-n+\dfrac{1}{2n^\alpha}) & \text{if}\quad t\in\left[n-\dfrac{1}{2n^\alpha},\,n\right],\\[2ex] 4n^\beta-4n^{\alpha+\beta}(t-n) & \text{if}\quad t\in\left]n,\,n+\dfrac{1}{2n^\alpha}\right],\\[2ex] 0 & \text{otherwise},\end{cases}\tag{3.52}$$

with $b_0^2<p$, $\alpha\ge1$, $\beta\in\,]\,\alpha-1,\,\alpha]$ and $n\in N$. (The case $\alpha=\beta=1$ has already been considered in [5]).

The graph of $(b(t)/b_0)^2$ consists of a sequence of isosceles triangles enumerated by n, having bases of length $1/n^\alpha$ and upper vertices with coordinates $(x,y)=(n,2n^\beta)$ (see the Figure 3.1). Their areas are $A_n=1/n^\gamma$, where $\gamma:=\alpha-\beta\in[0,1[$.

If $0\le t-t_0<2$ then we immediately find

$$\int\limits_{t_0}^{t} g(\tau)\,d\tau\le b_0^2\,2.\tag{3.53}$$

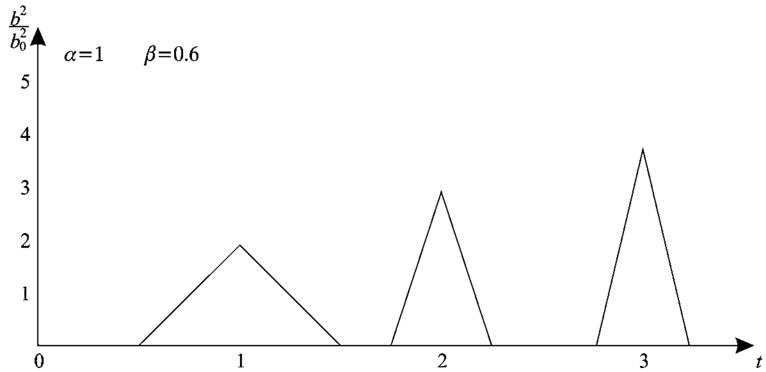

Figure 3.1

If on the contrary $t - t_0 \geq 2$, then there exist integers m, n with $0 \leq m \leq n-2$ and $t > t_0 \geq 0$ such that $t \in]n-1/2, n+1/2]$ and $t_0 \in]m-1/2, m+1/2]$. Then we find

$$\int_{m+1/2}^{n-1/2} g(\tau)\, d\tau \leq \int_{t_0}^{t} g(\tau)\, d\tau \leq \int_{m-1/2}^{n+1/2} g(\tau)\, d\tau,$$

namely

$$\sum_{k=m+1}^{n-1} \frac{b_0^2}{k^\gamma} = b_0^2 \sum_{k=m+1}^{n-1} A_k \leq \int_{t_0}^{t} g(\tau)\, d\tau \leq b_0^2 \sum_{\substack{k=m\\k\geq 1}}^{n} A_k = \sum_{\substack{k=m\\k\geq 1}}^{n} \frac{b_0^2}{k^\gamma}. \tag{3.54}$$

Consider the function $e(y) := y^{1-\gamma}$, $\gamma \in [0,1[$. Applying Lagrange's theorem we find that for any $h \in N$ there exists a $\xi_h \in]h, h+1[$ such that

$$(h+1)^{1-\gamma} - h^{1-\gamma} = (1-\gamma)\frac{1}{\xi_h^\gamma},$$

whence, taking $h = k$ and $h = k-1$ respectively,

$$(k+1)^{1-\gamma} - k^{1-\gamma} < (1-\gamma)\frac{1}{k^\gamma},$$
$$k^{1-\gamma} - (k-1)^{1-\gamma} > (1-\gamma)\frac{1}{k^\gamma};$$

therefore

$$\frac{1}{1-\gamma}\,[(k+1)^{1-\gamma} - k^{1-\gamma}] < \frac{1}{k^\gamma} < \frac{1}{1-\gamma}\,[k^{1-\gamma} - (k-1)^{1-\gamma}]. \tag{3.55}$$

From (3.54), (3.55) we find

$$\frac{b_0^2[n^{1-\gamma} - (m+1)^{1-\gamma}]}{1-\gamma} < \int_{t_0}^{t} g(\tau)\, d\tau < \frac{b_0^2[n^{1-\gamma} - (m-1)^{1-\gamma}(1-\delta_0^m)]}{1-\gamma}. \tag{3.56}$$

where δ_0^m denotes a Kronecker δ. Hence,

$$\int_{t_0}^{t} g(\tau)\,d\tau = \frac{b_0^2}{1-\gamma}\left[n^{1-\gamma}-(m+1)^{1-\gamma}\right]+L_{m,n}(t) \tag{3.57}$$

where the remainder $L_{m,n}(t)$ is bounded by the difference d_m on the right-hand side and left-hand side of (3.56),

$$0<L_{m,n}(t)<d_m:=\frac{b_0^2}{1-\gamma}\left[(m+1)^{1-\gamma}-(m-1)^{1-\gamma}(1-\delta_0^m)\right].$$

The expression in square bracket equals 1 for $m=0$ and $2^{1-\gamma}$ for $m=1$. It is immediate to check that the function $\tilde{e}(y):=(y+1)^{1-\gamma}-(y-1)^{1-\gamma}$ is decreasing for $y\geq 1$ and therefore takes its maximum in $y=1$. We therefore derive the bound

$$0<L_{m,n}(t)<d_m\leq \frac{b_0^2\,\tilde{e}(1)}{1-\gamma}=\frac{b_0^2\,2^{1-\gamma}}{1-\gamma}. \tag{3.58}$$

Moreover, since $t>n-1$, $t_0<m+1$ and g is nonnegative, from (3.57) we find

$$\int_{t_0}^{t} g(\tau)\,d\tau < \frac{b_0^2}{1-\gamma}\left[(t+1)^{1-\gamma}-t_0^{1-\gamma}\right]+L_{m,n}(t).$$

If $t_0\geq 1$, applying again Lagrange's theorem to the function $e(t)=t^{1-\gamma}$ we find

$$\frac{b_0^2}{1-\gamma}\left[(t+1)^{1-\gamma}-t_0^{1-\gamma}\right]=b_0^2\frac{t-t_0+1}{\bar{t}^{\gamma}}<b_0^2(t-t_0+1)$$

with a suitable $\bar{t}\in\,]t_0,\,t+1[$, and therefore

$$\int_{t_0}^{t} g(\tau)\,d\tau - b_0^2(t-t_0) < b_0^2\left(1+\frac{2^{1-\gamma}}{1-\gamma}\right). \tag{3.59}$$

If $0\leq t_0<1$,

$$\int_{t_0}^{t} g(\tau)\,d\tau - b_0^2(t-t_0)\leq \int_{0}^{1} g(\tau)\,d\tau - b_0^2(1-t_0)+\int_{1}^{t} g(\tau)\,d\tau - b_0^2(t-1) < b_0^2\left(2+\frac{2^{1-\gamma}}{1-\gamma}\right)=:\sigma,$$

where we have used (3.59) with $t_0=1$ and $\int_0^1 g(\tau)d\tau\leq b_0^2$, showing (together with (3.59) itself and (3.53)) that g fulfills condition (3.17) in any case.

On the other hand, choosing $t_0=0$ (and therefore $m=0$) in (3.57), dividing by $1+t^{1-\gamma}$ and subtracting $b_0^2/(1-\gamma)$ we find

$$\frac{\int_0^t g(\tau)\,d\tau}{1+t^{1-\gamma}}-\frac{b_0^2}{1-\gamma}=\frac{b_0^2}{1-\gamma}\left[\frac{n^{1-\gamma}-(1+t^{1-\gamma})-1}{1+t^{1-\gamma}}\right]+\frac{L_{0,n}(t)}{1+t^{1-\gamma}}.$$

But it is $n-1<t<n+1$, what implies

$$1-2^{1-\gamma} \le n^{1-\gamma}-(n+1)^{1-\gamma} < n^{1-\gamma}-t^{1-\gamma} < (t+1)^{1-\gamma}-t^{1-\gamma} < 1$$

(in fact the function $\hat{e}(y) := (y+1)^{1-\gamma}-y^{1-\gamma}$ is decreasing and therefore has maximum at the lower extremum of the interval in which we define it); hence, using also (3.58), we find

$$-\frac{b_0^2}{1-\gamma}\left[\frac{2^{1-\gamma}+1}{1+t^{1-\gamma}}\right] < \frac{\int\limits_0^t g(\tau)d\tau}{1+t^{1-\gamma}} - \frac{b_0^2}{1-\gamma} < \frac{b_0^2}{1-\gamma}\left[\frac{2^{1-\gamma}-1}{1+t^{1-\gamma}}\right] < \frac{b_0^2}{1-\gamma}\left[\frac{2^{1-\gamma}+1}{1+t^{1-\gamma}}\right].$$

We have proved these inequalities under the current assumption $t \ge 2$, showing that in this domain also condition (3.18), with $q = b_0^2/(1-\gamma)$, $\chi=\kappa=1-\gamma$ and $M = b_0^2(2^{1-\gamma}+1)/(1-\gamma)$, is satisfied. For $0 \le t \le 2$ the left-hand side of (3.18) is certainly bounded by $b_0^2\,3/[2(1-\gamma)]$, therefore it is sufficient to choose e.g. $M = b_0^2\,9/[2(1-\gamma)]$ to fulfill (3.18) for any $t \ge 0$.

4 Exponential-Asymptotic Stability for Special f's via a Family of Liapunov Functionals

In this section we specialize the function f of (3.1) as $f = F(u) - a(x,t,u,u_x,u_t,u_{xx})u_t$, where $F \in C(R)$ and $a \in C(\,]0,1[\, \times J \times R^4)$, and examine the particular problem

$$\begin{gathered} Lu = F(u) - a(x,t,u,u_x,u_t,u_{xx})u_t, \quad x \in]0,1[, \quad t > t_0, \\ u(0,t)=0, \quad u(1,t)=0, \quad t>t_0, \end{gathered} \tag{4.1}$$

with initial and consistency conditions (3.2)–(3.3). We shall use the one-parameter family of modified Liapunov functionals

$$\begin{aligned} W_\gamma(\varphi,\psi) = {} & \frac{1}{2}\int\limits_0^1 \left\{(\varepsilon\varphi_{xx}-\psi)^2 + \gamma\psi^2 + (1+\gamma)\varphi_x^2\right\} dx \\ & - (1+\gamma)\int\limits_0^1 \left(\int\limits_0^{\varphi(x)} F(z)dz\right) dx \end{aligned} \tag{4.2}$$

where $\gamma > 1/2$ is for the moment an unspecified parameter.

Theorem 2 *Under the following assumptions*

(1) $F(u) \in C^1(R)$, $F(0)=0$, *and moreover there exists a positive constant K such that*

$$F_u \le K < 3\pi^2/4; \tag{4.3}$$

(2) *the function a satisfies*

$$\nu := \varepsilon\pi^2 + \inf a > 0; \tag{4.4}$$

(3) *there exist $\tau \in [0,2[$ and constants $A>0$, $A' \ge 0$ such that*

$$a(x,t,\varphi,\varphi_x,\varphi_{xx},\psi) \le A[d(\varphi,\psi)]^\tau + A', \tag{4.5}$$

the zero solution of the problem (4.1) is exponential-asymptotically stable in the large.

As anticipated in the introduction, this should be compared with Theorem 3.3. in the main reference, [2]: by replacing the requirement that $\sup a < \infty$ and adding the assumption (4.5) we are still able to prove the exponential-asymptotic stability in the large of the zero solution. The trick is to associate to each neighbourhood of the origin with radius σ (the 'error') a Liapunov functional (4.2) with parameter γ adapted to σ, instead of fixing γ once and for all.

Proof We start by improving or recalling some inequalities proved in [2]. From (4.3) we find

$$\int_0^{\varphi} F(z)\,dz = \int_0^{\varphi} dz \int_0^{z} F_s(s)\,ds \leq K \int_0^{\varphi} dz \int_0^{z} ds = K\varphi^2/2. \tag{4.6}$$

Employing this inequality and the estimate (3.9) we find

$$\begin{aligned} W_\gamma(\varphi,\psi) = \frac{1}{2}\int_0^1 \Big\{ &(\varepsilon\varphi_{xx} - 2\psi)^2/4 + (\varepsilon\varphi_{xx} - \psi)^2/2 + (\gamma - 1/2)\psi^2 \\ &+ (1+\gamma)\varphi_x^2 + \varepsilon^2\varphi_{xx}^2/4 - 2(1+\gamma)\int_0^{\varphi} F(z)\,dz \Big\}\,dx. \end{aligned} \tag{4.7}$$

It easy to see that

$$\begin{aligned} W_\gamma(\varphi,\psi) &\geq \frac{1}{2}\int_0^1 \left[\left(\gamma - \frac{1}{2}\right)\psi^2 + (1+\gamma)\varphi_x^2 + \frac{\varepsilon^2}{4}\varphi_{xx}^2 - 2(1+\gamma)\int_0^{\varphi} F(z)\,dz \right] dx \\ &\geq \frac{1}{2}\int_0^1 \left[\left(\gamma - \frac{1}{2}\right)\psi^2 + (1+\gamma)\pi^2\varphi^2 + \frac{\varepsilon^2}{4}\omega_3(\varphi_{xx}^2 + \varphi_x^2) - (1+\gamma)K\varphi^2 \right] dx \\ &\geq k_1^2 d^2(\varphi,\psi), \end{aligned} \tag{4.8}$$

where we have used again (4.3) and we have introduced the constant k_1^2

$$k_1^2 = \min\{\varepsilon^2\omega_3/8,\, (2\gamma - 1)/4\}, \quad \gamma > 1/2. \tag{4.9}$$

Another inequality of [2] reads

$$W_\gamma(\varphi,\psi) \leq c_2^2\,[1 + m(d(\varphi,\psi))]\,d^2(\varphi,\psi), \tag{4.10}$$

where

$$m(|\varphi|) = \max\{|F_\zeta(\zeta)|:\ |\zeta| \leq |\varphi|\}. \tag{4.11}$$

The map $B(d) := [1 + m(d)]^{1/2}d$ is increasing and continuous, therefore invertible. Finally,

$$\frac{dW_\gamma(u,u_t)}{dt} = -\int_0^1 \{\varepsilon u_{xx}^2 + \varepsilon\gamma u_{xt}^2 + a(1+\gamma)u_t^2 + \varepsilon F(u)u_{xx} - \varepsilon a u_{xx}u_t\}\,dx$$

$$= -\int_0^1 \left\{\frac{3}{4}\varepsilon u_{xx}^2 + \varepsilon\left[\frac{c}{2}u_{xx} - \frac{a}{c}u_t\right]^2 + \varepsilon\gamma u_{tx}^2 + a\left[1+\gamma-\varepsilon a\right]u_t^2 - \varepsilon F_u u_x^2\right\}\,dx \tag{4.12}$$

$$\le -\int_0^1 \{3\varepsilon(1-\lambda)u_{xx}^2/4 + \varepsilon(3\lambda\pi^2/4 - K)u_x^2 + [(\varepsilon\pi^2 + a)\gamma + a(1-\varepsilon a)]u_t^2\}dx,$$

$$\frac{dW_\gamma(u,u_t)}{dt} \le -\int_0^1 \{3\varepsilon(1-\lambda)\omega_1(u_{xx}^2 + u^2)/4 + \varepsilon(3\lambda\pi^2/4 - K)u_x^2 + [(\varepsilon\pi^2 + a)\gamma + a(1-\varepsilon a)]u_t^2\}\,dx, \tag{4.13}$$

where $\lambda \in]0,1[$ is a constant chosen in such a way that $3\lambda\pi^2/4 - K > 0$, and we have used (3.9), (4.3).

Now we are going to show that for any "error" $\sigma > 0$ there exists a $\delta \in]0,\sigma[$ such that $d(t_0) \equiv d(u_0,u_1) < \delta$ implies

$$d(t) \equiv d(u(x,t),u_t(x,t)) < \sigma \quad \forall\, t \ge t_0. \tag{4.14}$$

To this end we associate to the neighbourhood with radius σ of the zero solution the Liapunov functional (4.2) choosing the parameter γ and δ as the following functions of σ:

$$\gamma(\sigma) = (A\sigma^\tau + A')\varepsilon + M, \qquad M := \frac{1+\varepsilon\pi^2+\varepsilon^3\pi^4}{\nu} + \frac{1}{\varepsilon\pi^2} + \frac{1}{2}, \tag{4.15}$$

$$\delta(\sigma) = B^{-1}\left(\frac{\sigma k_1(\gamma(\sigma))}{c_2(\gamma(\sigma))}\right); \tag{4.16}$$

we shall call the corresponding Liapunov functional W_σ. *Per absurdum*, assume that there exist a $t_1 > t_0$ such that (4.14) is fulfilled for any $t \in [t_0,t_1[$, whereas

$$d(t_1) = \sigma. \tag{4.17}$$

Consider the term in the square bracket on the right-hand side of (4.13). From (4.15), (4.4), (4.5) considering separately the cases $a > 0$, $-\varepsilon\pi^2 < a \le 0$, we find

$$-\left[(\varepsilon\pi^2 + a)\gamma + a(1-\varepsilon a)\right] \le -1, \tag{4.18}$$

whence

$$\frac{dW_\sigma(u(t),u_t(t))}{dt} \le -k_3^2 d^2(u(t),u_t(t)) < 0, \tag{4.19}$$

where

$$k_3^2 = \min\{3\varepsilon(1-\lambda)\omega_1/4,\ \varepsilon(3\lambda\pi^2/4-K),\ 1\}. \tag{4.20}$$

From (4.8), (4.19), (4.10), (4.16), it follows

$$\begin{aligned} k_1^2 d^2(t_1) &\le W_\sigma(u(t_1),u_t(t_1)) < W_\sigma(u(t_0),u_t(t_0)) \le c_2^2\,[1+m(d(t_0))]\,d^2(t_0) \\ &< c_2^2\,[1+m(\delta)]\,\delta^2 = c_2^2\,[B(\delta)]^2 = c_2^2\left[B\left(B^{-1}\left(\frac{\sigma k_1}{c_2}\right)\right)\right]^2 = k_1^2\sigma^2, \end{aligned}$$

against (4.17).

Having proved (4.14), it follows $m(d(t)) < m(\sigma)$, which replaced in (4.10) gives

$$W_\sigma \le c_2^2(\sigma)\,[1+m(\sigma)]\,d^2(t);$$

together with (4.19) this in turn implies

$$\frac{dW_\sigma(u(t),u_t(t))}{dt} \le -C(\sigma)W_\sigma(u(t),u_t(t)),$$

with $C(\sigma) := k_3^2/[c_2^2(\sigma)(1+m(\sigma))]$. Using the comparison principle we find that $d(t_0) \equiv d(u_0,u_1) < \delta$ implies

$$d(u(t),u_t(t)) \le D(\sigma)e^{-\frac{C(\sigma)}{2}(t-t_0)}d(u_0,u_1), \tag{4.21}$$

with $D(\sigma) := \dfrac{c_2}{k_1}\sqrt{1+m(\delta(\sigma))}$.

Last, we show that under the present assumptions the function (4.16) can be inverted. It is evident from (4.9) that $k_1(\sigma)$ is non-decreasing, from (3.7) and (4.5) that $\sigma/c_2(\gamma(\sigma))$ is strictly increasing, therefore that $\sigma k_1(\sigma)/c_2(\gamma(\sigma))$ is strictly increasing too, hence invertible. Since B^{-1} is invertible, $\delta(\sigma)$ is invertible and its range is J.

Thus we can express $D(\sigma)$, $C(\sigma)$ as functions of δ, proving the exponential asymptotic stability of the zero solution.

Remark 4 The theorem holds also if we replace the right-hand side of (4.5) with $A(d)$, where $A\colon [0,+\infty[\to R^+$ is any nondecreasing function such that $A(\sigma)/\sigma^2 \xrightarrow{\sigma\to+\infty} 0$.

Remark 5 If (4.5) holds with $\tau = 2$ the function $\dfrac{\sigma}{c_2(\gamma(\sigma))}$ is still increasing but its range is $[0,\, 2/\varepsilon A]$, implying that the function $\dfrac{\sigma k_1(\gamma(\sigma))}{c_2(\gamma(\sigma))}$ is still increasing but its range is $[0,\, \sqrt{\omega_3}/\sqrt{2}A]$. Therefore the condition (3.5) of Definition 3.4 is fulfilled only for $\alpha \in\,]0,\, B^{-1}(\sqrt{\omega_3}/\sqrt{2}A)[$, and the attraction region includes the set $d(u_0,u_1) < B^{-1}(\sqrt{\omega_3}/\sqrt{2}A)$.

We now give a variant of the preceding theorem, based on a hypothesis slightly different from (4.5). Beside the distance (3.4), we need also a "weaker" distance $d_1(u,u_t)$ between the zero and a nonzero solution $u(x,t)$ of the problem (3.1)−(3.2): for any $(\varphi,\psi) \in C_0^2([0,1]) \times C_0([0,1])$ we define

$$d_1^2(\varphi,\psi) = \int_0^1 (\varphi^2+\varphi_x^2+\psi^2)\,dx. \tag{4.22}$$

Clearly,

$$d_1(\varphi,\psi) \le d(\varphi,\psi). \tag{4.23}$$

The “Hamiltonian” Liapunov functional $v(u,u_t)$, with

$$v(\varphi,\psi) := \frac{1}{2}\int_0^1 \left\{\psi^2 + \varphi_x^2 - 2\left(\int_0^{\varphi(x)} F(z)\,dz\right)\right\} dx, \tag{4.24}$$

will play w.r.t. the distance d_1 a role similar to the one played by the Liapunov functionals V or W_γ w.r.t. the distance d.

Theorem 3 *Under the following assumptions*

(1) $F(u) \in C^1(R)$, $F(0) = 0$, *and there exists a positive constant* K *such that*

$$F_u \le K < 3\pi^2/4; \tag{4.25}$$

(2) *the function* a *satisfies*

$$\inf a > -\varepsilon\pi^2; \tag{4.26}$$

(3) *there exists a nondecreasing map* $A\colon J \to J$ *such that*

$$|a(x,t,\varphi,\varphi_x,\varphi_{xx},\psi)| \le A\left[d_1(\varphi,\psi)\right], \tag{4.27}$$

the zero solution of the problem (4.1) is exponential-asymptotically stable in the large.

Proof Some steps of the proof are exactly as in the previous theorem. Employing inequality (4.6) and the estimate (3.9) we find

$$v \ge \frac{1}{2}\int_0^1 \left\{\left(\frac{1}{8}\,u_x^2 + \frac{7}{8}\,u^2\pi^2\right) + u_t^2 - \frac{3}{4}\,\pi^2 u^2\right\} dx \ge \frac{1}{16}\,d_1^2. \tag{4.28}$$

Setting $v(t) \equiv v(u,u_t)$, integrating by parts and using (4.1), (4.26), (3.9) we also find

$$\begin{aligned} \frac{dv}{dt} &= \int_0^1 \{u_t[-u_{xx} + u_{tt} - F(u)]\}\,dx = -\int_0^1 \left\{\varepsilon u_{xt}^2 + a u_t^2\right\} dx \\ &\le -\int_0^1 (\varepsilon\pi^2 + a)u_t^2\,dx < 0 \end{aligned} \tag{4.29}$$

Now we are going to prove the uniform boundedness of the solutions of the problem (4.1). To this end first note that from the definition (4.11) it follows

$$\left|\int_0^{\varphi} F(z)dz\right| \le m(|\varphi|)\frac{\varphi^2}{2};$$

employing this inequality and the one $\varphi^2 \leq d_1^2(\varphi,\psi)$ we find

$$v \leq \frac{1}{2}\left[1+m\left(d_1(u,u_t)\right)\right] d_1^2(u,u_t). \tag{4.30}$$

From (4.29) we derive the inequality $v(t) < v(t_0)$ for any $t > t_0$, whence

$$\frac{1}{16}\, d_1^2(t) \leq v(t) < v(t_0) \leq \frac{1}{2}\left[1+m\left(d_1(t_0)\right)\right] d_1^2(t_0).$$

Therefore, for any $t > t_0$

$$d(t_0) \leq \alpha \quad \Longrightarrow \quad d_1(t_0) \leq \alpha \quad \Longrightarrow \quad d_1(t) < \beta_1(\alpha) := 2\sqrt{2}\left[1+m(\alpha)\right]^{1/2}\alpha,$$

so that, in view of the assumption (4.27),

$$d(t_0) \leq \alpha \quad \Longrightarrow \quad |a(x,t,u,u_x,u_t,u_{xx})| \leq A\left[\beta_1(\alpha)\right] \equiv A(\alpha). \tag{4.31}$$

Now we associate to any $\alpha > 0$ the Liapunov functional (4.2) with the parameter γ chosen as the following function of α:

$$\gamma(\alpha) = A(\alpha)\varepsilon + M, \qquad M := \frac{1+\varepsilon\pi^2+\varepsilon^3\pi^4}{\nu} + \frac{1}{\varepsilon\pi^2} + \frac{1}{2}; \tag{4.32}$$

we shall call the corresponding Liapunov functional W_α. Consider the term in the square bracket on the right-hand side of (4.13). From (4.31), (4.32), we find again (4.18), whence

$$\frac{dW_\alpha(u(t),u_t(t))}{dt} \leq -k_3^2 d^2(u(t),u_t(t)) < 0, \tag{4.33}$$

with the same k_3^2 of (4.20). From (4.8), (4.33), (4.10), it follows for any $t > t_0$

$$\begin{aligned} k_1^2 d^2(t) &\leq W_\alpha(u(t),u_t(t)) < W_\alpha(u(t_0),u_t(t_0)) \leq c_2^2\left[1+m(d(t_0))\right] d^2(t_0) \\ &< c_2^2(\gamma(\alpha))\left[1+m(\alpha)\right]\alpha^2 = c_2^2(\gamma(\alpha))B^2(\alpha), \end{aligned}$$

proving the uniform boundedness of u:

$$d(u(t),u_t(t)) < \frac{c_2(\gamma(\alpha))}{k_1(\gamma(\alpha))}\, B(\alpha) \equiv \beta(\alpha). \tag{4.34}$$

Having proved this, it follows $m(d(t)) < m(\beta(\alpha))$, which replaced in (4.10) gives

$$W_\alpha \leq c_2^2(\gamma(\alpha))[1+m(\beta(\alpha))]d^2(t);$$

together with (4.33) this in turn implies

$$\frac{dW_\alpha(u(t),u_t(t))}{dt} \leq -C(\alpha)W_\alpha(u(t),u_t(t)),$$

with $C(\alpha) := k_3^2(\gamma(\alpha))/\{c_2^2(\gamma(\alpha))[1+m(\beta(\alpha))]\}$. Using the comparison principle we find that $d(t_0) \equiv d(u_0,u_1) \leq \alpha$ implies

$$d(u(t),u_t(t)) \leq D(\alpha)e^{-C(\alpha)(t-t_0)} d(u_0,u_1), \tag{4.35}$$

with $D(\alpha) := \dfrac{c_2(\gamma(\alpha))}{k_1(\gamma(\alpha))}\sqrt{1+m(\beta(\alpha))}$, namely the exponential-asymptotical stability.

5 Uniform Asymptotic Stability in the Large for a Class of Non-Analytic f's

Here we give a generalization of Theorem 2 in [5]. As in the preceding sections, using the trick of the one-parameter family of Liapunov functionals we are able to replace the boundedness assumption for the function a by a weaker one.

Theorem 4 *Under the following assumptions*

$$F(\varphi) \in C(R) \ \text{such that} \ F(0) = 0, \tag{5.1}$$

$$\text{there exist } \tau \in [0,1[\ \text{and} \ D > 0 \ \text{such that, for any } \varphi, \psi \tag{5.2}$$

$$0 \le -\int_0^1 \left(\int_0^{\varphi(x)} F(z)dz \right) dx \le D d^{\tau+1}(\varphi, \psi),$$

$$\int_0^1 F(\varphi(x))\varphi_{xx}(x)\, dx \ge 0 \ \ \text{for any} \ \varphi \in C_0^2([0,1]), \tag{5.3}$$

$$\text{the function } a \text{ satisfies } \ \inf a > -\varepsilon\pi^2, \tag{5.4}$$

$$\text{there exists a nondecreasing map } A\colon [0,\infty[\to R^+ \ \text{such that} \tag{5.5}$$

$$|a(x,t,\varphi,\varphi_x,\varphi_{xx})| \le A\left(d(\varphi,\psi)\right),$$

the zero solution of the problem (4.1) is uniformly asymptotically stable in the large.

Proof From (4.7), (5.2)

$$\begin{aligned} W_\gamma(\varphi,\psi) &\ge \frac{1}{2}\int_0^1 \{(\gamma - 1/2)\psi^2 + (1+\gamma)\varphi_x^2 + \varepsilon^2\varphi_{xx}^2/4\}\, dx \\ &\ge \frac{1}{2}\int_0^1 \{(\gamma-1/2)\psi^2 + (1+\gamma)\omega_3(\varphi^2+\varphi_x^2) + \varepsilon^2\varphi_{xx}^2/4\}\, dx \ge {k_1'}^2 d^2(\varphi,\psi), \end{aligned} \tag{5.6}$$

where

$${k_1'}^2 := \frac{1}{2}\min\left\{\gamma - \frac{1}{2},\, \frac{\varepsilon^2}{4},\, (1+\gamma)\omega_3\right\}, \quad \gamma > \frac{1}{2}. \tag{5.7}$$

Moreover, taking into account (4.2), assumption (5.2), noting that $(\varepsilon\varphi_{xx} - \psi)^2 \le \varepsilon^2\varphi_{xx}^2 + \psi^2 + \varepsilon(\varphi_{xx}^2 + \psi^2)$, and considering (3.7) it follows

$$W_\gamma(\varphi,\psi) \le G_\gamma(d(\varphi,\psi)), \tag{5.8}$$

where

$$G_\gamma(d) := c_2^2(\gamma)d^2 + D(\gamma+1)d^{\tau+1}. \tag{5.9}$$

For any choice of $\gamma > \dfrac{1}{2}$ the map $G_\gamma(d)$ is increasing and continuous in d, therefore invertible. Finally, with the help of (3.9) we obtain from (4.12)

$$\begin{aligned} \frac{dW_\gamma(u,u_t)}{dt} &\le -\int_0^1 \{(3/4)\varepsilon u_{xx}^2 + [\varepsilon\gamma + a(1+\gamma-\varepsilon a)]u_t^2\}\, dx \\ &\le -\int_0^1 \{\varepsilon\omega_2(u_{xx}^2 + u_x^2 + u^2)/4 + [(\varepsilon + a)\gamma + a(1-\varepsilon a)]u_t^2\}\, dx. \end{aligned} \tag{5.10}$$

Now we are going to show that for any "error" $\sigma > 0$ there exists a $\delta \in]0, \sigma[$ such that $d(t_0) \equiv d(u_0, u_1) < \delta$ implies

$$d(t) \equiv d\left(u(x,t), u_t(x,t)\right) < \sigma \quad \forall\, t \geq t_0. \tag{5.11}$$

To this end we choose the parameter γ in the Liapunov functional (4.2) as in (4.32) and δ as the following function of the error σ:

$$\delta(\sigma) = G^{-1}_{\gamma(\sigma)}\left(\sigma^2 {k'_1}^2\left(\gamma(\sigma)\right)\right); \tag{5.12}$$

we shall indicate the corresponding Liapunov functional $W_{\gamma(\sigma)}$ simply by W_σ. *Per absurdum*, assume that there exist a $t_1 > t_0$ such that (4.14) is fulfilled for any $t \in [t_0, t_1[$, whereas (4.17) holds for $t = t_1$. Consider the term in the square bracket on the right-hand side of (5.10). From (4.32), (4.4), (5.5) we get again (4.18), whence

$$\frac{dW_\sigma(u(t), u_t(t))}{dt} \leq -{k'_3}^2 d^2(u(t), u_t(t)) < 0, \tag{5.13}$$

where now ${k'_3}^2 := \min\{\varepsilon\omega_2/4,\, 1\}$. From (5.6), (5.8), (5.13),(5.12), it follows

$$\begin{aligned} {k'_1}^2 d^2(t_1) &\leq W_\sigma(u(t_1), u_t(t_1)) < W_\sigma(u(t_0), u_t(t_0)) \\ &\leq G_{\gamma(\sigma)}(d(t_0)) < G_{\gamma(\sigma)}(\delta(\sigma)) = {k'_1}^2 \sigma^2, \end{aligned}$$

against (4.17). So we have proved the uniform stability of the zero solution.

Note now that the function $\delta(\sigma)$ is invertible, since it is the composition of two increasing functions. Therefore W_σ can be expressed as a function W_δ of the parameter δ. By (5.13) it is $W_\delta(t) \leq W_\delta(t_0)$ so by (5.6), (5.8) we find that for $d(t_0) \equiv d(u_0, u_1) \leq \delta$

$$d^2(t) \leq \frac{W_\delta(t)}{{k'_1}^2} \leq \frac{W_\delta(t_0)}{{k'_1}^2} \leq \frac{G_{\gamma(d(t_0))}}{{k'_1}^2} \leq \frac{G_{\gamma(\delta)}}{{k'_1}^2\big(\gamma(\delta)\big)} =: \beta^2(\delta),$$

proving the uniform boundedness of u.

Employing an argument of [5] one can now show that for any choice of the initial condition $d(t_0) < \delta$ the functional W_δ decreases to zero (at least) as a negative power of $(t - t_0)$ as $(t - t_0) \to \infty$. From (5.8) we find

$$d^2 \geq \min\left\{\frac{W_\delta}{2c_2^2(\gamma(\sigma))}, \left(\frac{W_\delta}{2D(\gamma+1)}\right)^{\frac{2}{\tau+1}}\right\},$$

which considered in (5.13) gives

$$\frac{dW_\delta(u, u_t)}{dt} \leq -k'_3 \min\left\{\frac{W_\delta}{2c_2^2}, \left(\frac{W_\delta}{2D(\gamma+1)}\right)^{\frac{2}{\tau+1}}\right\} \leq 0. \tag{5.14}$$

If at $t = t_0$

$$\frac{W_\delta}{2c_2^2} \geq \left(\frac{W_\delta}{2D(\gamma+1)}\right)^{\frac{2}{\tau+1}}, \tag{5.15}$$

then setting

$$E(\delta) := \frac{k'_3}{\left[2D\big(\gamma(\delta)+1\big)\right]^{\frac{2}{\tau+1}}} \frac{1-\tau}{1+\tau} > 0$$

one finds

$$d^2(t) \leq \frac{W_\delta(t)}{{k'_1}^2} \leq \frac{1}{{k'_1}^2[W_\delta(t_0) + E(t-t_0)]^{\frac{1+\tau}{1-\tau}}} \leq \frac{1}{{k'_1}^2[E(t-t_0)]^{\frac{1+\tau}{1-\tau}}} \tag{5.16}$$

for $t \geq t_0$. If on the contrary

$$\frac{W_\delta(t_0)}{2c_2^2} < \left(\frac{W_\delta(t_0)}{2D(\gamma+1)}\right)^{\frac{2}{\tau+1}},$$

(5.14) will imply for some time

$$\frac{dW_\delta(u,u_t)}{dt} \leq -k_3' W_\delta$$

and by the comparison principle an (at least) exponential decrease of W_δ. Hence there will exist a $\widetilde{T}(\delta) > 0$ such that

$$\frac{W_\delta(t_0+\tilde{T})}{2c_2^2} = \left(\frac{W_\delta(t_0+\widetilde{T})}{2D(\gamma+1)}\right)^{\frac{2}{\tau+1}},$$

after which (5.14) will take again the form considered in the previous case and thus imply

$$\begin{aligned} d^2(t) \leq \frac{W_\delta(t)}{k_1'} &\leq \frac{1}{k_1'^2[W_\delta(t_0+\widetilde{T}) + E(t-t_0-\widetilde{T})]^{\frac{1+\tau}{1-\tau}}} \\ &\leq \frac{1}{k_1'^2[E(t-t_0-\widetilde{T})]^{\frac{1+\tau}{1-\tau}}} \end{aligned} \tag{5.17}$$

for $t \geq t_0 + \widetilde{T}$. Formula (5.17) will be valid also if δ is so small that inequality (5.15) occurs, provided we correspondingly define $\widetilde{T} := 0$, so that it reduces to (5.16). Formula (5.17) evidently implies the quasi-uniform asymptotic stability in the large of the zero solution.

References

[1] Cannon, J.R. *The One-Dimensional Heat Equation.* Addison-Wesley, 1984.

[2] D'Acunto, B. and D'Anna, A. Stability for a third order Sine-Gordon equation. *Rend. Mat. Serie VII* **18** (1998) 347–365.

[3] D'Acunto, B. and Renno, P. On some nonlinear visco-elastic models. *Ricerche di Mat.* **41** (1992) 101–122.

[4] D'Acunto, B. and Renno, P. On the operator $\varepsilon\partial_{xxt} + c^2\partial_{xx} - \partial_{tt}$ in general domains. *Atti del Seminario Matematico e Fisico dell'universitá di Modena* **XLVII** (1999) 191–202.

[5] D'Anna, A. and Fiore, G. Stability and attractivity for a class of dissipative phenomena. *Rend. Mat. Serie VII* **21** (2000) 191–206.

[6] Davydov, A.S. *Solitons in Molecular Systems.* Reidel Publishing Company, 1985.

[7] Ghidaglia, J.M. and Marzocchi, A. Long time behaviour of strongly damped wave equations, global attractors and their dimensions. *SIAM J. Math. Anal.* **22** (1991) 879–895.

[8] Greenberg, J.M., MacCamy, R.C. and Mizel, V.J. On the existence, uniqueness, and stability of solutions of the equation $\sigma'(u_x)u_{xx} + \lambda u_{xtx} = \rho_0 u_{tt}$. *J. Math. Mech.* **17**(7) (1968) 707–728.

[9] Greenberg, J.M. and MacCamy, R.C. On the exponential stability of solutions of $E(u_x)u_{xx} + \lambda u_{xtx} = \rho u_{tt}$. *J. Math. Anal. Appl.* **31** (1970) 406–417.

[10] Hale, J.K. Asymptotic behaviour of solutions of differential-difference equations. *Proc. of Symp. Nonlinear Oscillations, IUTAM*, Kiev, September 1961, **II**, P.409–426.

[11] Lomdhal, P.S., Soerensen, O.H. and Christiansen, P.L. Soliton excitations in Josephson tunnel junctions. *Phys. Rev. B* **25** (1982) 5737–5748.

[12] Morrison, J.A. Wave propagations in rods of Voigt material and visco-elastic materials with three-parameters models. *Quart. Appl. Math.* **14** (1956) 153–169.

[13] Flavin, N. and Rionero, S. *Qualitative Estimates for Partial Differential Equations*. CRC Press, 1996.

[14] Yoshizawa, T. *Stability Theory by Liapunov's Second Method*. The Mathematical Society of Japan, 1966.

Nonlinear Dynamics and Systems Theory, **5**(1) (2005) 39–48

Optimal Maneuvers Using a Three Dimensional Gravity Assist

G. Felipe and A.F.B.A. Prado

Instituto Nacional de Pesquisas Espaciais,
São José dos Campos, 12227-010, Brazil

Received: January 12, 2004; Revised: June 29, 2004

Abstract: In the present paper the swing-by maneuvers are studied and classified under the model given by the three-dimensional restricted three-body problem. The modification in the orbit of the spacecraft due to the close approach is shown in plots that specify from which type of orbits the spacecraft is coming and to which type it is going. The results generated here are used to solve optimal problems, such as finding trajectories that satisfy some given constraints (such as achieving an escape or a capture) with some parameters being extremized (position, velocity, etc...).

Keywords: *Astrodynamics; swing-by; orbital maneuvers; gravity-assist; restricted problem.*

Mathematics Subject Classification (2000): 70M20, 70Q05.

1 Introduction

Applications of the swing-by technique can be found in several publications in the literature [1−9]. In the present paper the swing-by maneuvers are studied and classified under the model given by the three-dimensional circular restricted three-body problem. The goal is to simulate a large variety of initial conditions for those orbits and classify them according to the effects caused by the close approach in the orbit of the spacecraft. This swing-by is assumed to be performed around the secondary body of the system. For a large number of values of these three variables, the equations of motion are integrated numerically forward and backward in time, until the spacecraft is at a distance that can be considered far enough from M_2. It is necessary to integrate in both directions of time because the set of initial conditions used gives information about the spacecraft exactly at the moment of the closest approach. At these two points, the effect of M_2 can be neglected and the system formed by M_1 and the spacecraft can be considered a two-body system. At these two points, two-body celestial mechanics formulas are valid to compute the energy and the angular momentum before and after the close approach. With those results, the orbits are classified in four categories: elliptic direct (negative energy and positive angular momentum), elliptic retrograde (negative energy and angular momentum), hyperbolic direct (positive energy and angular momentum) and hyperbolic retrograde (positive energy and negative angular momentum). Then, the problem is to identify the category of the orbit of the spacecraft before and after the close encounter with M_2. After that, those results are used to identify up to sixteen classes of transfers, accordingly to the changes in the category of the orbit caused by the close encounter. They are named with the first sixteen letters of the alphabet. After that, several optimal problems involving this maneuver can be formulated and solved with the help of the plots shown. Some examples include finding specific types of orbits (escape, capture, etc.) that have maximum or minimum velocity at periapsis (or any other parameters, such as the distance of the periapsis or the angle of approach).

2 The Swing-By in Three Dimensions

This maneuver can be identified by four independent parameters: i) V_p, the magnitude of the velocity of the spacecraft at periapsis. For the most general case, it would be necessary to give an information about the direction of the velocity. In this paper, only velocities parallel to the x–y plane are considered. This constraint is assumed, because it is the most usual situation in interplanetary research, since the planets have orbits that are almost coplanar. Under this approximation, and taking into account that the velocity at periapse is perpendicular to the periapsis vector, the information about the magnitude of the velocity is enough to completely specify the velocity vector; ii) R_p, the distance between the spacecraft and the celestial body during the closest approach; iii) α, the angle between the projection of the periapsis line in the x–y plane and the line that connects the two primaries; iv) β, the angle between the periapsis line and the x–y plane.

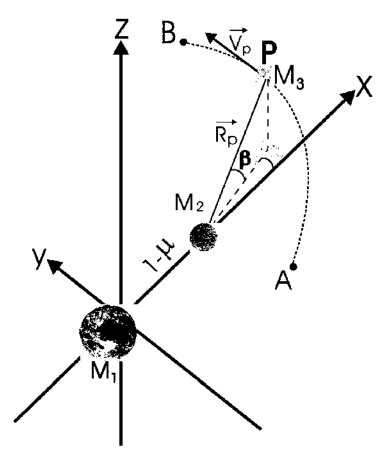

Figure 2.1. The swing-by in the three-dimensional space.

Figure 2.1 shows the sequence for this maneuver and some of those and other important variables. It is assumed that the system has three bodies: a primary (M_1) and a secondary (M_2) body with finite masses that are in circular orbits around their common center of mass and a third body with negligible mass (the spacecraft) that has its motion governed by the two other bodies. The spacecraft leaves the point A, passes by the point P (the periapsis of the trajectory of the spacecraft in its orbit around M_2) and goes to the point B. The points A and B are chosen in a such way that the influence of M_2 at those two points can be neglected and, consequently, the energy can be assumed to remain constant after B and before A (the system follows the two-body celestial mechanics). The initial conditions are clearly identified in the Figure 2.1: the periapsis distance R_p (distance measured between the point P and the center of M_2), the angles α and β and the velocity V_p. The distance R_p is not to scale, to make the figure easier to understand. The result of this maneuver is a change in velocity, energy and angular momentum in the keplerian orbit of the spacecraft around the central body.

3 The Three-Dimensional Circular Restricted Problem

For the research performed in this paper, the equations of motion for the spacecraft are assumed to be the ones valid for the well-known three-dimensional restricted circular three-body problem. The standard dimensionless canonical system of units is used, which implies that: the unit of distance is the distance between M_1 and M_2; the mean angular velocity (ω) of the motion of M_1 and M_2 is assumed to be one; the mass of the smaller primary (M_2) is given by $\mu = \frac{m_2}{m_1+m_2}$ (where m_1 and m_2 are the real masses of M_1 and M_2, respectively) and the mass of M_2 is $(1-\mu)$; the unit of time is defined such that the period of the motion of the two primaries is 2π and the gravitational constant is one. There are several systems of reference that can be used to describe the three-dimensional restricted three-body problem [10; Chapter 10]. In this paper the rotating system is used. In this system of reference, the origin is the center of mass of the two massive primaries. The horizontal axis (x) is the line that connects the two primaries at any time. It rotates with a variable angular velocity in a such way that the two massive primaries are always on this axis. The vertical axis (y) is perpendicular to the (x) axis. In this system, the positions of the primaries are: $x_1 = -\mu$, $x_2 = 1-\mu$, $y_1 = y_2 = 0$. In this system, the equations of motion for the massless particle are [10; Chapter 10]:

$$\ddot{x} - 2\dot{y} = x - (1-\mu)\,\frac{x+\mu}{r_1^3} - \mu\,\frac{x-1+\mu}{r_2^3}, \tag{1}$$

$$\ddot{y} + 2\dot{x} = y - (1-\mu)\,\frac{y}{r_1^3} - \mu\,\frac{y}{r_2^3}, \tag{2}$$

$$\ddot{z} = -(1-\mu)\,\frac{z}{r_1^3} - \mu\,\frac{z}{r_2^3}, \tag{3}$$

where r_1 and r_2 are the distances from M_1 and M_2.

4 Algorithm to Solve the Problem

A numerical algorithm to solve the problem has the following steps: 1. Arbitrary values for the three parameters R_p, V_p, α and β are given; 2. With these values the initial conditions in the rotating system are computed. The initial position is the point (X_i, Y_i, Z_i) and the initial velocity is (V_{Xi}, V_{Yi}, V_{Zi}), where:

$$X_i = 1 - \mu + R_p \cos(\beta)\cos(\alpha), \tag{4}$$

$$Y_i = R_p \cos(\beta)\sin(\alpha), \tag{5}$$

$$Z_i = R_p \sin(\beta), \tag{6}$$

$$V_{Xi} = -V_p \sin(\alpha) + R_p \cos(\beta)\sin(\alpha), \tag{7}$$

$$V_{Yi} = V_p \cos(\alpha) - R_p \cos(\beta)\cos(\alpha), \tag{8}$$

$$V_{Zi} = 0, \tag{9}$$

where the last equation comes from the decision of studying the maneuvers with V_p parallel to the orbital plane of the primaries; 3. With these initial conditions, the equations of motion are integrated forward in time until the distance between M_2 and the spacecraft is larger than a specified limit d. At this point the numerical integration is stopped and the energy (E_+) and the angular momentum (C_+) after the encounter are calculated; 4. Then, the particle goes back to its initial conditions at the point P, and the equations of motion are integrated backward in time, until the distance d is reached again. Then the energy (E_-) and the angular momentum (C_-) before the encounter are calculated.

For all the simulations shown, a Runge–Kutta of 8^{th} order was used for numerical integration. The criteria to stop numerical integration is the distance between the spacecraft and M_2. When this distance reaches the value $d = 0.5$ (half of the semimajor axis of the two primaries) the numerical integration is stopped. The value 0.5 is larger than the sphere of influence of M_2, which avoids any important effects of M_2 at these points. Simulations using larger values for this distance were performed, and it increased the integration time, but did not significantly change the results. To study the effects of numerical accuracy, several cases were simulated using different integration methods and/or different values for the accuracy required with no effects in the results.

With this algorithm available, the given initial conditions (values of R_p, V_p, α, β) are varied in any desired range and the effects of the close approach in the orbit of the spacecraft are studied.

5 Classification of the Orbits

The main results consist of plots that show the change of the orbit of the spacecraft, due to the close encounter with M_2. The Earth–Moon, Sun–Uranus and the Sun–Saturn systems of primaries are used. Any mission using a swing-by with one of those system can use those results. First of all, it is necessary to classify all the close encounters between M_2 and the spacecraft, according to the change obtained in the orbit of the spacecraft. The letters A – P are used for this classification. They are assigned to the orbits according to the rules showed in Table 5.1.

With those rules defined, the results consist of assigning one of those letters to a position in a two-dimensional diagram that has the angle α (in degrees) in the vertical axis and the angle β (in degrees) in the horizontal axis. One plot is made for every value of R_p and V_p. This type of diagram is called here a “letter-plot” and it was used before in [2].

Table 5.1. Rules to assign letters to orbits.

After / Before	Direct Ellipse	Retrograde Ellipse	Direct Hyperbola	Retrograde Hyperbola
Direct Ellipse	A	E	I	M
Retrograde Ellipse	B	F	J	N
Direct Hyperbola	C	G	K	O
Retrograde Hyperbola	D	H	L	P

In the present paper several simulations were made and they are shown in Figures 5.1 – 5.3. For each plot a total of 961 trajectories were generated, dividing each axis in 31 segments. The interval plotted for α is $180 \le \alpha \le 360\,deg$ because there is a symmetry with respect to the vertical line $\alpha = 180\,deg$. The plot for the interval $0 \le \alpha \le 180\,deg$ is a mirror image of the region $180 \le \alpha \le 360\,deg$ with the following letter substitutions: L becomes O, N becomes H, I becomes C, B becomes E, M becomes D and J becomes G. The letters K, P, F and A remain unchanged.

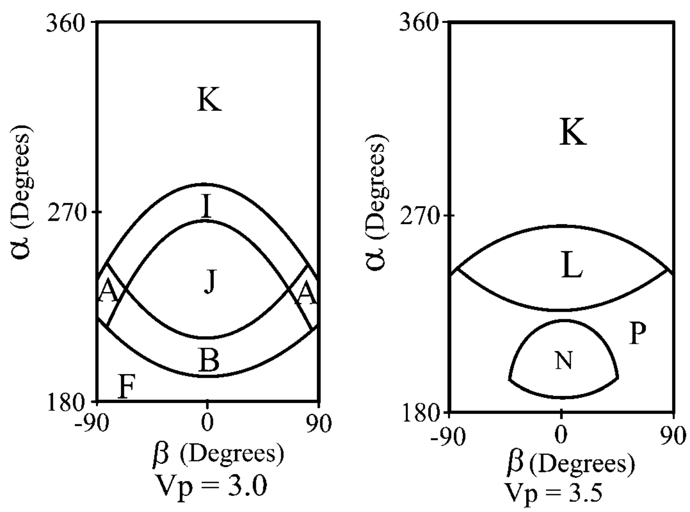

Figure 5.1. Simulations for $R_p = 0.00008464$ in the Sun–Saturn system.

By examining Figures 5.1 – 5.3 it is possible to identify the following families of orbits: a) Orbits that result in an escape (transfer from elliptic to hyperbolic), that are represented by the letters I, J, M, N and that appear between the center ($\alpha = 270°$) and the bottom part of some of the plots (the ones for lower velocities); b) Orbits that result in a capture (transfer from hyperbolic to elliptic), that are represented by the letters C, D, G, H that do not appear in the plots shown in this paper (but exist in the symmetric part not shown here); c) Elliptic orbits (transfer from elliptic to elliptic), that are represented by the letters A, B, E, F and that appear at the bottom of some of the plots (the ones for lower velocities); d) Hyperbolic orbits (transfer from hyperbolic to hyperbolic), that

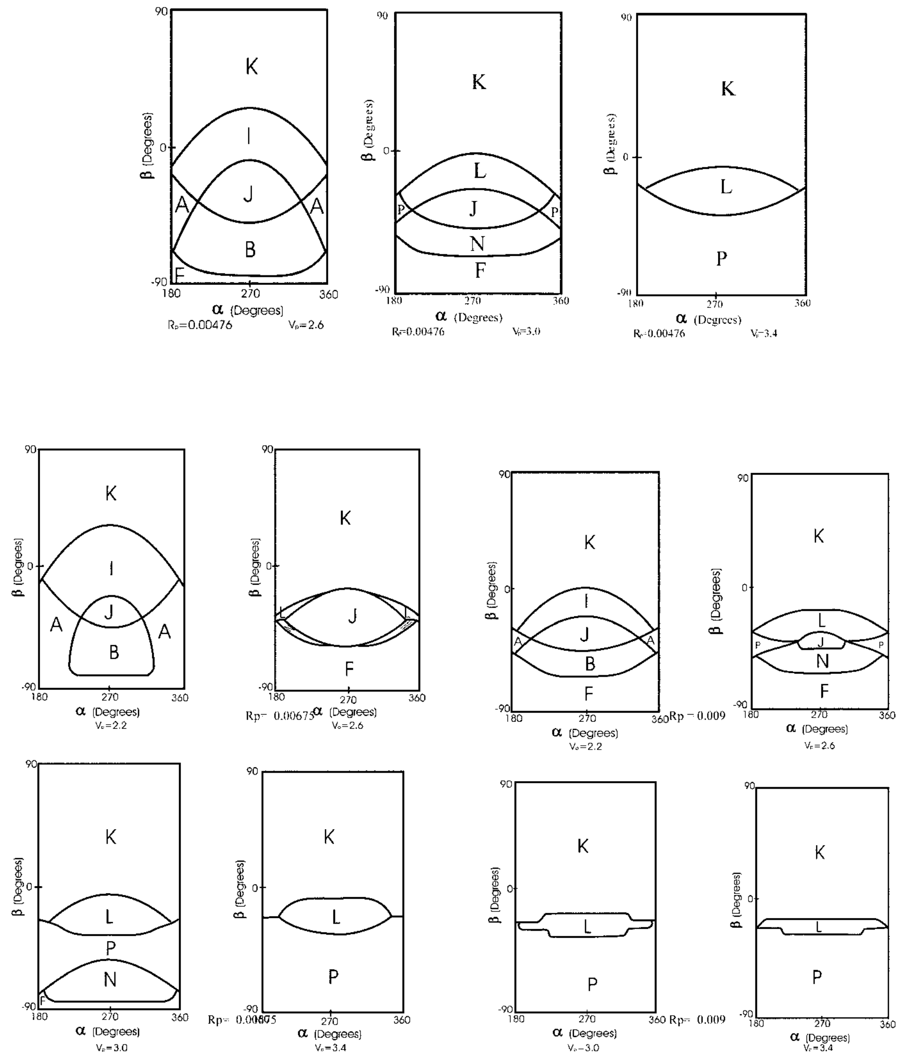

Figure 5.2. Simulations in the Earth–Moon system.

are represented by the letters K, L, O, P and that appears at the upper part of the plots and that are the only families available for higher velocities; e) Orbits that change the direction of motion from direct to retrograde that are represented by the letters E, M, G, O and that do not appear in the plots shown in this paper (but exist in the symmetric part not shown here); f) Orbits that change the direction of motion from retrograde to direct, that are represented by the letters B, D, J, L, that appear in the lower-center of the plot; g) Retrograde orbits that are represented by the letters F, H, N, P that appear in the majority of the bottom part of the plots; h) Direct orbits that are represented by the letters A, C, I, K that appear in the top of the plots.

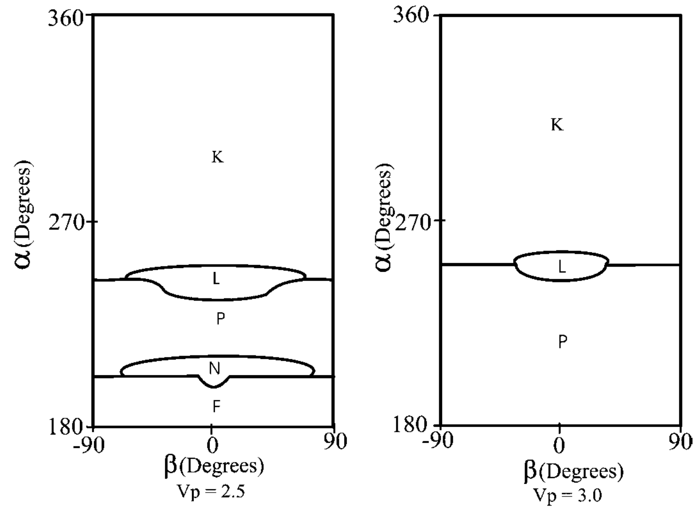

Figure 5.3. Simulations for $R_p = 0.000082$ in the Sun–Uranus system.

6 Optimal Problems

The results generated in this research can be used to help mission designers to plan missions that involve optimization of parameters. It is possible to use the plots made here to find situations where a specific case (represented by the letters A – P) can be obtained with one or more variables (like V_p or R_p) extremized. The parameters V_p and R_p are important parameters to be extremized. If the goal of the mission is to collect data from M_2, it is interesting to minimize R_p (to get closer to M_2) and V_p (to stay more time close to M_2). In the opposite, if M_2 is necessary to be used to change the trajectory of the spacecraft, but it represents a risk to the vehicle due to the presence of an atmosphere and/or radiation, etc., it is necessary to maximize R_p and/or V_p, subject to the restriction of obtaining the desired change in the trajectory. To use a real case as an example, the Earth–Moon, Sun–Saturn and the Sun–Uranus systems are used to solve the problems described below.

Problem 1: It is desired to find a trajectory of type N (a retrograde escape) in the Earth–Moon system, subject to the constraints $V_p = 3.0$ and requiring that R_p is maximized. Figure 5.2 shows that the trajectory type N, in the case $V_p = 3.0$, appear for $R_p = 0.00476$ and $R_p = 0.00675$, but do not appear for $R_p = 0.009$. Figure 6.1 shows plots of the sequence made to find the solution. The solution to this problem is $R_p = 0.0075234375$. The complete values for the set of variables are: $\alpha = 192°$; $\beta = 0°$.

Problem 2: It is desired to find a trajectory of type B (an ellipse that changes the motion from retrograde to direct) in the Sun–Saturn system, subject to the constraints $R_p = 0.00008464$ (2.0 radius of Saturn) and requiring that the velocity at periapsis be a maximum. Figure 5.1 shows that the trajectory type B appears for $V_p = 3.0$, but do not appear for $V_p = 3.5$. To find the solution, plots were made for several values of V_p in this interval. Figure 6.2 shows two plots of this sequence. The solution to this problem is $V_p = 3.12$, since for $V_p = 3.13$, B does not occur anymore. It is also possible to see that this problem has four solutions: $\alpha = 216°$, $\beta = -54°$; $\alpha = 210°$, $\beta = -24°$; $\alpha = 210°$, $\beta = 24°$; $\alpha = 216°$, $\beta = 54°$.

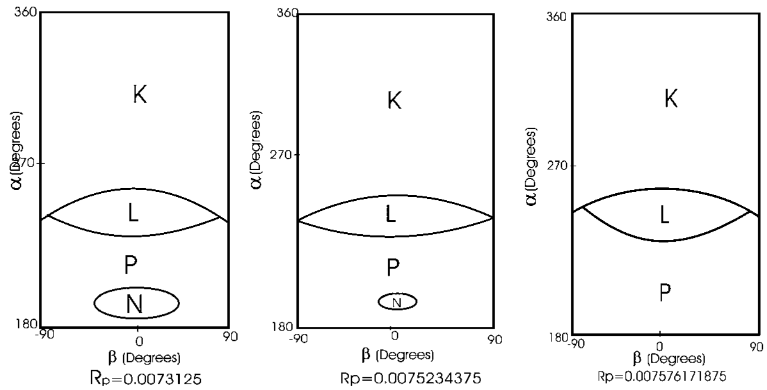

Figure 6.1. Solution for the Problem 1 in the Earth–Moon system.

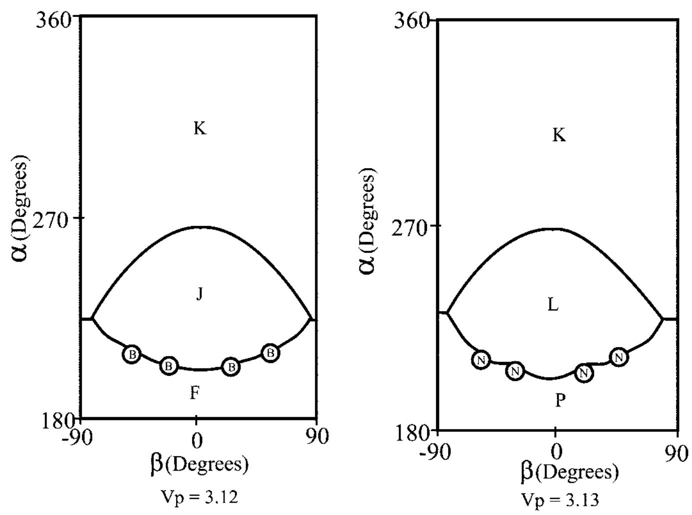

Figure 6.2. Solution for the Problem 2 for $R_p = 0.00008464$ in the Sun–Saturn system.

Problem 3: It is desired to obtain a trajectory of type N (a retrograde ellipse before the swing-by and a retrograde hyperbola after) in the Sun–Uranus system, subject to the constraints $R_p = 0.000082$ (10.0 radius of Uranus) and requiring that the velocity at periapsis be a maximum. Figure 5.3 shows that the trajectory type N appears for $V_p = 2.5$, but do not appear for $V_p = 3.0$. To find the solution, plots were made for several values of V_p in this interval. Figure 6.3 shows two plots of this sequence. The solution to this problem is $V_p = 2.62$, since for $V_p = 2.63$, N does not occur anymore. In this example, it is possible to see that there is a range of values of β that allows solutions. So, the complete values for the set of variables are: $-48° \leq \beta \leq 48°$; $\alpha = +186°$.

This information constitutes a set of initial conditions to design the trajectory. Several improvements can be made: 1) more plots can be generated to get more accuracy for the data, in particular in the solutions of the optimal problems; 2) many other types of

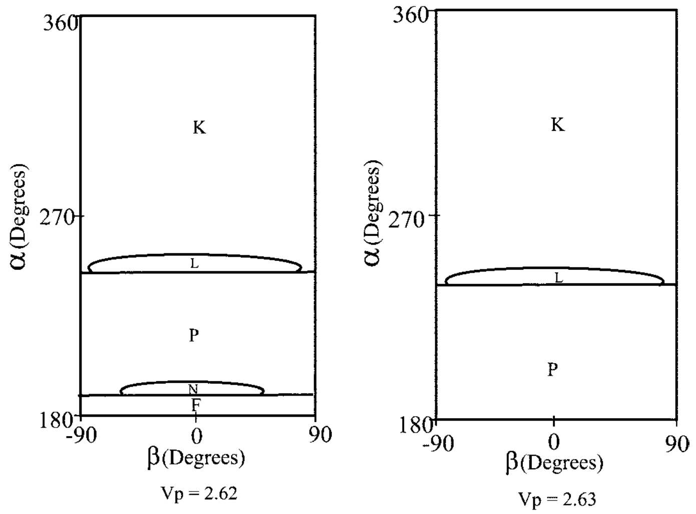

Figure 6.3. Solution for the Problem 3 for $R_p = 0.000082$ in the Sun–Uranus system.

optimization problems can be solved, combining different constraints and/or variables to be extremized; 3) others systems can be used; etc.

7 Conclusions

In this paper the three-dimensional restricted three-body problem is described and used to study the swing-by maneuver. Several letter-plot type of graphics are made to represent the effect of a close approach in the orbit of a spacecraft. In particular, the effects of the third dimension in this maneuver are studied. It is shown that the hyperbolic orbits (family K) dominate the region where α is larger than 270° and that when the velocity increases, the families K, L and P dominate the plots. Families with particularities, like parabolic or zero angular momentum orbits, are shown to exist in the borders between the main families. After that, the results available here were used in the solution of optimal problems. In this type of problem, it is necessary to find the initial conditions that generates a given orbit change, subject to the extremization of some parameters like V_p or R_p.

References

[1] Broucke, R.A. and Prado, A.F.B.A. Jupiter swing-by trajectories passing near the Earth. *Advances in the Astronautical Sciences* **82**(2) (1993) 1159–1176.

[2] Broucke, R.A. The celestial mechanics of gravity assist. In: *AIAA/AAS Astrodynamics Conference, Minneapolis, MN, 15–17 August*, Paper 88-4220, 1988.

[3] D'Amario, L.A., Byrnes, D.V. and Stanford, R.H. A new method for optimizing multiple-flyby trajectories. *Journal of Guidance, Control, and Dynamics* **4**(6) (1981) 591–596.

[4] D'Amario, L.A., Byrnes, D.V. and Stanford, R.H. Interplanetary trajectory optimization with application to galileo. *Journal of Guidance, Control, and Dynamics* **5**(5) (1982) 465–471.

[5] Farquhar, R.W. and Dunham, D.W. A new trajectory concept for exploring the Earth's geomagnetic tail. *Journal of Guidance, Control and Dynamics* **4**(2) (1981) 192–196.

[6] Felipe, G. and Prado, A.F.B.A. Classification of out of plane swing-by trajectories. *Journal of Guidance, Control and Dynamics* **22**(5) (1999) 643–649.

[7] Flandro, G. Fast reconnaissance missions to the outer solar system utilizing energy derived from the gravitational field of Jupiter. *Astronautical Acta* **12**(4) (1966) 329–337.

[8] Muhonen, D., Davis, S. and Dunham, D. Alternative gravity-assist sequences for the ISEE-3 escape trajectory. *Journal of Astronautical Sciences* **33**(3) (1985) 255–273.

[9] Prado, A.F.B.A. Close-approach trajectories in the elliptic restricted problem. *Journal of Guidance, Control, and Dynamics* **20**(4) (1997) 797–802.

[10] Szebehely, V. *Theory of Orbits*. Academic Press, New York, 1967.

Nonlinear Dynamics and Systems Theory, **5**(1) (2005) 49–59

A Constant-Gain Nonlinear Estimator for Linear Switching Systems

S. Ibrir and E.K. Boukas

Mechanical Engineering Department, École Polytechnique de Montréal,
P.O. Box 6079, station "Centre-ville", Montréal, Québec, Canada H3C 3A7

Received: December 19, 2003; Revised: June 1, 2004

Abstract: In this paper, we develop a nonlinear observer for switching linear systems. The developed state observer offers the properties of being self-tuning without any information about the switching instants and the current mode of the switching system to be observed. Numerical example is provided to highlight the usefulness of the developed results.

Keywords: *Estimation; switching systems; nonlinear observers; LMIs.*

Mathematics Subject Classification (2000): 93B36, 93C05, 93E03; 93E15.

1 Introduction

Estimation of switching systems has rapidly increased in importance with the development of new circuits technologies. Recently, we witnessed an increasing interest in the so-called *switching systems*. We call herein switching systems all dynamical systems described by differential inclusions of the form

$$\dot{x}(t) \in \{f_\sigma(x(t),\, u(t))\}_{\sigma \in \mathcal{A}}\,,$$

where $x(t)$ is the state variable, $u(t)$ is the control input, and $f_\sigma(\,\cdot\,,\,\cdot\,)$ is a collection of continuously differentiable functions parameterized by σ belonging to some given set $\mathcal{A}$. Such systems are composed of both discrete and continuous subsystems. Control, observation, and supervision of this kind of systems appear in many ongoing research projects such as multimedia protocols, electrical circuits, systems subject to failure and so on.

Numerous control procedures are based on the knowledge of all state variables of the considered system. This assumption is not always true since the measurements of the states variables are, in most cases, not possible or simply too expensive. For this reason, observer design has received widespread attention since the introduction of Kalman theory and remains of great importance nowadays.

Estimation of hybrid systems is one of the challenging research problems that necessitate a particular attention. Extension of available results in observation of linear systems to hybrid linear systems is not quite easy due to the variation of nominal models and others technical problems. Switching between different models to compensate or analyze system variations is a well-known technique in modern control theory. It is obvious that if both the switching instants and the switching modes are known, then it is easy to construct a switching gain observer that switches among different gains. We refer the reader to the references [1−5], and [6] for more details.

The question we are addressing in this paper is how one can estimate the unmeasured states of a given switching system if the current mode is unknown? The answer to this question will be detailed in the present work where we assume that there is no switching law that defines the passage of the switching system from a mode to another. The goal of this paper is to develop a new observation technique for switching linear systems. The developed observers are nonlinear and do not necessitate the mode estimation of the system to be observed. We mean by mode estimation, the ability to track a system's discrete dynamics as it moves between different behavioral modes. We show that a constant high-gain observer is sufficient to observe the unmeasured dynamics whatever the changes in the nominal matrices of the considered switching system. The present work eliminates two major frequently-faced problems: detection of the switching instants and identification of the current mode. The whole observer design is efficiently accomplished by using an LMI procedure.

The paper is organized as follows. Section 2 is devoted to the design of the observer for regular switching systems. In Section 3, the results of the previous section are then extended to uncertain switching systems. Section 4 treats a numerical observation example of a switching system. The paper ends with general conclusions and some concluding remarks. Throughout this paper, we note by I and 0 the identity matrix and the null matrix of appropriate dimensions, respectively. $A > 0$ (resp. $A < 0$) denotes that the matrix A is a symmetric and positive-definite (resp. symmetric and negative-definite). We note by A' the matrix transpose of the matrix A. $\|\cdot\|$ stands for the Euclidean norm.

2 Constant-Gain Observer for Switching Systems

Our objective is to conceive an observer for the following switching system

$$\begin{aligned} \frac{dx(t)}{dt} &= A(\sigma(t))x(t) + B(\sigma(t))u(t), \\ y(t) &= Cx(t), \end{aligned} \tag{2}$$

where $x(t) \in R^n$ is the state vector, $u(t) \in R^m$ is the control input, and $y(t) \in R^p$ is the system output. $\sigma(t)$ is a switching signal that maps the index time $[0, +\infty[$ into an index set $\mathcal{S} = \{1, 2, \ldots, s\}$. Each mode $j \in \mathcal{S}$ corresponds to a specific model characterized by $A(j) \in \mathcal{A} = \{A(1), A(2), \ldots, A(s)\}$ and $B(j) \in \mathcal{B} = \{B(1), B(2), \ldots, B(s)\}$. We assume that the switchs in the output matrix C are absent. For the observer design, we suppose that the following assumptions are verified.

Assumption 1 The switch between two different modes is instantaneous and arbitrary.

Assumption 2 There is no information on the current mode of the switching system, and the switching instants are not known.

Assumption 3 For any time t, the control input $u(t)$ is smooth, i.e., it can be written as

$$u(t) = \int_0^t v(\tau)\, d\tau. \tag{3}$$

where $v(t) \in R^m$ is the new control input.

For the class of systems we are considering, different types of observability have been studied in the past and for more details on this subject, we refer the reader to [1] and the references therein. Here, we will assume that the pairs $(A(\sigma(t)), C)$, $\forall\, \sigma(t)$ are observable. This means that the system is observable, in the sense of Kalman, for each mode.

Based on the last assumptions, the switching system is rewritten in the following form:

$$\begin{aligned} \frac{dx(t)}{dt} &= A(\sigma(t))x(t) + B(\sigma(t))u(t), \\ \frac{du(t)}{dt} &= v(t), \\ y(t) &= Cx(t). \end{aligned} \tag{4}$$

For the simplicity of the representation, let

$$\widetilde{A}(\sigma(t)) = \begin{bmatrix} A(\sigma(t)) & B(\sigma(t)) \\ 0 & 0 \end{bmatrix}, \qquad \widetilde{B} = \begin{bmatrix} 0 \\ I \end{bmatrix},$$

$$z = \begin{bmatrix} x(t) \\ u(t) \end{bmatrix}, \qquad \widetilde{C} = [\, C \quad 0\,],$$

then the dynamics (4) is rewritten as:

$$\begin{aligned} \frac{dz(t)}{dt} &= \widetilde{A}(\sigma(t))z + \widetilde{B}v(t), \\ y(t) &= \widetilde{C}z(t). \end{aligned} \tag{5}$$

We propose an observer of the following form:

$$\frac{d\hat{z}(t)}{dt} = \left(\sum_{j=1}^{s} \widetilde{A}(j)\right)\hat{z}(t) + \widetilde{B}v(t) + \left(\sum_{i=1}^{s} P_i\right)^{-1} Y\left(y(t) - \widetilde{C}\hat{z}(t)\right) - \rho(y(t), \hat{z}(t)), \tag{6}$$

where $P_1, P_2, \dots, P_s$ are $(m+n) \times (n+m)$ symmetric and positive definite matrices, Y is a constant matrix of appropriate dimensions, and $\rho(y(t),\, \hat{z}(t))$ is a nonlinear additive term that depends on the output $y(t)$, and the observer state vector $\hat{z}(t)$. The dynamics of the proposed observer is the sum of the dynamics of classical Luemberger observers written for each mode plus a nonlinear additive term $\rho(\cdot,\cdot)$ that attenuates the effects of the difference between the observer and the system outputs. The design of $(P_i)_{1\le i\le s}$, and $\rho(\cdot,\cdot)$ will be given latter. Let $e(t)=\hat{z}(t)-z(t)$ be the observation error, and let

$$\frac{de(t)}{dt}=\left(\sum_{j=1}^{s}\widetilde{A}(j)\right)\hat{z}(t)-\widetilde{A}(\sigma(t))z(t)-\rho(y(t),\hat{z}(t))-\left(\sum_{i=1}^{s}P_i\right)^{-1}Y\widetilde{C}e(t),$$

be the dynamics of the observer error, then we can write

$$\frac{de(t)}{dt}=\left(\widetilde{A}(\sigma(t))-\left(\sum_{i=1}^{s}P_i\right)^{-1}Y\widetilde{C}\right)e(t)+\sum_{\substack{j\in\mathcal{S}\\ j\neq\sigma}}\widetilde{A}(j)\hat{z}(t)-\rho(y(t),\hat{z}(t)). \tag{7}$$

The time derivative of the Lyapunov function $V(e(t))=e^{\mathrm{T}}(t)\left(\sum_{i=1}^{s}P_i\right)e(t)$ along the trajectory of (7) is

$$\begin{aligned}\frac{dV(e(t))}{dt}&=\frac{de^{\mathrm{T}}(t)}{dt}\left(\sum_{i=1}^{s}P_i\right)e(t)+e^{\mathrm{T}}(t)\left(\sum_{i=1}^{s}P_i\right)\frac{de(t)}{dt}\\
&=e^{\mathrm{T}}(t)\left(\widetilde{A}^{\mathrm{T}}(\sigma(t))\left(\sum_{i=1}^{s}P_i\right)+\left(\sum_{i=1}^{s}P_i\right)\widetilde{A}(\sigma(t))-\widetilde{C}^{\mathrm{T}}Y^{\mathrm{T}}-Y\widetilde{C}\right)e(t)\\
&\quad-2e^{\mathrm{T}}(t)\left(\sum_{i=1}^{s}P_i\right)\rho(y(t),\hat{z}(t))+e^{\mathrm{T}}(t)\left(\sum_{i=1}^{s}P_i\right)\sum_{\substack{j\in\mathcal{S}\\ j\neq\sigma}}\widetilde{A}(j)\hat{z}(t)\\
&\quad+\hat{z}^{\mathrm{T}}(t)\sum_{\substack{j\in\mathcal{S}\\ j\neq\sigma}}\widetilde{A}^{\mathrm{T}}(j)\left(\sum_{i=1}^{s}P_i\right)e(t).\end{aligned}$$

We have for arbitrary vectors w_1 and w_2 and a given positive definite matrix Z of appropriate dimensions [7]

$$2w_1^{\mathrm{T}}w_2\le w_1^{\mathrm{T}}Z^{-1}w_1+w_2^{\mathrm{T}}Zw_2.$$

If we take

$$w_1=\left(\sum_{i=1}^{s}P_i\right)e(t),\quad w_2=\sum_{\substack{j\in\mathcal{S}\\ j\neq\sigma}}\widetilde{A}(j)\hat{z}(t),\quad Z=\mu_\sigma I,$$

then

$$\begin{aligned}&e^{\mathrm{T}}(t)\left(\sum_{i=1}^{s}P_i\right)\sum_{\substack{j\in\mathcal{S}\\ j\neq\sigma}}\widetilde{A}(j)\hat{z}(t)+\hat{z}^{\mathrm{T}}(t)\sum_{\substack{j\in\mathcal{S}\\ j\neq\sigma}}\widetilde{A}^{\mathrm{T}}(j)\left(\sum_{i=1}^{s}P_i\right)e(t)\\
&\quad=2e^{\mathrm{T}}(t)\left(\sum_{i=1}^{s}P_i\right)\sum_{\substack{j\in\mathcal{S}\\ j\neq\sigma}}\widetilde{A}(j)\hat{z}(t)\\
&\le\mu_\sigma^{-1}e^{\mathrm{T}}(t)\left(\sum_{i=1}^{s}P_i\right)\left(\sum_{i=1}^{s}P_i\right)e(t)+\mu_\sigma\hat{z}^{\mathrm{T}}(t)\sum_{\substack{j\in\mathcal{S}\\ j\neq\sigma}}\widetilde{A}^{\mathrm{T}}(j)\sum_{\substack{j\in\mathcal{S}\\ j\neq\sigma}}\widetilde{A}(j)\hat{z}(t).\end{aligned}$$

If the matrices $(P_i)_{1\le i\le s}$ are selected so as to

$$\begin{aligned}&\widetilde{A}^{\mathrm{T}}(\sigma(t))\Big(\sum_{i=1}^{s}P_i\Big)+\Big(\sum_{i=1}^{s}P_i\Big)\widetilde{A}(\sigma(t))-\widetilde{C}^{\mathrm{T}}Y^{\mathrm{T}}-Y\widetilde{C}\\&\quad+\mu_\sigma^{-1}\Big(\sum_{i=1}^{s}P_i\Big)\Big(\sum_{i=1}^{s}P_i\Big)=-Q(\sigma)<0,\end{aligned}\tag{8}$$

then we obtain

$$\frac{dV(e(t))}{dt}=-e^{\mathrm{T}}(t)Q(\sigma)e(t)+\mu\hat{z}^{\mathrm{T}}(t)\sum_{\substack{j\in\mathcal{S}\\ j\neq\sigma}}\widetilde{A}^{\mathrm{T}}(j)\sum_{\substack{j\in\mathcal{S}\\ j\neq\sigma}}\widetilde{A}(j)\hat{z}(t)-\rho(y(t),\hat{z}(t)).$$

If we choose $\mu_{\max}=\max\limits_{\sigma}\mu_\sigma$ and

$$\rho(y(t),\hat{z}(t))=\begin{cases}\mu_{\max}\varpi\hat{z}^{\mathrm{T}}(t)\hat{z}(t)\dfrac{\Big(\sum\limits_{i=1}^{s}P_i\Big)^{-1}\widetilde{C}^{\mathrm{T}}\widetilde{C}e(t)}{2\|\widetilde{C}e(t)\|^2} & \text{if}\quad \|\widetilde{C}e(t)\|\neq 0,\\ 0 & \text{if}\quad \|\widetilde{C}e(t)\|=0,\end{cases}\tag{9}$$

where

$$\varpi=\sup_{\sigma(t)}\left\|\sum_{\substack{j\in\mathcal{S}\\ j\neq\sigma}}\widetilde{A}^{\mathrm{T}}(j)\sum_{\substack{j\in\mathcal{S}\\ j\neq\sigma}}\widetilde{A}(j)\right\|,\tag{10}$$

then

$$\frac{dV(e(t))}{dt}\le -e^{\mathrm{T}}(t)Q(\sigma)e(t),$$

which implies that the observer error decays exponentially to the origin.

Remark 1 The formulae of $\rho(\cdot,\cdot)$ given by equation (9) is just a conceptual one. When the observation error is close to zero, it is recommended to modify the nonlinear term $\rho(\cdot,\cdot)$ as follows:

$$\rho(y(t),\hat{z}(t))=\begin{cases}\mu_{\max}\varpi\hat{z}^{\mathrm{T}}(t)\hat{z}(t)\dfrac{\Big(\sum\limits_{i=1}^{s}P_i\Big)^{-1}\widetilde{C}^{\mathrm{T}}\widetilde{C}e(t)}{2\|\widetilde{C}e(t)\|^2} & \text{if}\quad \|\widetilde{C}e(t)\|>\bar{\epsilon},\\ 0 & \text{if}\quad \|\widetilde{C}e(t)\|\le\bar{\epsilon},\end{cases}$$

where $\bar{\epsilon}>0$ is some prescribed small parameter. We summarize the result in the following statement.

Theorem 1 *System*

$$\frac{\hat{z}(t)}{dt}=\left(\sum_{j=1}^{s}\widetilde{A}(j)\right)\hat{z}(t)+\widetilde{B}v(t)+\left(\sum_{i=1}^{s}P_i\right)^{-1}Y\left(y(t)-\widetilde{C}\hat{z}(t)\right)-\rho(y(t),\hat{z}(t)),\tag{11}$$

is an asymptotic observer for system (5) if there exist a set of positive constants $\mathcal{M} = \{\mu_1, \mu_2, \dots, \mu_s\}$*, and a set of symmetric and positive definite matrices* $\mathcal{P} = \{P_1, P_2, \dots, P_s\}$ *such that the following coupled LMIs are feasible*

$$\begin{bmatrix} \mathcal{J}(P_1, \dots, P_s, Y, j) & \left(\sum\limits_{i=1}^{s} P_i\right) \\ \left(\sum\limits_{i=1}^{s} P_i\right) & -\mu_j I \end{bmatrix} < 0, \quad 1 \leq j \leq s,$$

where

$$\mathcal{J}(P_1, \dots, P_s, Y, j) = \widetilde{A}^{\mathrm{T}}(j)\left(\sum_{i=1}^{s} P_i\right) + \left(\sum_{i=1}^{s} P_i\right)\widetilde{A}(j) - \widetilde{C}^{\mathrm{T}} Y^{\mathrm{T}} - Y\widetilde{C}.$$

Proof The LMIs conditions (12) and (8) are equivalent by the Schur complement lemma.

3 Extension to Uncertain Switching Systems

Consider the uncertain switching system

$$\begin{aligned} \frac{dx(t)}{dt} &= (A(\sigma(t)) + \Delta A(\sigma(t)))x(t) + (B(\sigma(t)) + \Delta B(\sigma(t)))u(t), \\ \frac{du(t)}{dt} &= v(t), \\ y(t) &= Cx(t), \end{aligned} \tag{13}$$

which satisfies the assumptions of system (4). The aim of this section is to design a robust nonlinear observer that can estimates the states of (13) without a priori knowledge of the current mode and or the switching instants. The uncertain terms $\Delta A(\sigma(t))$ and $\Delta B(\sigma(t))$ are written respectively as $E_A^{\mathrm{T}} F_A(\sigma(t)) D_A$ and $E_B^{\mathrm{T}} F_B(\sigma(t)) D_B$. he matrices E_A, E_B, D_A, and D_B are constant known matrices and $F_A(\sigma(t))$, $F_B(\sigma(t))$ are unknown matrices satisfying the inequalities $F_A^{\mathrm{T}}(\sigma(t))F_A(\sigma(t)) < I$, $F_B^{\mathrm{T}}(\sigma(t))F_B(\sigma(t)) < I$, respectively. In matrix notation system (13) is rewritten as

$$\begin{aligned} \frac{d\xi(t)}{dt} &= \left(\widetilde{A}(\sigma(t)) + \Delta\widetilde{A}(\sigma(t))\right)\xi(t) + \widetilde{B}v(t), \\ y &= \widetilde{C}\xi(t), \end{aligned} \tag{14}$$

where

$$\Delta\widetilde{A}(\sigma(t)) = \begin{bmatrix} \Delta A(\sigma(t)) & \Delta B(\sigma(t)) \\ 0_{m\times n} & 0_{m\times m} \end{bmatrix}, \quad \xi(t) = \begin{bmatrix} x(t) \\ u(t) \end{bmatrix}, \tag{15}$$

and $\widetilde{A}(\sigma(t))$, $\widetilde{B}$, $\widetilde{C}$ are defined as in equation (5). The uncertain term $\Delta\widetilde{A}(\sigma(t))$ can be rewritten as $\widetilde{E}_A^{\mathrm{T}} \widetilde{F}_A(\sigma(t)) \widetilde{D}_A$ where

$$\widetilde{E}_A = \begin{bmatrix} E_A & 0 \\ E_B & 0 \end{bmatrix}, \quad \widetilde{F}_A(\sigma(t)) = \begin{bmatrix} F_A(\sigma(t)) & 0 \\ 0 & F_B(\sigma(t)) \end{bmatrix}, \quad \widetilde{D}_A = \begin{bmatrix} D_A & 0 \\ 0 & D_B \end{bmatrix}.$$

The observer design is given in the following statement:

Theorem 2 *Consider system (13). If there exist a set of* $(n+m)\times(n+m)$ *symmetric and positive-definite matrices* $(P_i)_{1\le i\le s} > 0$*, a matrix* Y *of appropriate dimensions, and positive constants* $(\mu_i)_{1\le i\le s}$*,* $(\epsilon_A(i))_{1\le i\le s}$*,* $(\epsilon_B(i))_{1\le i\le s}$ *such that the following coupled LMIs hold*

$$\begin{bmatrix} \mathcal{K}(P_1, P_2, \dots, P_s, Y) & \widetilde{E}_A\Big(\sum_{i=1}^{s} P_i\Big) & \Big(\sum_{i=1}^{s} P_i\Big) & \Big(\sum_{i=1}^{s} P_i\Big) \\ \Big(\sum_{i=1}^{s} P_i\Big)\widetilde{E}_A{}^{\mathrm{T}} & -\epsilon_A(j)I & 0 & 0 \\ \Big(\sum_{i=1}^{s} P_i\Big) & 0 & -\epsilon_B(j)I & 0 \\ \Big(\sum_{i=1}^{s} P_i\Big) & 0 & 0 & -\mu_j I \end{bmatrix} < 0, \quad 1 \le j \le s, \tag{16}$$

where

$$\mathcal{K}(P_1, P_2, \dots, P_s, Y) = \widetilde{A}^{\mathrm{T}}(j)\Big(\sum_{i=1}^{s} P_i\Big) + \Big(\sum_{i=1}^{s} P_i\Big)\widetilde{A}(j) - \widetilde{C}^{\mathrm{T}}Y^{\mathrm{T}} - Y\widetilde{C} + \epsilon_A(j)\widetilde{D}_A^{\mathrm{T}}\widetilde{D}_A.$$

Then system

$$\frac{\hat{\xi}(t)}{dt} = \Big(\sum_{j=1}^{s}\widetilde{A}(j)\Big)\hat{\xi}(t) + \widetilde{B}v(t) + \Big(\sum_{i=1}^{s} P_i\Big)^{-1} Y\left(y(t) - \widetilde{C}\hat{\xi}(t)\right) - \rho(y(t),\hat{\xi}(t)),$$

is an asymptotic observer for the uncertain switching system (13) where $\rho(y(t),\hat{\xi}(t))$ *is defined as*

$$\rho(y(t),\hat{z}(t)) = \begin{cases} \left(\varpi\mu_{\max} + \epsilon_{\max}\|\widetilde{E}_A\|^2\|\widetilde{D}_A\|^2\right)\hat{\xi}^{\mathrm{T}}(t)\hat{\xi}(t)\dfrac{\Big(\sum_{i=1}^{s} P_i\Big)^{-1}\widetilde{C}^{\mathrm{T}}\widetilde{C}e(t)}{\|\widetilde{C}e(t)\|^2} & \\ \qquad\qquad \text{if} \quad \|\widetilde{C}e(t)\| \ne 0, & \\ 0 \qquad\qquad \text{if} \quad \|\widetilde{C}e(t)\| = 0, & \end{cases}$$

where $\epsilon_{\max} = \max_{\sigma}(\epsilon_A(\sigma))$ *and* ϖ *is defined as in Theorem 1.*

Proof Let $e(t) = \hat{\xi}(t) - \xi(t)$ be the observation error. Then its dynamics is given by

$$\begin{aligned} \frac{de(t)}{dt} &= \Big(\sum_{j=1}^{s}\widetilde{A}(j)\Big)\hat{\xi}(t) - \widetilde{A}(\sigma)\xi(t) - \Delta A(\sigma(t))\xi(t) - \Big(\sum_{i=1}^{s} P_i\Big)^{-1} Y\widetilde{C}e - \rho(y,\hat{\xi}(t)) \\ &= \left(\widetilde{A}(\sigma(t)) - \Big(\sum_{i=1}^{s} P_i\Big)^{-1} Y\widetilde{C} + \Delta\widetilde{A}(\sigma(t))\right)e(t) \\ &\quad + \sum_{\substack{j\in\mathcal{S}\\ j\ne\sigma}}\widetilde{A}(j)\hat{\xi}(t) - \Delta\widetilde{A}(\sigma(t))\hat{\xi}(t) - \rho(y,\hat{\xi}(t)). \end{aligned} \tag{17}$$

Choosing the Lyapunov function as $V(e(t)) = e^{\mathrm{T}}(t)\Big(\sum_{i=1}^{s} P_i\Big)e(t)$, we obtain

$$\begin{aligned}\frac{dV(e(t))}{dt} &= e^{\mathrm{T}}(t)\left(\widetilde{A}^{\mathrm{T}}(\sigma(t)) - \widetilde{C}^{\mathrm{T}}Y^{\mathrm{T}}\left(\sum_{i=1}^{s} P_i\right)^{-1} + \Delta\widetilde{A}^{\mathrm{T}}(\sigma(t))\right)\left(\sum_{i=1}^{s} P_i\right)e(t)\\ &\quad + e^{\mathrm{T}}(t)\left(\sum_{i=1}^{s} P_i\right)\left(\widetilde{A}(\sigma(t)) - \left(\sum_{i=1}^{s} P_i\right)^{-1} Y\widetilde{C} + \Delta\widetilde{A}(\sigma(t))\right)e(t)\\ &\quad + 2e(t)\left(\sum_{i=1}^{s} P_i\right)\sum_{\substack{j\in\mathcal{S}\\ j\neq\sigma}}\widetilde{A}(j)\hat{\xi}(t) + 2e^{\mathrm{T}}(t)\left(\sum_{i=1}^{s} P_i\right)\Delta\widetilde{A}(\sigma(t))\hat{\xi}(t)\\ &\quad - 2e^{\mathrm{T}}(t)\left(\sum_{i=1}^{s} P_i\right)\rho(y,\, \hat{\xi}(t)).\end{aligned}$$

We have for any $\mu_\sigma > 0$

$$\begin{aligned}2e^{\mathrm{T}}(t)\left(\sum_{i=1}^{s} P_i\right)\sum_{\substack{j\in\mathcal{S}\\ j\neq\sigma}}\widetilde{A}(j)\,\hat{\xi}(t) &\le \mu_\sigma^{-1}e^{\mathrm{T}}(t)\left(\sum_{i=1}^{s} P_i\right)\left(\sum_{i=1}^{s} P_i\right)e(t)\\ &\quad + \mu_\sigma\hat{\xi}^{\,\mathrm{T}}(t)\sum_{\substack{j\in\mathcal{S}\\ j\neq\sigma}}\widetilde{A}^{\,\mathrm{T}}(j)\sum_{\substack{j\in\mathcal{S}\\ j\neq\sigma}}\widetilde{A}(j)\hat{\xi}(t),\end{aligned}$$

furthermore,

$$\begin{aligned}&\Delta\widetilde{A}^{\,\mathrm{T}}(\sigma(t))\left(\sum_{i=1}^{s} P_i\right) + \left(\sum_{i=1}^{s} P_i\right)\Delta\widetilde{A}(\sigma(t))\\ &= \widetilde{D}_A^{\mathrm{T}}\widetilde{F}_A^{\mathrm{T}}(\sigma(t))\widetilde{E}_A\left(\sum_{i=1}^{s} P_i\right) + \left(\sum_{i=1}^{s} P_i\right)\widetilde{E}_A^{\mathrm{T}}\widetilde{F}_A\widetilde{D}_A\\ &\le \epsilon_A(\sigma)\widetilde{D}_A^{\mathrm{T}}\widetilde{D}_A + \epsilon_A^{-1}(\sigma)\left(\sum_{i=1}^{s} P_i\right)\widetilde{E}_A^{\mathrm{T}}\widetilde{E}_A\left(\sum_{i=1}^{s} P_i\right).\end{aligned}$$

In addition, we have

$$\begin{aligned}2e^{\mathrm{T}}(t)\left(\sum_{i=1}^{s} P_i\right)\Delta\widetilde{A}(\sigma(t))\hat{\xi}(t) &\le \epsilon_B^{-1}(\sigma)e^{\mathrm{T}}(t)\left(\sum_{i=1}^{s} P_i\right)\left(\sum_{i=1}^{s} P_i\right)e(t)\\ &\quad + \epsilon_B(\sigma)\hat{\xi}^{\,\mathrm{T}}(t)\Delta\widetilde{A}^{\,\mathrm{T}}(\sigma(t))\Delta\widetilde{A}(\sigma(t))\hat{\xi}(t).\end{aligned}$$

Using the definition of $\rho(\cdot,\cdot)$, we obtain

$$\begin{aligned}&\epsilon_B(\sigma)\hat{\xi}^{\,\mathrm{T}}(t)\Delta\widetilde{A}^{\,\mathrm{T}}(\sigma(t))\Delta\widetilde{A}(\sigma(t))\hat{\xi}(t) + \mu_\sigma\hat{\xi}^{\,\mathrm{T}}(t)\sum_{\substack{j\in\mathcal{S}\\ j\neq\sigma}}\widetilde{A}^{\,\mathrm{T}}(j)\sum_{\substack{j\in\mathcal{S}\\ j\neq\sigma}}\widetilde{A}(j)\hat{\xi}(t)\\ &\qquad - 2e^{\mathrm{T}}(t)\left(\sum_{i=1}^{s} P_i\right)\rho(\cdot,\,\cdot)\\ &\le \left(\varpi\mu_{\max} + \epsilon_{\max}\|\widetilde{E}_A\|^2\|\widetilde{D}_A\|^2\right)\hat{\xi}^{\,\mathrm{T}}(t)\hat{\xi}(t) - 2e^{\mathrm{T}}(t)\left(\sum_{i=1}^{s} P_i\right)\rho(\cdot,\cdot) \le 0.\end{aligned}$$

This implies that

$$\frac{dV(e(t))}{dt} \le e^{\mathrm{T}}(t)\Bigg(\widetilde{A}^{\mathrm{T}}(\sigma(t))\Big(\sum_{i=1}^{s}P_i\Big) + \Big(\sum_{i=1}^{s}P_i\Big)\widetilde{A}(\sigma(t)) - \widetilde{C}^{\mathrm{T}}Y^{\mathrm{T}} - Y\widetilde{C}$$
$$+ \epsilon_A(\sigma)\widetilde{D}_A'\widetilde{D}_A + \epsilon_B^{-1}(\sigma)\Big(\sum_{i=1}^{s}P_i\Big)\Big(\sum_{i=1}^{s}P_i\Big) + \mu_\sigma^{-1}\Big(\sum_{i=1}^{s}P_i\Big)\Big(\sum_{i=1}^{s}P_i\Big)\Bigg)e(t). \tag{18}$$

If for each mode $1 \le j \le s$ the matrices

$$\widetilde{A}^{\mathrm{T}}(j)\Big(\sum_{i=1}^{s}P_i\Big) + \Big(\sum_{i=1}^{s}P_i\Big)\widetilde{A}(j) - \widetilde{C}^{\mathrm{T}}Y^{\mathrm{T}} - Y\widetilde{C} + \epsilon_A(j)\widetilde{D}_A^{\mathrm{T}}\widetilde{D}_A$$
$$+ \epsilon_A^{-1}(j)\Big(\sum_{i=1}^{s}P_i\Big)\widetilde{E}_A^{\mathrm{T}}\widetilde{E}_A\Big(\sum_{i=1}^{s}P_i\Big) + \epsilon_B^{-1}\Big(\sum_{i=1}^{s}P_i\Big)\Big(\sum_{i=1}^{s}P_i\Big) \tag{19}$$
$$+ \mu_j^{-1}\Big(\sum_{i=1}^{s}P_i\Big)\Big(\sum_{i=1}^{s}P_i\Big) < 0,$$

then $dV(e(t))/dt$ becomes always negative and the observer error decays exponentially to the origin. The last inequality is equivalent by the Schur complement to (12). This ends the proof.

An observer for uncertain single-mode systems can be deduced from result of Theorem 2. It is sufficient to replace $\Big(\sum\limits_{i=1}^{s}P_i\Big)$ by a one positive definite matrix X in the LMIs of Theorem 2 to deliver a sufficient conditions for the existence of the observer gain. We summarize the result in the following corollary.

Corollary 1 Consider the uncertain system

$$\begin{aligned}\frac{dx(t)}{dt} &= (A+\Delta A)x(t) + (B+\Delta B)u(t),\\ \frac{du(t)}{dt} &= v(t),\\ y(t) &= Cx(t),\end{aligned} \tag{20}$$

where $x(t) \in R^n$, $u(t) \in R^m$, and $y \in R^p$. The uncertain parts of $\Delta A = E_A^{\mathrm{T}}F_A(x(t))D_A$ and $\Delta B(\sigma(t)) = E_B^{\mathrm{T}}F_B(x(t))D_B$ are supposed to satisfy the inequalities $F_A^{\mathrm{T}}(\sigma(t)) \times F_A(\sigma(t)) < I$, $F_B^{\mathrm{T}}(\sigma(t))F_B(\sigma(t)) < I$, respectively. If there exist a matrix $X > 0$, a matrix Y of appropriate dimensions, and positive constants ϵ_A, and ϵ_B such that the following LMI is feasible

$$\begin{bmatrix}\mathcal{H}(X,Y) & \widetilde{E}_A X & X\\ X\widetilde{E}_A^{\mathrm{T}} & -\epsilon_A I & 0\\ X & 0 & -\epsilon_B I\end{bmatrix} < 0, \tag{21}$$

where

$$\mathcal{H}(X, Y) = \widetilde{A}^{\mathrm{T}} X + X\widetilde{A} - \widetilde{C}^{\mathrm{T}} Y^{\mathrm{T}} - Y\widetilde{C} + \epsilon_A \widetilde{D}_A^{\mathrm{T}} \widetilde{D}_A,$$

then system

$$\frac{\hat{\xi}(t}{dt}) = A\hat{\xi}(t) + \widetilde{B}v(t) + X^{-1}Y\left(y(t) - \widetilde{C}\hat{\xi}(t)\right) - \varphi(y(t), \hat{\xi}(t)), \tag{22}$$

is an asymptotic observer for the uncertain switching system (22) where $\varphi(y(t), \hat{\xi}(t))$ is defined as

$$\varphi(y(t), \hat{z}(t)) = \begin{cases} \epsilon_B \|\widetilde{E}_A\|^2 \|\widetilde{D}_A\|^2 \hat{\xi}^{\mathrm{T}}(t)\hat{\xi}(t) \dfrac{X^{-1}\widetilde{C}^{\mathrm{T}}\widetilde{C}e(t)}{\|\widetilde{C}e(t)\|^2} & \text{if} \quad \|\widetilde{C}e(t)\| \neq 0, \\ 0 & \text{if} \quad \|\widetilde{C}e(t)\| = 0, \end{cases}$$

where $\widetilde{E}_A$ and $\widetilde{D}_A$ are defined as in Section 3.

4 Illustrative Example

4.1 Observation of a switching system without uncertainties

Consider the following switching system described by:

$$A(1) = \begin{bmatrix} 0.1 & -0.5 \\ 0 & -1 \end{bmatrix}, \quad A(2) = \begin{bmatrix} -1 & -1 \\ 0.9 & -1 \end{bmatrix}, \quad B(1) = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$
$$B(2) = \begin{bmatrix} 0.1 \\ -1 \end{bmatrix}, \quad C = [\,1 \quad 0\,].$$

Applying the result of Theorem 1 with $\epsilon_A(j) = \epsilon_B(j) = 1 \; \forall\, j, \; \varpi = 1.7818,$ we obtain $\mu = 10^3$ and

$$P_1 = \begin{bmatrix} 27.8287 & -0.1325 & 5.6453 \\ -0.1325 & 53.9905 & 2.6926 \\ 5.6453 & 2.6926 & 39.0959 \end{bmatrix}, \quad P_2 = \begin{bmatrix} 571.9220 & -99.1576 & -201.2708 \\ -99.1576 & 140.3128 & 48.5049 \\ -201.2708 & 48.5049 & 176.2018 \end{bmatrix},$$
$$Y = \begin{bmatrix} 389.7613 \\ -242.2155 \\ 260.9596 \end{bmatrix},$$

and the observer gain is

$$(P_1 + P_2)^{-1} Y = \begin{bmatrix} 1.3246 \\ -1.2869 \\ 2.7216 \end{bmatrix}.$$

The nonlinear term in the observer dynamics can be computed in terms of the solutions P_1, P_2, μ, and ϖ.

4.2 Observation of a switching system with uncertainties

Taking the same switching model with the following additional data:

$$E_A = \begin{bmatrix} 0.2 & 0.5 \\ 0.4 & 0.4 \end{bmatrix}, \qquad D_A = \begin{bmatrix} 0.1 & 0.2 \\ 0.3 & 0 \end{bmatrix}, \qquad E_B = [\,0.3 \quad 0.6\,], \qquad D_B = 0.2.$$

By the application of result of Theorem 2, with $\mu = 10$, and $\epsilon_A(j) = \epsilon_B(j) = 1\ \forall\, j$, we have

$$P_1 = \begin{bmatrix} 0.1337 & -0.0303 & -0.0486 \\ -0.0303 & 0.0486 & 0.0143 \\ -0.0486 & 0.0143 & 0.0605 \end{bmatrix}, \qquad P_2 = \begin{bmatrix} 0.2767 & -0.0625 & -0.1029 \\ -0.0625 & 0.1010 & 0.0303 \\ -0.1029 & 0.0303 & 0.1248 \end{bmatrix},$$

$$Y = \begin{bmatrix} 0.9465 \\ -0.1669 \\ 0.1946 \end{bmatrix}, \qquad (P_1 + P_2)^{-1} Y = \begin{bmatrix} 3.8608 \\ 0.0266 \\ 4.1977 \end{bmatrix}.$$

5 Conclusion

A new observer design methodology is proposed to estimate the unmeasured states of switching systems and uncertain switching systems. We showed that a constant-gain observer is sufficient to observe the system states whatever the switch in the nominal matrices, and the existence of the observer gain is related to the feasibility of a set of coupled LMIs. The proposed observer design is an alternative to the technique of *switching observers* that necessitates both the construction of several observers and estimation of the current modes of the switching system being observed.

References

[1] Li, Z.G., Wen, C.Y. and Soh, Y. Observer-based stabilization of switching linear systems. *Automatica* **39** (2003) 517–524.

[2] Liu, Y. Switching observer design for uncertain nonlinear systems. *IEEE Trans. Automat. Control* **42**(12) (1997) 1699–1703.

[3] Kamas, L.A. and Sanders, S.R. Parameter and state estimation in power electronic circuits. *IEEE Trans. Circuits and Systems-I: Fundamental Theory and Appl.* **40**(12) (1993) 920–928.

[4] García, R.A. and Mancilla-Aguilar, J.L. State-norm estimation of switched nonlinear systems. *In Proc. of the Amer. Control Conf.* May 2002, P.1892–1896.

[5] Husain, I. and Islam, M.S. Observers for position and speed estimations in switched reluctance motors. *In Proc. of the 40th IEEE Conf. on Decision and Contr.* December 2001, P.2217–2222.

[6] Alessandri, A. and Coletta, P. Switching observers for continuous-time and discrete-time linear systems. *In Proc. of the Amer. Control Conf.*, June 2001, Arlington, P.2516–2521.

[7] Boukas, E.-K. and Liu, Z.K. *Deterministic and Stochastic Time-Delay Systems.* Birkhäuser, 2002.

Nonlinear Dynamics and Systems Theory, **5**(1) (2005) 61–90

A Nonlinear Model for Dynamics of Delaminated Composite Beam with Account of Contact of the Delamination Crack Faces, Based on the First Order Shear Deformation Theory

V.Y. Perel

Department of Aeronautics and Astronautics,
Air Force Institute of Technology,
6130 Noranda Dr., Dayton, Ohio 45415, USA

Received: October 14, 2004; Revised: January 9, 2005

Abstract: In this work, a new approach is developed for dynamic analysis of a composite beam with an interplay crack, in which a physically impossible interpenetration of the crack faces is prevented by imposing a special constraint, leading to taking account of a force of contact interaction of the crack faces and to nonlinearity of the formulated boundary value problem. Longitudinal force resultants in the delaminated parts of the beam are taken into account also, which is another source of the nonlinearity. The shear deformation and rotary inertia terms are included into the formulation, to achieve better accuracy. The model is based on the first order shear deformation theory, i.e. the longitudinal displacement is assumed to vary linearly through the beam's thickness. A variational formulation of the problem, nonlinear partial differential equations of motion with boundary conditions, a weak form for the partial differential equations and a finite element formulation on the basis of the weak form are developed. An example problem of a clamped-free beam with a piezoelectric actuator is considered, and its finite element solution is obtained. A noticeable difference of forced vibrations of the delaminated and undelaminated beams due to the contact interaction of the crack faces is predicted by the developed model. Besides, linear eigenvalue analysis shows decrease of natural frequencies upon increase of the crack length, and crack opening and closing during the vibration in higher mode shapes, beginning from the fifth one.

Keywords: *Composite delaminated beam; piezoelectric actuator; contact of crack faces; Lagrange multipliers; penalty function method; shear deformation theory; nonlinear partial differential equations; nonlinear finite element analysis.*

Mathematics Subject Classification (2000): 74A40, 74F15, 74S99.

1 Introduction

A model of a composite beam with the delamination and with a piezoelectric actuator, with account of contact interaction of the delamination crack faces, based on the classical beam theory, was presented in Reference [1]. This model did not take the shear strain energy into account, and, therefore, produced sufficiently accurate results only for thin beams. To model thicker beams with delamination, one needs to use a beam theory, based on simplifying assumptions, which do not lead to vanishing of the shear strains. The first order shear deformation theory [2], based on assumed linear variation of a longitudinal displacement in the thickness direction, is the simplest approach that satisfies the requirement of a non-zero shear strain. This approach is used in the present paper for modeling a composite delaminated beam with a piezoelectric actuator. In this model, the interpenetration of the crack faces is prevented by imposing a constraint, written with the use of the Heaviside function in one of its analytical forms, leading to taking account of a force of contact interaction of the crack faces and to nonlinearity of the formulated boundary value problem.

2 Variational Formulation of The Problem

Total Potential Energy for Zone 0 (Part 0), i.e. for $0 \leq x \leq a$ (Figure 2.1).

Assumptions of the first-order shear deformation beam theory:

$$u_0(x,z,t) = z\phi_0(x,t), \quad w_0(x,z,t) = w_0(x,t), \tag{1}$$

where $u_0(x,z,t)$ and $w_0(x,z,t)$ are longitudinal and transverse displacements of Zone 0 (Part 0). The subscript 0 in the notations $u_0(x,z,t)$ and $w_0(x,z,t)$ indicates that the quantities u_0 and w_0 are associated with the Zone 0 (Part 0). The notation $u_0 = u_0(x,z,t)$ is not a notation for the axial longitudinal displacement (at $z = 0$). The axial longitudinal displacement is considered to be negligibly small here, because this model is developed for the beam to which an external longitudinal force is not applied.

Strain-displacement relations:

$$\varepsilon_{xx}^{(0)} = \frac{\partial u_0}{\partial x}, \quad \varepsilon_{xz}^{(0)} = \frac{1}{2}\left(\frac{\partial u_0}{\partial z} + \frac{\partial w_0}{\partial x}\right). \tag{2a}$$

In this text, ε_{xz} is a notation for a component of the strain tensor, not an engineering strain. With account of Equation (1), Equations (2a) take the form

$$\varepsilon_{xx}^{(0)} = z\phi_0', \quad \varepsilon_{xz}^{(0)} = \frac{1}{2}(\phi_0 + w_0'), \tag{2b}$$

where prime denotes differentiation with respect to x.

Stress-strain relations for an orthotropic piezoelectric layer of a composite beam (plane stress with respect to the y-direction), Appendix A:

$$\sigma_{xx}^{(p)} = \frac{1}{\overline{S}_{11}}\varepsilon_{xx}^{(p)} - \frac{\overline{d}_{31}}{\overline{S}_{11}}\frac{V}{\tau}, \quad \sigma_{xz}^{(p)} = \frac{1}{\overline{S}_{55}}2\varepsilon_{xz}^{(p)}, \tag{3}$$

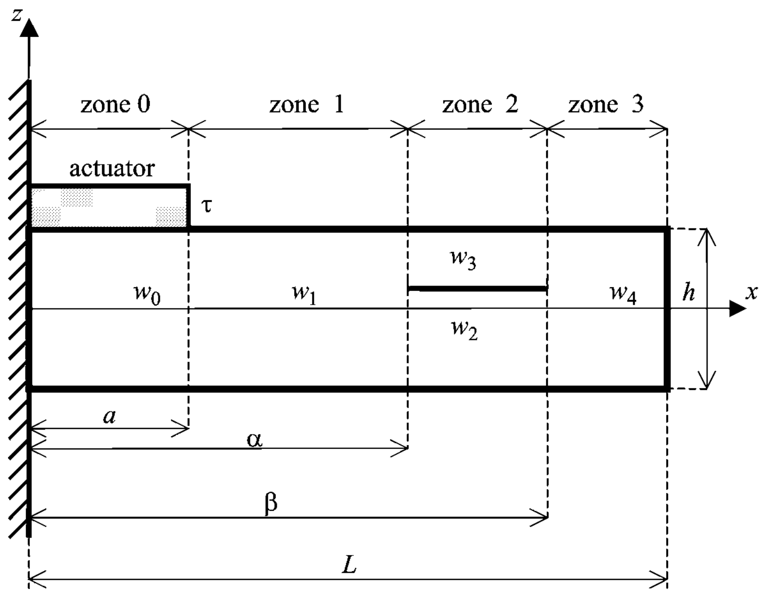

Figure 2.1. Cantilever beam with delamination and piezoelectric actuator.

a is length of the actuator; α is x-coordinate of the left crack tip; β is x-coordinate of the right crack tip; γ is z-coordinate of the crack (distance from x-axis to crack); τ is thickness of the actuator; w_0 is transverse displacement of zone 0; w_1 is transverse displacement of zone 1; w_2 is transverse displacement of lower part of zone 2 (under the crack); w_3 is transverse displacement of upper part of zone 2 (above the crack); w_4 is transverse displacement of zone 3.

where τ is thickness of the actuator, and $V = V(x,t)$ is a voltage, applied to the actuator. It is implied that this voltage creates an electric field in the z-direction.

Stress-strain relations for an orthotropic layer that does not have piezoelectric properties, Appendix A:

$$\sigma_{xx}^{(0)} = \frac{1}{\overline{S}_{11}}\varepsilon_{xx}^{(0)}, \quad \sigma_{xz}^{(0)} = \frac{1}{\overline{S}_{55}}2\varepsilon_{xz}^{(0)}. \tag{4}$$

Total potential energy where K is a shear correction factor and b is the beam's width

$$\begin{aligned} U_0 &= \frac{1}{2}\,b\int_0^a\int_{-h/2}^{h/2}\frac{1}{\overline{S}_{11}^{(0)}(z)}\big(\varepsilon_{xx}^{(0)}\big)^2\,dz\,dx \\ &+ \frac{1}{2}\,b\int_0^a\int_{h/2}^{h/2+\tau}\bigg(\frac{1}{\overline{S}_{11}^{(p)}(z)}\big(\varepsilon_{xx}^{(p)}\big)^2 - \frac{2\overline{d}_{31}(z)}{\overline{S}_{11}(z)}\,\frac{V}{\tau}\,\varepsilon_{xx}^{(p)}\bigg)dz\,dx \\ &+ 2Kb\int_0^a\int_{-h/2}^{h/2}\frac{1}{\overline{S}_{55}^{(0)}(z)}\big(\varepsilon_{xz}^{(0)}\big)^2\,dz\,dx + 2Kb\int_0^a\int_{-h/2}^{h/2+\tau}\frac{1}{\overline{S}_{55}^{(p)}(z)}\big(\varepsilon_{xz}^{(p)}\big)^2\,dz\,dx, \end{aligned} \tag{5}$$

where K is a shear correction factor and b is the beam's width. Substitution of equation (2b) into Equation (5) yields

$$U_0 = \int_0^a \left(\frac{A_0}{2}(\phi_0')^2 + K\frac{G_0}{2}(\phi_0 + w_0')^2 - I_p V \phi_0' \right) dx, \tag{6}$$

where the constants A_0, I_p and G_0 are defined as

$$A_0 = b \int_{-h/2}^{h/2} \frac{z^2}{\overline{S}_{11}^{(0)}(z)}\, dz + b \int_{h/2}^{h/2+\tau} \frac{z^2}{\overline{S}_{11}^{(p)}(z)}\, dz,$$
$$I_p = \left(\frac{b}{\tau} \int_{h/2}^{h/2+\tau} \frac{\overline{d}_{31}(z)}{\overline{S}_{11}^{(p)}(z)}\, z\, dz \right), \qquad G_0 = b \int_{-h/2}^{h/2} \frac{1}{\overline{S}_{55}^{(0)}(z)}\, dz + b \int_{h/2}^{h/2+\tau} \frac{1}{\overline{S}_{55}^{(p)}(z)}\, dz. \tag{7}$$

Kinetic Energy for Zone 0 (Part 0), i.e. for $0 \le x \le a$:

$$T_0 = \frac{1}{2}\, b \int_0^a \int_{-h/2}^{h/2} \rho^{(0)}(z)(\dot{w}_0^2 + \dot{u}_0^2)\, dz\, dx + \frac{1}{2}\, b \int_0^a \int_{h/2}^{h/2+\tau} \rho^{(p)}(z)(\dot{w}_0^2 + \dot{u}_0^2)\, dz\, dx, \tag{8}$$

where $\rho^{(0)}(z)$ is a mass density of composite layers of Zone 0 without piezoelectric properties and $\rho^{(p)}(z)$ is the mass density of the piezoelectric actuator ($\rho^{(p)}$ may depend on the z-coordinate if the actuator has plies with different densities).

Substitution of Equations (1) into Equation (8) produces the result

$$T_0 = \int_0^a \left(\frac{1}{2} B_0 \dot{w}_0^2 + \frac{1}{2} C_0 \dot{\phi}_0^2 \right) dx, \tag{9}$$

where the constants B_0 and C_0 are defined as follows:

$$B_0 = b \left(\int_{-h/2}^{h/2} \rho^{(0)}(z)\, dz + \int_{h/2}^{h/2+\tau} \rho^{(p)}(z)\, dz \right),$$
$$C_0 = b \left(\int_{-h/2}^{h/2} \rho^{(0)}(z)\, z^2\, dz + \int_{h/2}^{h/2+\tau} \rho^{(p)}(z)\, z^2\, dz \right). \tag{10}$$

In a similar manner we obtain the strain and kinetic energy for Zone 1 (Part 1) and Zone 3 (Part 4).

Strain Energy for Zone 1 (Part 1), i.e. for $a \le x \le \alpha$:

$$U_1 = \int_a^\alpha \left(\frac{A_1}{2}(\phi_1')^2 + K\frac{G_1}{2}(\phi_1 + w_1')^2 \right) dx, \tag{11}$$

where the constants A_1 and G_1 are defined as follows:

$$A_1 \equiv b \int_{-h/2}^{h/2} \frac{z^2}{\overline{S}_{11}^{(1)}(z)}\, dz \qquad G_1 \equiv b \int_{-h/2}^{h/2} \frac{1}{\overline{S}_{55}^{(1)}(z)}\, dz. \tag{12}$$

Kinetic energy for Zone 1 (Part 1), i.e. for $a \le x \le \alpha$:

$$T_1 = \int_a^{\alpha} \left(\frac{1}{2} B_1 \dot{w}_1^2 + \frac{1}{2} C_1 \dot{\phi}_1^2 \right) dx, \tag{13}$$

where the constants B_1 and C_1 are defined as follows:

$$B_1 = b \int_{-h/2}^{h/2} \rho^{(1)}(z)\, dz, \qquad C_1 = b \int_{-h/2}^{h/2} \rho^{(1)}(z)\, z^2\, dz. \tag{14}$$

Strain Energy for Zone 3 (Part 4), i.e. for $\beta \le x \le L$:

$$U_4 = \int_{\beta}^{L} \left(\frac{A_4}{2} (\phi_4')^2 + K \frac{G_4}{2} (\phi_4 + w_4')^2 \right) dx, \tag{15}$$

where the constants A_4 and G_4 are defined as follows:

$$A_4 = b \int_{-h/2}^{h/2} \frac{z^2}{\overline{S}_{11}^{(4)}(z)}\, dz \qquad G_4 = b \int_{-h/2}^{h/2} \frac{1}{\overline{S}_{55}^{(4)}(z)}\, dz. \tag{16}$$

Kinetic Energy for Zone 3 (Part 4), i.e. for $\beta \le x \le L$:

$$T_4 = \int_{\beta}^{L} \left(\frac{1}{2} B_4 \dot{w}_4^2 + \frac{1}{2}\, C_4 \dot{\phi}_4^2 \right) dx, \tag{17}$$

where the constants B_4 and C_4 are defined as follows:

$$B_4 = b \int_{-h/2}^{h/2} \rho^{(4)}(z)\; dz, \quad C_4 = b \int_{-h/2}^{h/2} \rho^{(4)}(z)\; z^2\; dz. \tag{18}$$

Strain Energy for Zone 2 (Part 2 and Part 3), i.e. for $\alpha \leq x \leq \beta$.

In the Zone 2, which contains the delamination crack, the longitudinal force resultants in the delaminated lower and upper parts (Part 2 and Part 3),

$$N_x^{(2)} = b\int\limits_{-h/2}^{\gamma} \sigma_{xx}^{(2)}\, dz, \qquad N_x^{(3)} = b\int\limits_{\gamma}^{h/2} \sigma_{xx}^{(3)}\, dz,$$

can be not negligibly small, even if external longitudinal forces are not applied to the beam. In order for these force resultants to be taken into account, a nonlinear term $\frac{1}{2}(w')^2$ in the Green-Lagrange strain-displacement relation for the strain component ε_{xx} must be taken into account. So, for the Part 2 (lower part of Zone 2) the following relations are used:

strain-displacement relations:

$$\varepsilon_{xx}^{(2)} = \frac{\partial u_2}{\partial x} + \frac{1}{2}\left(\frac{\partial w_2}{\partial x}\right)^2, \tag{19a}$$

$$\varepsilon_{xz}^{(2)} = \frac{1}{2}\left(\frac{\partial u_2}{\partial z} + \frac{\partial w_2}{\partial x}\right); \tag{19b}$$

simplifying assumptions:

$$u_2(x,z,t) = z\phi_2(x,t) \quad w_2(x,z,t) = w_2(x,t); \tag{20}$$

stress-strain relations:

$$\sigma_{xx}^{(2)} = \frac{1}{\overline{S}_{11}^{(2)}}\varepsilon_{xx}^{(2)}, \quad \sigma_{xz}^{(2)} = \frac{1}{\overline{S}_{55}^{(2)}}2\varepsilon_{xz}^{(2)}; \tag{21}$$

strain energy:

$$U_2 = \frac{1}{2}\,b\int\limits_{\alpha}^{\beta}\int\limits_{-h/2}^{\gamma} \sigma_{xx}^{(2)}\varepsilon_{xx}^{(2)}\, dz\, dx + Kb\int\limits_{\alpha}^{\beta}\int\limits_{-h/2}^{\gamma} \sigma_{xz}^{(2)}\varepsilon_{xz}^{(2)}\, dz\, dx. \tag{22}$$

From Equations (18) – (22) we obtain the following expression for the strain energy:

$$U_2 = \int\limits_{\alpha}^{\beta}\left[\frac{1}{2}A_2(\phi_2')^2 + \frac{1}{2}KG_2(\phi_2 + w_2')^2 + \frac{1}{4}H_2(w_2')^2\phi_2' + \frac{1}{4}N_x^{(2)}(w_2')^2\right]dx, \tag{23}$$

where A_2, G_2, H_2 are constants, defined as

$$A_2 = b\int\limits_{-h/2}^{\gamma}\frac{1}{\overline{S}_{11}^{(2)}(z)}z^2\, dz, \quad G_2 = b\int\limits_{-h/2}^{\gamma}\frac{1}{\overline{S}_{55}^{(2)}(z)}\, dz, \quad H_2 = b\int\limits_{-h/2}^{\gamma}\frac{1}{\overline{S}_{11}^{(2)}(z)}\, z\, dz, \tag{24}$$

and $N_x^{(2)}$ is a longitudinal force resultant in the lower delaminated part (Part 2):

$$N_x^{(2)} = b \int_{-h/2}^{\gamma} \sigma_{xx}^{(2)}\, dz = H_2 \phi_2' + \frac{1}{2}\, Q_2 (w_2')^2, \tag{25}$$

where Q_2 is a constant, defined as

$$Q_2 = b \int_{-h/2}^{\gamma} \frac{1}{\overline{S}_{11}^{(2)}(z)}\, dz. \tag{26}$$

Similarly, for the Part 3 (upper part of Zone 2) the expression for the strain energy has the form

$$U_3 = \int_{\alpha}^{\beta} \left[\frac{1}{2} A_3 (\phi_3')^2 + \frac{1}{2} K G_3 (\phi_3 + w_3')^2 + \frac{1}{4} H_3 (w_3')^2 \phi_3' + \frac{1}{4} N_x^{(3)} (w_3')^2 \right] dx, \tag{27}$$

where

$$A_3 = b \int_{\gamma}^{h/2} \frac{1}{\overline{S}_{11}^{(3)}(z)}\, z^2\, dz, \qquad G_3 = b \int_{\gamma}^{h/2} \frac{1}{\overline{S}_{55}^{(3)}(z)}\, dz, \qquad H_3 = b \int_{\gamma}^{h/2} \frac{1}{\overline{S}_{11}^{(3)}(z)}\, z\, dz, \tag{28}$$

$$N_x^{(3)} = b \int_{\gamma}^{h/2} \sigma_{xx}^{(3)}\, dz = H_3 \phi_3' + \frac{1}{2}\, Q_3 (w_3')^2, \tag{29}$$

where

$$Q_3 = b \int_{\gamma}^{h/2} \frac{1}{\overline{S}_{11}^{(3)}(z)}\, dz. \tag{30}$$

Kinetic Energy for Zone 2 (Part 2 and Part 3), i.e. for $\alpha \le x \le \beta$.

Expressions for kinetic energy of Part 2 and Part 3 are obtained similarly to the expressions for the kinetic energies of all other parts, and they have the form:

$$\begin{aligned} T_2 &= \int_{\alpha}^{\beta} \left(\frac{1}{2} B_2 \dot{w}_2^2 + \frac{1}{2} C_2 \dot{\phi}_2^2 \right) dx, \\ T_3 &= \int_{\alpha}^{\beta} \left(\frac{1}{2} B_3 \dot{w}_3^2\, dx + \frac{1}{2} C_3 \dot{\phi}_3^2 \right) dx, \end{aligned} \tag{31}$$

where

$$\begin{aligned} B_2 &= b \int_{-h/2}^{\gamma} \rho^{(2)}(z)\, dz, \qquad C_2 = b \int_{-h/2}^{\gamma} \rho^{(2)}(z) z^2\, dz, \\ B_3 &= b \int_{\gamma}^{h/2} \rho^{(3)}(z)\, dz, \qquad C_3 = b \int_{\gamma}^{h/2} \rho^{(3)}(z) z^2\, dz. \end{aligned} \tag{32}$$

In view of the expressions for strain and kinetic energies, derived above, the Lagrangian function density (potential energy minus kinetic energy per unit length) for the delaminated composite beam with the piezoelectric actuator (Figure 2.1) can be written as follows

$$\widetilde{\mathfrak{L}} = \begin{cases} \widetilde{\mathfrak{L}}_0(\dot{w}_0,\, w_0',\, \phi_0,\, \dot{\phi}_0,\, \phi_0') & \text{in Zone 0} \;\; (0 \le x \le a) \\ \widetilde{\mathfrak{L}}_1(\dot{w}_1,\, w_1',\, \phi_1,\, \dot{\phi}_1,\, \phi_1') & \text{in Zone 1} \;\; (a \le x \le \alpha) \\ \widetilde{\mathfrak{L}}_2(\dot{w}_2,\, w_2',\, \phi_2,\, \dot{\phi}_2,\, \phi_2',\, \dot{w}_3,\, w_3',\, \phi_3,\, \dot{\phi}_3,\, \phi_3') & \text{in Zone 2} \;\; (\alpha \le x \le \beta) \\ \widetilde{\mathfrak{L}}_3(\dot{w}_4,\, w_4',\, \phi_4,\, \dot{\phi}_4,\, \phi_4') & \text{in Zone 3} \;\; (\beta \le x \le L), \end{cases} \tag{33}$$

where

$$\widetilde{\mathfrak{L}}_0 = \frac{A_0}{2}(\phi_0')^2 + K\frac{G_0}{2}(\phi_0 + w_0')^2 - I_p V \phi_0' - \frac{B_0}{2}\dot{w}_0^2 - \frac{C_0}{2}\dot{\phi}_0^2, \tag{34a}$$

$$\widetilde{\mathfrak{L}}_1 = \frac{A_1}{2}(\phi_1')^2 + K\frac{G_1}{2}(\phi_1 + w_1')^2 - \frac{B_1}{2}\dot{w}_1^2 - \frac{C_1}{2}\dot{\phi}_1^2, \tag{34b}$$

$$\begin{aligned} \widetilde{\mathfrak{L}}_2 = {} & \frac{1}{2}A_2(\phi_2')^2 + \frac{1}{2}KG_2(\phi_2 + w_2')^2 + \frac{1}{2}H_2(w_2')^2\phi_2' + \frac{1}{8}Q_2(w_2')^4 \\ & - \frac{1}{2}B_2\dot{w}_2^2 - \frac{1}{2}C_2\dot{\phi}_2^2 + \frac{1}{2}A_3(\phi_3')^2 + \frac{1}{2}KG_3(\phi_3 + w_3')^2 \\ & + \frac{1}{2}H_3(w_3')^2\phi_3' + \frac{1}{8}Q_3(w_3')^4 - \frac{1}{2}B_3\dot{w}_2^2 - \frac{1}{2}C_3\dot{\phi}_3^2, \end{aligned} \tag{34c}$$

$$\widetilde{\mathfrak{L}}_3 = \frac{A_4}{2}(\phi_4')^2 + K\frac{G_4}{2}(\phi_4 + w_4')^2 - \frac{B_4}{2}\dot{w}_4^2 - \frac{C_4}{2}\dot{\phi}_4^2. \tag{34d}$$

A variational formulation of the problem includes essential boundary conditions at the ends of each zone, which will be treated as point-wise constraints, and a nonpenetration condition for the delamination crack faces (subdomain constraints for Zone 2), Reference [1]. For a clamped-free beam, the *point-wise constraints* have the form

$$R_i(t) = 0 \quad (i = 1,\, 2, \ldots,\, 12), \tag{35a}$$

where

$$\begin{aligned} R_1 &\equiv w_0(0,t), & R_2 &\equiv \phi_0(0,t), \\ R_3 &\equiv w_0(a,t) - w_1(a,t) & R_4 &\equiv \phi_0(a,t) - \phi_1(a,t), \\ R_5 &\equiv w_1(\alpha,t) - w_2(\alpha,t), & R_6 &\equiv \phi_1(\alpha,t) - \phi_2(\alpha,t), \\ R_7 &\equiv w_1(\alpha,t) - w_3(\alpha,t), & R_8 &\equiv \phi_1(\alpha,t) - \phi_3(\alpha,t), \\ R_9 &\equiv w_2(\beta,t) - w_4(\beta,t), & R_{10} &\equiv \phi_2(\beta,t) - \phi_4(\beta,t), \\ R_{11} &\equiv w_3(\beta,t) - w_4(\beta,t), & R_{12} &\equiv \phi_3(\beta,t) - \phi_4(\beta,t). \end{aligned} \tag{35b}$$

In case of other kinds of fixation of the beam's ends, the first two point-wise constraints will be different, of course, but the other point-wise constraints will be the same.

During the vibration of the delaminated beam, the upper and lower delaminated parts touch each other, and the force of their interaction needs to be taken into account. This force enters into the differential equations of motion as a reaction of constraint, which prevents overlapping of the upper and lower delaminated parts. Such a constraint can be expressed by a relationship between w_2 and w_3 (i.e. displacements of the lower and upper delaminated parts) that prevents the difference $w_3 - w_2$ to take on negative values:

$$f(w_2(x,t),\, w_3(x,t)) = f(x,t) \equiv (w_3 - w_2)\,[1 - \mathfrak{H}_0(w_3 - w_2)] = 0, \tag{36a}$$

where $\mathfrak{H}_0$ is a Heaviside function, defined in Appendix B. If delaminated sublaminates "attempt" to overlap during the vibration (if $w_3 - w_2 < 0$), or if the crack is closed $(w_3 - w_2 = 0)$, then $\mathfrak{H}_0(w_3 - w_2) = 0$, and, therefore, due to equation (10), the difference $w_3 - w_2$ is set equal to zero. If the crack is open $(w_3 - w_2 > 0)$, then $\mathfrak{H}_0(w_3 - w_2) = 1$, and no constraints are imposed on the difference $w_3 - w_2$. With the use of the analytical representation of the Heaviside function (equation B-5), the nonpenetration constraint, expressed by equation (36a), can be written as follows:

$$f(x,t) \equiv (w_3 - w_2)\left(\frac{1}{2} - \frac{1}{\pi}\arctan\frac{w_3 - w_2}{\epsilon}\right) = 0, \tag{36b}$$

where ϵ is some small number. The nonpenetration constraint (36) is a *subdomain constraint* for the Zone 2 $(\alpha \le x \le \beta)$.

Now, the problem can be formulated as a problem of finding a constrained (conditional) extremum of the functional

$$J = \int_{t_1}^{t_2}\int_0^L \widetilde{\mathfrak{L}}\,dx\,dt \tag{37}$$

with constraints expressed by Equations (35) and (36). The constraints (35) and (36) can be included into the functional by the method of Lagrange multipliers. This will produce a modified functional $\overline{J}$:

$$\overline{J} = J + \int_{t_1}^{t_2}\sum_{i=1}^{12}\lambda_i(t)R_i(t) + \int_{t_1}^{t_2}\int_{\alpha}^{\beta}\mu(x,t)f(x,t)\,dx\,dt, \tag{39}$$

where $\lambda_i(t)$ and $\mu(x,t)$ are the Lagrange multipliers. Now we have a problem of an unconstrained (unconditional) extremum of the modified functional $\overline{J}$. Derivation of the partial differential equations of motion and natural boundary conditions from the condition of extremum of the functional (39) can be performed using standard methods of calculus of variations. In the following text, partial differential equations of motion with boundary conditions, a weak form of the partial differential equations and a finite element formulation on the basis of the weak form will be obtained.

3 Partial Differential Equations with Boundary Conditions

To derive the partial differential equations of motion with boundary conditions, the condition of unconstrained extremum of the functional $\overline{J}$ (Equation (39)) will be used. The condition $\delta\overline{J} = 0$ leads to the following partial differential equations and natural boundary conditions.

Partial differential equations:

$$-\frac{\partial}{\partial t}\frac{\partial \widetilde{\mathfrak{L}}_0}{\partial \dot{w}_0} - \frac{\partial}{\partial x}\frac{\partial \widetilde{\mathfrak{L}}_0}{\partial w_0'} = 0 \quad \text{in} \quad x \in [0,\, a]\,, \tag{41}$$

$$\frac{\partial \widetilde{\mathfrak{L}}_0}{\partial \phi_0} - \frac{\partial}{\partial t}\frac{\partial \widetilde{\mathfrak{L}}_0}{\partial \dot{\phi}_0} - \frac{\partial}{\partial x}\frac{\partial \widetilde{\mathfrak{L}}_0}{\partial \phi_0'} = 0 \quad \text{in} \quad x \in [0,\, a]\,, \tag{42}$$

$$-\frac{\partial}{\partial t}\frac{\partial \widetilde{\mathfrak{L}}_1}{\partial \dot{w}_1} - \frac{\partial}{\partial x}\frac{\partial \widetilde{\mathfrak{L}}_1}{\partial w_1'} = 0 \quad \text{in} \quad x \in [a,\, \alpha]\,, \tag{43}$$

$$\frac{\partial \widetilde{\mathfrak{L}}_1}{\partial \phi_1} - \frac{\partial}{\partial t}\frac{\partial \widetilde{\mathfrak{L}}_1}{\partial \dot{\phi}_1} - \frac{\partial}{\partial x}\frac{\partial \widetilde{\mathfrak{L}}_1}{\partial \phi_1'} = 0 \quad \text{in} \quad x \in [a,\, \alpha], \tag{44}$$

$$\mu\frac{\partial f}{\partial w_2} - \frac{\partial}{\partial t}\frac{\partial \widetilde{\mathfrak{L}}_2}{\partial \dot{w}_2} - \frac{\partial}{\partial x}\frac{\partial \widetilde{\mathfrak{L}}_2}{\partial w_2'} = 0 \quad \text{in} \quad x \in [\alpha,\, \beta]\,, \tag{45}$$

$$\frac{\partial \widetilde{\mathfrak{L}}_2}{\partial \phi_2} - \frac{\partial}{\partial t}\frac{\partial \widetilde{\mathfrak{L}}_2}{\partial \dot{\phi}_2} - \frac{\partial}{\partial x}\frac{\partial \widetilde{\mathfrak{L}}_2}{\partial \phi_2'} = 0 \quad \text{in} \quad x \in [\alpha,\, \beta]\,, \tag{46}$$

$$\mu\frac{\partial f}{\partial w_3} - \frac{\partial}{\partial t}\frac{\partial \widetilde{\mathfrak{L}}_2}{\partial \dot{w}_3} - \frac{\partial}{\partial x}\frac{\partial \widetilde{\mathfrak{L}}_2}{\partial w_3'} = 0 \quad \text{in} \quad x \in [\alpha,\, \beta]\,, \tag{47}$$

$$\frac{\partial \widetilde{\mathfrak{L}}_2}{\partial \phi_3} - \frac{\partial}{\partial t}\frac{\partial \widetilde{\mathfrak{L}}_2}{\partial \dot{\phi}_3} - \frac{\partial}{\partial x}\frac{\partial \widetilde{\mathfrak{L}}_2}{\partial \phi_3'} = 0 \quad \text{in} \quad x \in [\alpha,\, \beta]\,, \tag{48}$$

$$-\frac{\partial}{\partial t}\frac{\partial \widetilde{\mathfrak{L}}_3}{\partial \dot{w}_4} - \frac{\partial}{\partial x}\frac{\partial \widetilde{\mathfrak{L}}_3}{\partial w_4'} = 0 \quad \text{in} \quad x \in [\beta,\, L]\,, \tag{49}$$

$$\frac{\partial \widetilde{\mathfrak{L}}_3}{\partial \phi_4} - \frac{\partial}{\partial t}\frac{\partial \widetilde{\mathfrak{L}}_3}{\partial \dot{\phi}_4} - \frac{\partial}{\partial x}\frac{\partial \widetilde{\mathfrak{L}}_3}{\partial \phi_4'} = 0 \quad \text{in} \quad x \in [\beta,\, L]\,. \tag{50}$$

Natural boundary conditions:

$$\frac{\partial \widetilde{\mathfrak{L}}_0}{\partial w_0'} + \lambda_3 = 0 \quad \text{at} \quad x = a, \qquad \frac{\partial \widetilde{\mathfrak{L}}_0}{\partial \phi_0'} + \lambda_4 = 0 \quad \text{at} \quad x = a, \tag{51}$$

$$-\frac{\partial \widetilde{\mathfrak{L}}_1}{\partial w_1'} - \lambda_3 = 0 \quad \text{at} \quad x = a, \qquad \frac{\partial \widetilde{\mathfrak{L}}_1}{\partial w_1'} + \lambda_5 + \lambda_7 = 0 \quad \text{at} \quad x = \alpha, \tag{52}$$

$$-\frac{\partial \widetilde{\mathfrak{L}}_1}{\partial \phi_1'} - \lambda_4 = 0 \quad \text{at} \quad x = a, \qquad \frac{\partial \widetilde{\mathfrak{L}}_1}{\partial \phi_1'} + \lambda_6 + \lambda_8 = 0 \quad \text{at} \quad x = \alpha \tag{53}$$

$$-\frac{\partial \widetilde{\mathfrak{L}}_2}{\partial w_2'} - \lambda_5 = 0 \quad \text{at} \quad x = \alpha, \qquad \frac{\partial \widetilde{\mathfrak{L}}_2}{\partial w_2'} + \lambda_9 = 0 \quad \text{at} \quad x = \beta, \tag{54}$$

$$-\frac{\partial \widetilde{\mathfrak{L}}_2}{\partial \phi_2'} - \lambda_6 = 0 \quad \text{at} \quad x = \alpha, \qquad \frac{\partial \widetilde{\mathfrak{L}}_2}{\partial \phi_2'} + \lambda_{10} = 0 \quad \text{at} \quad x = \beta, \tag{55}$$

$$-\frac{\partial \widetilde{\mathfrak{L}}_2}{\partial w_3'} - \lambda_7 = 0 \quad \text{at} \quad x = \alpha, \qquad \frac{\partial \widetilde{\mathfrak{L}}_2}{\partial w_3'} + \lambda_{11} = 0 \quad \text{at} \quad x = \beta, \tag{56}$$

$$-\frac{\partial \widetilde{\mathfrak{L}}_2}{\partial \phi_3'} - \lambda_8 = 0 \quad \text{at} \quad x = \alpha, \qquad \frac{\partial \widetilde{\mathfrak{L}}_2}{\partial \phi_3'} + \lambda_{12} = 0 \quad \text{at} \quad x = \beta, \tag{57}$$

$$-\frac{\partial\widetilde{\mathfrak{L}}_3}{\partial w_4'} - \lambda_9 - \lambda_{11} = 0 \quad \text{at} \quad x = \beta, \qquad \frac{\partial\widetilde{\mathfrak{L}}_3}{\partial w_4'} = 0 \quad \text{at} \quad x = L, \tag{58}$$

$$-\frac{\partial\widetilde{\mathfrak{L}}_3}{\partial \phi_4'} - \lambda_{10} - \lambda_{12} = 0 \quad \text{at} \quad x = \beta, \qquad \frac{\partial\widetilde{\mathfrak{L}}_3}{\partial \phi_4'} = 0 \quad \text{at} \quad x = L. \tag{59}$$

Elimination of the Lagrange multipliers from Equations (51) – (59) leads to the following eight natural boundary conditions:

$$\frac{\partial\widetilde{\mathfrak{L}}_0}{\partial w_0'} - \frac{\partial\widetilde{\mathfrak{L}}_1}{\partial w_1'} = 0 \quad \text{at} \quad x = a, \tag{60}$$

$$\frac{\partial\widetilde{\mathfrak{L}}_0}{\partial \phi_0'} - \frac{\partial\widetilde{\mathfrak{L}}_1}{\partial \phi_1'} = 0 \quad \text{at} \quad x = a, \tag{61}$$

$$\frac{\partial\widetilde{\mathfrak{L}}_1}{\partial w_1'} - \frac{\partial\widetilde{\mathfrak{L}}_2}{\partial w_2'} - \frac{\partial\widetilde{\mathfrak{L}}_2}{\partial w_3'} = 0 \quad \text{at} \quad x = \alpha, \tag{62}$$

$$\frac{\partial\widetilde{\mathfrak{L}}_1}{\partial \phi_1'} - \frac{\partial\widetilde{\mathfrak{L}}_2}{\partial \phi_2'} - \frac{\partial\widetilde{\mathfrak{L}}_2}{\partial \phi_3'} = 0 \quad \text{at} \quad x = \alpha, \tag{63}$$

$$\frac{\partial\widetilde{\mathfrak{L}}_2}{\partial w_2'} + \frac{\partial\widetilde{\mathfrak{L}}_2}{\partial w_3'} - \frac{\partial\widetilde{\mathfrak{L}}_3}{\partial w_4'} = 0 \quad \text{at} \quad x = \beta, \tag{64}$$

$$\frac{\partial\widetilde{\mathfrak{L}}_2}{\partial \phi_2'} + \frac{\partial\widetilde{\mathfrak{L}}_2}{\partial \phi_3'} - \frac{\partial\widetilde{\mathfrak{L}}_3}{\partial \phi_4'} = 0 \quad \text{at} \quad x = \beta, \tag{65}$$

$$\frac{\partial\widetilde{\mathfrak{L}}_3}{\partial w_4'} = 0 \quad \text{at} \quad x = L, \tag{66}$$

$$\frac{\partial\widetilde{\mathfrak{L}}_3}{\partial \phi_4'} = 0 \quad \text{at} \quad x = L. \tag{67}$$

Substitution of Equations (34) into Equations (41) – (50) and into Equations (60) – (67) produces the following result.

Partial differential equations:

$$KG_0(w_0'' + \phi_0') - B_0\ddot{w}_0 = 0 \quad \text{in} \quad x \in [0,\, a]\,, \tag{68}$$

$$A_0\phi_0'' - KG_0(w_0' + \phi_0) - C_0\ddot{\phi}_0 = I_pV' \quad \text{in} \quad x \in [0,\, a]\,, \tag{69}$$

$$KG_1(w_1'' + \phi_1') - B_1\ddot{w}_1 = 0 \quad \text{in} \quad x \in [a,\, \alpha]\,, \tag{70}$$

$$A_1\phi_1'' - KG_1(w_1' + \phi_1) - C_1\ddot{\phi}_1 = 0 \quad \text{in} \quad x \in [a,\, \alpha]\,, \tag{71}$$

$$\begin{aligned} KG_2(w_2'' + \phi_2') - B_2\ddot{w}_2 - \mu\left(\frac{1}{2} - \frac{1}{\pi}\arctan\frac{w_3 - w_2}{\epsilon}\right) \\ - H_2\phi_2'w_2'' - \frac{3}{2}\,Q_2(w_2')^2w_2'' = 0 \quad \text{in} \quad x \in [\alpha,\, \beta]\,, \end{aligned} \tag{72}$$

$$A_2\phi_2'' - KG_2(w_2' + \phi_2) - C_2\ddot{\phi}_2 - H_2w_2'w_2'' = 0 \quad \text{in} \quad x \in [\alpha,\, \beta]\,, \tag{73}$$

$$\begin{aligned} KG_3(w_3'' + \phi_3') - B_3\ddot{w}_3 + \mu\left(\frac{1}{2} - \frac{1}{\pi}\arctan\frac{w_3 - w_2}{\epsilon}\right) \\ - H_3\phi_3'w_3'' - \frac{3}{2}\,Q_3(w_3')^2w_3'' = 0 \quad \text{in} \quad x \in [\alpha,\, \beta]\,, \end{aligned} \tag{74}$$

$$A_3\phi_3'' - KG_3(w_3' + \phi_3) - C_3\ddot{\phi}_3 - H_3 w_3' w_3'' = 0 \quad \text{in} \quad x \in [\alpha,\, \beta]\,, \tag{75}$$

$$KG_4(w_4'' + \phi_4') - B_4\ddot{w}_4 = 0 \quad \text{in} \quad x \in [\beta,\, L]\,, \tag{76}$$

$$A_4\phi_4'' - KG_4(w_4' + \phi_4) - C_4\ddot{\phi}_4 = 0 \quad \text{in} \quad x \in [\beta,\, L]\,, \tag{77}$$

Natural boundary conditions:

$$G_0(\phi_0 + w_0') - G_1(\phi_1 + w_1') = 0 \quad \text{at} \quad x = a, \tag{78}$$

$$A_0\phi_0' - A_1\phi_1' = I_p V \quad \text{at} \quad x = a, \tag{79}$$

$$\begin{aligned} &KG_1(\phi_1 + w_1') - KG_2(\phi_2 + w_2') - H_2 w_2'\phi_2' \\ &- \frac{1}{2}\,Q_2(w_2')^3 - KG_3(\phi_3 + w_3') - H_3 w_3'\phi_3' - \frac{1}{2}\,Q_3(w_3')^3 = 0 \quad \text{at} \quad x = \alpha, \end{aligned} \tag{80}$$

$$A_1\phi_1' - A_2\phi_2' - \frac{1}{2}\,H_2(w_2')^2 - A_3\phi_3' - \frac{1}{2}\,H_3(w_3')^2 = 0 \quad \text{at} \quad x = \alpha, \tag{81}$$

$$\begin{aligned} &KG_2(\phi_2 + w_2') + H_2 w_2'\phi_2' + \frac{1}{2}\,Q_2(w_2')^3 + KG_3(\phi_3 + w_3') \\ &+ H_3 w_3'\phi_3' + \frac{1}{2}\,Q_3(w_3')^3 - KG_4(\phi_4 + w_4') = 0 \quad \text{at} \quad x = \beta, \end{aligned} \tag{82}$$

$$A_2\phi_2' + A_3\phi_3' - A_4\phi_4' = 0 \quad \text{at} \quad x = \beta, \tag{83}$$

$$\phi_4 + w_4' = 0 \quad \text{at} \quad x = L, \tag{84}$$

$$\phi_4' = 0 \quad \text{at} \quad x = L. \tag{85}$$

In computation of derivatives $\dfrac{\partial f}{\partial w_2}$ and $\dfrac{\partial f}{\partial w_3}$ in Equations (45) and (47), which led to Equations (72) and (74), the following chain of transformations was used:

$$\begin{aligned} \frac{\partial f}{\partial w_2} &= \lim_{\epsilon\to 0}\frac{\partial}{\partial w_2}\left((w_3 - w_2)\left(\frac{1}{2} - \frac{1}{\pi}\arctan\frac{w_3 - w_2}{\epsilon}\right)\right) \\ &= \lim_{\epsilon\to 0}\left(-\left(\frac{1}{2} - \frac{1}{\pi}\arctan\frac{w_3 - w_2}{\epsilon}\right) + \frac{1}{\epsilon\pi}\,\frac{\epsilon^2(w_3 - w_2)}{\epsilon^2 + (w_3 - w_2)^2}\right) \\ &= -\lim_{\epsilon\to 0}\left(\frac{1}{2} - \frac{1}{\pi}\arctan\frac{w_3 - w_2}{\epsilon}\right). \end{aligned}$$

So,

$$\frac{\partial f}{\partial w_2} \approx -\left(\frac{1}{2} - \frac{1}{\pi}\arctan\frac{w_3 - w_2}{\epsilon}\right), \tag{86}$$

where ϵ is some small number. Similarly

$$\frac{\partial f}{\partial w_3} \approx \left(\frac{1}{2} - \frac{1}{\pi}\arctan\frac{w_3 - w_2}{\epsilon}\right). \tag{87}$$

So, the formulation of the problem includes eleven equations for subdomains: ten partial differential equations (68)–(77) and one algebraic equation of constraint (36b) for Zone 2. The number of unknown functions is also eleven. The unknown functions are: $\mu(x,t)$, $w_k(x,t)$, $\phi_k(x,t)$ $(k = 0, 1, 2, 3, 4)$. The total order of spatial derivatives of the partial differential equations is twenty, and the number of boundary conditions is also twenty: twelve essential boundary conditions (Equations (35)) and eight natural boundary conditions (Equations (78) – (85)).

The formulation of the problem in terms of partial differential equations can be simplified, if the *penalty function method* [2] is used for the nonpenetration constraint, i.e. if the Lagrange multiplier $\mu(x,t)$, associated with the nonpenetration constraint (36), is written as

$$\mu(x,t) = \chi f(x,t), \tag{88}$$

where the function $f(x,t)$ is defined by Equation (36b), and χ is some large number, which has to be chosen by an analyst. Then, Equation (72) takes the form

$$KG_2(w_2'' + \phi_2') - B_2\ddot{w}_2 - \chi(w_3 - w_2)\left(\frac{1}{2} - \frac{1}{\pi}\arctan\frac{w_3 - w_2}{\epsilon}\right) = 0 \quad \text{in} \quad x \in [\alpha,\, \beta]\,, \tag{89}$$

and Equation (74) takes the form

$$KG_3(w_3'' + \phi_3') - B_3\ddot{w}_3 + \chi(w_3 - w_2)\left(\frac{1}{2} - \frac{1}{\pi}\arctan\frac{w_3 - w_2}{\epsilon}\right) = 0 \quad \text{in} \quad x \in [\alpha,\, \beta]\,. \tag{90}$$

In transition from Equations (72) and (74) to Equations (89) and (90) respectively, the following transformation was used

$$\begin{aligned}\left(\frac{1}{2} - \frac{1}{\pi}\arctan\frac{w_3 - w_2}{\epsilon}\right)^2 &= (1 - \mathfrak{H}_0(w_3 - w_2))^2 \\ &= 1 - \mathfrak{H}_0(w_3 - w_2) = \frac{1}{2} - \frac{1}{\pi}\arctan\frac{w_3 - w_2}{\epsilon}.\end{aligned} \tag{91}$$

Now, the formulation of the problem contains ten partial differential equations (68) – (71), (89), (73), (90), (75) – (77) with ten unknown functions $w_k(x,t)$, $\phi_k(x,t)$, $(k = 0,1,2,3,4)$.

The natural boundary condition (79) is nonhomogeneous, because the externally applied voltage $V(a,t)$ enters into it. To avoid having the nonhomogeneous boundary condition, one can consider that the voltage, applied to the actuator, is distributed not over the subdomain $x \in [0,a]$, but over the subdomain $x \in [0,\, a - \varepsilon]$, where ε is some very small positive number. Then the physics of the problem is not changed, and the voltage $V(a,t)$ does not enter into the boundary condition (79), i.e. this boundary condition takes a simpler homogeneous form

$$A_0\phi_0' - A_1\phi_1' = 0 \quad \text{at} \quad x = a. \tag{92}$$

Let us consider, for example, the *voltage distributed uniformly over the length of the actuator*, i.e.

$$V(x,t) = V(t) \quad \text{in} \quad x \in [0,a]. \tag{93a}$$

Then, without altering the physical formulation of the problem, we can write

$$V(x,t) = V(t) \quad \text{in} \quad x \in [0,\, a - \varepsilon]. \tag{93b}$$

Then, the voltage $V(x,t)$ can be presented in the form

$$V(x,t) = [1 - \mathfrak{H}_{a-\varepsilon}(x)]\; V(t) \quad \text{in} \quad x \in [0,a] \tag{94}$$

where $\mathfrak{H}_{a-\varepsilon}(x)$ is the Heaviside function (Appendix B). Then, the right side of the differential equation (69) takes the form

$$I_p V' = -I_p V(t)\ \mathfrak{H}'_{a-\varepsilon}(x) = -I_p V(t)\ \delta_{a-\varepsilon}(x), \tag{95}$$

where $\delta_{a-\varepsilon}(x)$ is the delta-function (Appendix B).

In computation of the example problems for the clamped-free beams, presented below, the formulation based on the penalty function method will be used, and the voltage will be distributed uniformly along the length of the actuator. For convenience, this formulation is summarized below.

Partial differential equations:

$$KG_0(w_0'' + \phi_0') - B_0\ddot{w}_0 = 0 \quad \text{in} \quad x \in [0, a], \tag{96}$$

$$A_0\phi_0'' - KG_0(w_0' + \phi_0) - C_0\ddot{\phi}_0 = -I_p V(t)\delta_{a-\varepsilon}(x) \quad \text{in} \quad x \in [0, a], \tag{97}$$

$$KG_1(w_1'' + \phi_1') - B_1\ddot{w}_1 = 0 \quad \text{in} \quad x \in [a, \alpha], \tag{98}$$

$$A_1\phi_1'' - KG_1(w_1' + \phi_1) - C_1\ddot{\phi}_1 = 0 \quad \text{in} \quad x \in [a, \alpha], \tag{99}$$

$$\begin{aligned} &KG_2(w_2'' + \phi_2') - B_2\ddot{w}_2 - \chi(w_3 - w_2)\left(\frac{1}{2} - \frac{1}{\pi}\arctan\frac{w_3 - w_2}{\epsilon}\right) \\ &\quad - H_2\phi_2' w_2'' - \frac{3}{2}Q_2(w_2')^2 w_2'' = 0 \quad \text{in} \quad x \in [\alpha, \beta], \end{aligned} \tag{100}$$

$$A_2\phi_2'' - KG_2(w_2' + \phi_2) - C_2\ddot{\phi}_2 - H_2 w_2' w_2'' = 0 \quad \text{in} \quad x \in [\alpha,\ \beta], \tag{101}$$

$$\begin{aligned} &KG_3(w_3'' + \phi_3') - B_3\ddot{w}_3 + \chi(w_3 - w_2)\left(\frac{1}{2} - \frac{1}{\pi}\arctan\frac{w_3 - w_2}{\epsilon}\right) \\ &\quad - H_3\phi_3' w_3'' - \frac{3}{2}Q_3(w_3')^2 w_3'' = 0 \quad \text{in} \quad x \in [\alpha,\ \beta]. \end{aligned} \tag{102}$$

$$A_3\phi_3'' - KG_3(w_3' + \phi_3) - C_3\ddot{\phi}_3 - H_3 w_3' w_3'' = 0 \quad \text{in} \quad x \in [\alpha,\ \beta], \tag{103}$$

$$KG_4(w_4'' + \phi_4') - B_4\ddot{w}_4 = 0 \quad \text{in} \quad x \in [\beta, L], \tag{104}$$

$$A_4\phi_4'' - KG_4(w_4' + \phi_4) - C_4\ddot{\phi}_4 = 0 \quad \text{in} \quad x \in [\beta, L]. \tag{105}$$

Essential boundary conditions:

$$R_i = 0 \quad (i = 1, 2, \ldots, 12), \tag{106a}$$

where

$$\begin{aligned} R_1 &\equiv w_0(0,t), & R_2 &\equiv \phi_0(0,t), \\ R_3 &\equiv w_0(a,t) - w_1(a,t), & R_4 &\equiv \phi_0(a,t) - \phi_1(a,t), \\ R_5 &\equiv w_1(\alpha,t) - w_2(\alpha,t), & R_6 &\equiv \phi_1(\alpha,t) - \phi_2(\alpha,t), \\ R_7 &\equiv w_1(\alpha,t) - w_3(\alpha,t), & R_8 &\equiv \phi_1(\alpha,t) - \phi_3(\alpha,t), \\ R_9 &\equiv w_2(\beta,t) - w_4(\beta,t), & R_{10} &\equiv \phi_2(\beta,t) - \phi_4(\beta,t), \\ R_{11} &\equiv w_3(\beta,t) - w_4(\beta,t), & R_{12} &\equiv \phi_3(\beta,t) - \phi_4(\beta,t). \end{aligned} \tag{106b}$$

Natural boundary conditions:

$$G_0(\phi_0 + w_0') - G_1(\phi_1 + w_1') = 0 \quad \text{at} \quad x = a, \tag{107}$$

$$A_0\phi_0' - A_1\phi_1' = 0 \quad \text{at} \quad x = a, \tag{108}$$

$$\begin{aligned} &KG_1(\phi_1 + w_1') - KG_2(\phi_2 + w_2') - H_2 w_2'\phi_2' - \frac{1}{2}\,Q_2(w_2')^3 \\ &- KG_3(\phi_3 + w_3') - H_3 w_3'\phi_3' - \frac{1}{2}\,Q_3(w_3')^3 = 0 \quad \text{at} \quad x = \alpha, \end{aligned} \tag{109}$$

$$A_1\phi_1' - A_2\phi_2' - \frac{1}{2}\,H_2(w_2')^2 - A_3\phi_3' - \frac{1}{2}\,H_3(w_3')^2 = 0 \quad \text{at} \quad x = \alpha, \tag{110}$$

$$\begin{aligned} &KG_2(\phi_2 + w_2') + H_2 w_2'\phi_2' + \frac{1}{2}\,Q_2(w_2')^3 + KG_3(\phi_3 + w_3') \\ &+ H_3 w_3'\phi_3' + \frac{1}{2}\,Q_3(w_3')^3 - KG_4(\phi_4 + w_4') = 0 \quad \text{at} \quad x = \beta, \end{aligned} \tag{111}$$

$$A_2\phi_2' + A_3\phi_3' - A_4\phi_4' = 0 \quad \text{at} \quad x = \beta, \tag{112}$$

$$\phi_4 + w_4' = 0 \quad \text{at} \quad x = L, \tag{113}$$

$$\phi_4' = 0 \quad \text{at} \quad x = L. \tag{114}$$

4 Finite Element Formulation

The finite element formulation is made on the basis of weak forms for the derived partial differential equations (96)–(105).

Finite element within Zone 0 (Part 0), i.e. within a subdomain $x \in [0, a]$.

The weak form for a finite element within Zone 0 is obtained by multiplying equations (96) and (97) by weight functions (variations) δw_0 and $\delta\phi_0$ respectively, integrating them over an element's length, performing integration by parts and adding the resulting equations. The weak form thus obtained is

$$\begin{aligned} 0 = \int_{X_A}^{X_B} &\Big[A_0\phi_0'\,\delta\phi_0' + KG_0(w_0' + \phi_0)\,\delta w_0' + KG_0(w_0' + \phi_0)\,\delta\phi_0 \\ &+ B_0\ddot{w}_0\,\delta w_0 + C_0\ddot{\phi}_0\,\delta\phi_0 - I_p V(t)\,\delta_{a-\varepsilon}(x)\,\delta\phi_0\Big]dx \\ &+ \Big[KG_0(w_0' + \phi_0)\,\delta w_0\Big]_{x=X_A} - \Big[KG_0(w_0' + \phi_0)\,\delta w_0\Big]_{x=X_B} \\ &+ \Big(A_0\phi_0'\,\delta\phi_0\Big)_{x=X_A} - \Big(A_0\phi_0'\,\delta\phi_0\Big)_{x=X_B}, \end{aligned} \tag{115}$$

where X_A and X_B are coordinates of the element's left and right boundary points. In the boundary terms of the weak form, variations of the unknown functions w_0 and ϕ_0 themselves (not their derivatives) are present, therefore, the Lagrange interpolation polynomials are appropriate for approximation of the unknown functions within a finite element [3]. In this analysis, the author chose to approximate both unknown functions $w_0(x,t)$ and $\phi_0(x,t)$, within a finite element, by the Lagrange interpolation polynomials of a fifth degree:

$$w_0(x,t) \approx \sum_{i=1}^{6} N_i(x) w_{0i}(t), \qquad \phi_0(x,t) \approx \sum_{i=1}^{6} N_i(x)\phi_{0i}(t), \tag{116}$$

where

$$w_{0i}(t) \equiv w_0(x_i,\, t), \quad \phi_{0i}(t) \equiv \phi_0(x_i,\, t), \tag{117}$$

$$N_i(x) = \prod_{\substack{j=1 \\ j\neq i}}^{6} \frac{x - x_j}{x_i - x_j}, \tag{118}$$

$$x_1 \equiv X_A, \quad x_6 \equiv X_B. \tag{119}$$

So, the finite element has six nodes, and two unknown nodal parameters w_{0i} and ϕ_{0i} are associated with each i-th node. The nodes are chosen to be equidistant. Denoting the element's length as l, the nodal coordinates, in the local element coordinate system (the origin of which coincides with the left boundary point of the element), can be written as

$$x_i = \frac{(i-1)l}{5} \quad (i = 1, \ldots, 6). \tag{120}$$

Explicit expressions for the shape functions are written below

$$\begin{aligned}
N_1(x) &= -\frac{625}{24l^5}x^5 + \frac{625}{8l^4}x^4 - \frac{2125}{24l^3}x^3 + \frac{375}{8l^2}x^2 - \frac{137}{12l}x + 1,\\
N_2(x) &= \frac{3125}{12}\frac{x^5}{l^5} - 625\frac{x^4}{l^4} + \frac{6625}{12}\frac{x^3}{l^3} - \frac{425}{2}\frac{x^2}{l^2} + 30\frac{x}{l},\\
N_3(x) &= -\frac{3125}{12}\frac{x^5}{l^5} + \frac{8125}{12}\frac{x^4}{l^4} - \frac{7375}{12}\frac{x^3}{l^3} + \frac{2675}{12}\frac{x^2}{l^2} - 25\frac{x}{l},\\
N_4(x) &= \frac{3125}{12}\frac{x^5}{l^5} - 625\frac{x^4}{l^4} + \frac{6125}{12}\frac{x^3}{l^3} - \frac{325}{2}\frac{x^2}{l^2} + \frac{50}{3}\frac{x}{l},\\
N_5(x) &= -\frac{3125}{24}\frac{x^5}{l^5} + \frac{6875}{24}\frac{x^4}{l^4} - \frac{5125}{24}\frac{x^3}{l^3} + \frac{1525}{24}\frac{x^2}{l^2} - \frac{25}{4}\frac{x}{l},\\
N_6(x) &= \frac{625}{24}\frac{x^5}{l^5} - \frac{625}{12}\frac{x^4}{l^4} + \frac{875}{24}\frac{x^3}{l^3} - \frac{125}{12}\frac{x^2}{l^2} + \frac{x}{l}.
\end{aligned} \tag{121}$$

The column-matrix of element nodal parameters is introduced as follows

$$\underset{(12\times1)}{\{\theta\}} \equiv \lfloor w_{01} \quad \phi_{01} \quad w_{02} \quad \phi_{02} \quad w_{03} \quad \phi_{03} \quad w_{04} \quad \phi_{04} \quad w_{05} \quad \phi_{05} \quad w_{06} \quad \phi_{06} \rfloor^{\mathrm{T}}. \tag{122}$$

Then, in view of formulas (116), the unknown functions $w_0(x,t)$ and $\phi_0(x,t)$ can be expressed in terms of the column-matrix of nodal parameters $\{\theta\}$ by the formulas

$$w_0 = \underset{(1\times12)}{\lfloor\Phi\rfloor}\{\theta\}, \qquad \phi_0 = \underset{(1\times12)}{\lfloor\Psi\rfloor}\{\theta\}, \tag{123}$$

where

$$\underset{(1\times12)}{\lfloor\Phi\rfloor} \equiv \lfloor N_1 \quad 0 \quad N_2 \quad 0 \quad N_3 \quad 0 \quad N_4 \quad 0 \quad N_5 \quad 0 \quad N_6 \quad 0 \rfloor, \tag{124a}$$

$$\underset{(1\times12)}{\lfloor\Psi\rfloor} \equiv \lfloor 0 \quad N_1 \quad 0 \quad N_2 \quad 0 \quad N_3 \quad 0 \quad N_4 \quad 0 \quad N_5 \quad 0 \quad N_6 \rfloor. \tag{124b}$$

Substitution of equations (124) into the integral part of the weak form (115) produces the result

$$\underset{(1\times 12)}{\{\delta\theta\}}^{\mathrm{T}}\Big(\underset{(12\times 12)}{[m]}\underset{(12\times 1)}{\{\ddot{\theta}\}}+\underset{(12\times 12)}{[k]}\underset{(12\times 1)}{\{\theta\}}-\underset{(12\times 1)}{\{f\}}\Big)=0, \tag{125}$$

where $\{\delta\theta\}$ is a column-matrix of variations of the nodal parameters, and the other matrices are defined as follows:

element mass matrix:

$$\underset{(12\times 12)}{[m]}=B_0\int_0^l \underset{(12\times 1)}{\lfloor\Phi\rfloor}^{\mathrm{T}}\underset{(1\times 12)}{\lfloor\Phi\rfloor}\,dx+C_0\int_0^l \underset{(12\times 1)}{\lfloor\Psi\rfloor}^{\mathrm{T}}\underset{(1\times 12)}{\lfloor\Psi\rfloor}\,dx, \tag{126}$$

element stiffness matrix:

$$\begin{aligned}\underset{(12\times 12)}{[k]}&=A_0\int_0^l\left(\frac{d}{dx}\underset{(12\times 1)}{\lfloor\Psi\rfloor}^{T}\right)\left(\frac{d}{dx}\underset{(1\times 12)}{\lfloor\Psi\rfloor}\right)dx\\&+KG_0\int_0^l\underset{(12\times 1)}{\left(\frac{d}{dx}\lfloor\Phi\rfloor+\lfloor\Psi\rfloor\right)^{\mathrm{T}}}\underset{(1\times 12)}{\left(\frac{d}{dx}\lfloor\Phi\rfloor+\lfloor\Psi\rfloor\right)}dx,\end{aligned} \tag{127}$$

element force vector for the element adjacent to the right boundary of Zone 0:

$$\underset{(12\times 1)}{\{f\}}=\left\{\begin{matrix}\underset{(11\times 1)}{\{0\}}\\ I_pV(t)\end{matrix}\right\}, \tag{128a}$$

element force vector for all other elements of Zone 0:

$$\underset{(12\times 1)}{\{f\}}=\underset{(12\times 1)}{\{0\}}\,. \tag{128b}$$

Similar derivations can be used for deriving equations of motion of a finite element within Zone 1 (Part 1) and Zone 3 (Part 4).

Finite element within Zone 2 (Part 2 and Part 3), i.e. within a subdomain $x\in[\alpha,\,\beta]$ and $z\in[-h/2,\,\gamma]$.

The weak form for a finite element within Zone 2 is obtained by multiplying equations (100) and (101) by weight functions (variations) δw_2 and $\delta\phi_2$ respectively, multiplying equations (102) and (103) by δw_3 and $\delta\phi_3$ respectively, integrating them over an element's length, performing integration by parts and adding the resulting equations. The integral part of the weak form thus obtained is

$$
\begin{aligned}
0 = \int_0^l & [A_2\phi_2'\,\delta\phi_2' + KG_2(w_2' + \phi_2)\,\delta w_2' + KG_2(w_2' + \phi_2)\,\delta\phi_2 \\
& + B_2\ddot{w}_2\,\delta w_2 + C_2\ddot{\phi}_2\,\delta\phi_2]\,dx \\
& + \int_0^l [A_3\phi_3'\,\delta\phi_3' + KG_3(w_3' + \phi_3)\,\delta w_3' + KG_3(w_3' + \phi_3)\,\delta\phi_3 \\
& + B_3\ddot{w}_3\,\delta w_3 + C_3\ddot{\phi}_3\,\delta\phi_3]\,dx \\
& - \int_0^l \left[\left(H_2\phi_2' + \frac{3}{2}\,Q_2(w_2')^2\right) w_2''\,\delta w_2 + H_2 w_2' w_2''\,\delta\phi_2\right] dx \qquad (129) \\
& - \int_0^l \left[\left(H_3\phi_3' + \frac{3}{2}\,Q_3(w_3')^2\right) w_3''\,\delta w_3 + H_3 w_3' w_3'',\ \delta\phi_3\right] dx \\
& - \int_0^l \chi(w_3 - w_2)\left(\frac{1}{2} - \frac{1}{\pi}\arctan\frac{w_3 - w_2}{\epsilon}\right)\delta w_2\,dx \\
& + \int_0^l \chi(w_3 - w_2)\left(\frac{1}{2} - \frac{1}{\pi}\arctan\frac{w_3 - w_2}{\epsilon}\right)\delta w_3\,dx.
\end{aligned}
$$

The same interpolation polynomials are used for the Zone 2 as for the Zone 0, i.e.

$$
\begin{aligned}
w_2(x,t) &\approx \sum_{i=1}^{6} N_i(x) w_{2i}(t), \qquad w_3(x,t) \approx \sum_{i=1}^{6} N_i(x) w_{3i}(t), \\
\phi_2(x,t) &\approx \sum_{i=1}^{6} N_i(x) \phi_{2i}(t), \qquad \phi_3(x,t) \approx \sum_{i=1}^{6} N_i(x) \phi_{3i}(t),
\end{aligned} \qquad (130)
$$

where

$$
w_{2i}(t) \equiv w_2(x_i,t), \quad w_{3i}(t) \equiv w_3(x_i,t), \quad \phi_{2i}(t) \equiv \phi_2(x_i,t), \quad \phi_{3i}(t) \equiv \phi_3(x_i,t), \quad (131)
$$

and shape functions $N_i(x)$ are defined by equations (121).

The column-matrix of the element nodal parameters for Zone 2 are introduced as follows:

$$
\underset{(24\times 1)}{\{\theta\}} = \left\{ \begin{array}{c} \underset{(12\times 1)}{\{\theta\}^{(2)}} \\ \underset{(12\times 1)}{\{\theta\}^{(3)}} \end{array} \right\}, \qquad (132)
$$

where

$$
\{\theta\}^{(2)} \equiv \lfloor\, w_{21} \quad \phi_{21} \quad w_{22} \quad \phi_{22} \quad w_{23} \quad \phi_{23} \quad w_{24} \quad \phi_{24} \quad w_{25} \quad \phi_{25} \quad w_{26} \quad \phi_{26} \,\rfloor^{\mathrm{T}}, \qquad (133)
$$

$$
\{\theta\}^{(3)} \equiv \lfloor\, w_{31} \quad \phi_{31} \quad w_{32} \quad \phi_{32} \quad w_{33} \quad \phi_{33} \quad w_{34} \quad \phi_{34} \quad w_{35} \quad \phi_{35} \quad w_{36} \quad \phi_{36} \,\rfloor^{\mathrm{T}} \qquad (134)
$$

are column-matrices of nodal parameters of Part 2 and Part 3 respectively (of lower and upper delaminated parts of Zone 2). Then, using the weak form (equation (129)) and following the same procedures as for an element in Zone 0, the following expressions for the element mass and stiffness matrices of Zone 2 are obtained.

Element mass matrix for Zone 2:

$$\underset{(24\times 24)}{[m]} = \begin{bmatrix} \underset{(12\times 12)}{[m]^{(2)}} & \underset{(12\times 12)}{[0]} \\ \underset{(12\times 12)}{[0]} & \underset{(12\times 12)}{[m]^{(3)}} \end{bmatrix}, \tag{135}$$

where

$$\underset{(12\times 12)}{[m]^{(i)}} = B_i \int_0^l \underset{(12\times 1)}{\lfloor \Phi \rfloor^{\mathrm{T}}} \underset{(1\times 12)}{\lfloor \Phi \rfloor} \, dx + C_i \int_0^l \underset{(12\times 1)}{\lfloor \Psi \rfloor^{\mathrm{T}}} \underset{(1\times 12)}{\lfloor \Psi \rfloor} \, dx \quad (i = 2, 3), \tag{136}$$

and row-matrices $\lfloor \Phi \rfloor$ and $\lfloor \Psi \rfloor$ are defined by equations (124).

Element stiffness matrix for Zone 2:

$$\underset{(24\times 24)}{[k]} = \begin{bmatrix} \underset{(12\times 12)}{[k]^{(2)}} & \underset{(12\times 12)}{[0]} \\ \underset{(12\times 12)}{[0]} & \underset{(12\times 12)}{[k]^{(3)}} \end{bmatrix}, \tag{137}$$

where

$$\begin{aligned} \underset{(12\times 12)}{[k]^{(i)}} &= A_i \int_0^l \left(\frac{d}{dx} \underset{(12\times 1)}{\lfloor \Psi \rfloor^{\mathrm{T}}} \right) \left(\frac{d}{dx} \underset{(1\times 12)}{\lfloor \Psi \rfloor} \right) dx \\ &+ KG_i \int_0^l \underset{(12\times 1)}{\left(\frac{d}{dx} \lfloor \Phi \rfloor + \lfloor \Psi \rfloor \right)^{\mathrm{T}}} \underset{(1\times 12)}{\left(\frac{d}{dx} \lfloor \Phi \rfloor + \lfloor \Psi \rfloor \right)} dx \quad (i = 2, 3). \end{aligned} \tag{138}$$

The last two integrals in the weak form (129) represent virtual works of forces of mutual impact of the crack's faces, acting, correspondingly, on the lower and upper crack's face. The computation of contribution of these integrals to the discretized equations of motion of a finite element within Zone 2 is presented below. Let us consider one of these integrals

$$I_2 \equiv \int_0^l \chi(w_3 - w_2) \left(\frac{1}{2} - \frac{1}{\pi} \arctan \frac{w_3 - w_2}{\epsilon} \right) (\delta w_2) \, dx, \tag{140}$$

which represent virtual work of force of impact acting on the lower face of the crack. Substitution of functions $w_2(x,t)$ and $w_3(x,t)$ by their polynomial approximation (equations (130)) yields

$$\begin{aligned} I_2 = \int_0^l \chi \left(\sum_{i=1}^{6} (\delta w_{2i}) N_i \right) \left(\sum_{j=1}^{6} (w_{3j} - w_{2j}) N_j \right) \\ \times \left(\frac{1}{2} - \frac{1}{\pi} \arctan \left(\epsilon^{-1} \sum_{m=1}^{6} (w_{3m} - w_{2m}) N_m \right) \right) dx. \end{aligned} \tag{141}$$

Let the function under the integral sign in the last integral be denoted as $g(x)$. Then, using the trapezoidal rule of numerical integration, with evaluation of the function $g(x)$ at the nodal points $x_1 = 0$, x_2, x_3, x_4, x_5, $x_6 = l$ of the finite element,

$$\int_0^l g(x)\,dx \approx \frac{l}{10}\left[g(0) + g(l) + 2\sum_{k=2}^{5} g(x_k)\right], \tag{142}$$

and using the property $N_i(x_j) = \delta_{ij}$ of the shape functions, defined by equation (118), one can obtain

$$I_2 = \chi \underset{(1\times 12)}{\left\{\delta\theta^{(2)}\right\}^{\mathrm{T}}} \underset{(12\times 1)}{\{f\}}\,, \tag{143}$$

where

$$\begin{gathered} f_i = \frac{l}{10}(w_{3i} - w_{2i})\left(\frac{1}{2} - \frac{1}{\pi}\arctan\frac{w_{3i} - w_{2i}}{\epsilon}\right) \quad \text{for} \quad i = 1, 11, \\ f_i = \frac{l}{5}(w_{3i} - w_{2i})\left(\frac{1}{2} - \frac{1}{\pi}\arctan\frac{w_{3i} - w_{2i}}{\epsilon}\right) \quad \text{for} \quad i = 3,\ 5,\ 7,\ 9, \\ f_i = 0 \quad \text{for} \quad i = 2,\ 4,\ 6,\ 8,\ 10,\ 12. \end{gathered} \tag{144}$$

Similarly, the last integral in equation (129) can be written as

$$I_3 \equiv \int_0^l \chi(w_3 - w_2)\left(\frac{1}{2} - \frac{1}{\pi}\arctan\frac{w_3 - w_2}{\epsilon}\right)(\delta w_3)\,dx = \chi \underset{(1\times 12)}{\left\{\delta\theta^{(3)}\right\}^{\mathrm{T}}} \underset{(12\times 1)}{\{f\}}\,. \tag{145}$$

The nonlinear terms

$$\begin{aligned} &-\int_0^l \left[\left(H_2\phi_2' + \frac{3}{2}\,Q_2(w_2')^2\right)w_2''\,\delta w_2 + H_2 w_2' w_2''\,\delta\phi_2\right]dx \\ &-\int_0^l \left[\left(H_3\phi_3' + \frac{3}{2}\,Q_3(w_3')^2\right)w_3''\,\delta w_3 + H_3 w_3' w_3''\,\delta\phi_3\right]dx \end{aligned}$$

in the weak form (129), which are due to taking account of longitudinal force resultants in the delaminated parts (i.e. due to the von Karman nonlinearity of the strain-displacement relations), lead to the presence of a column-matrix in the equations of motion of a finite element, the components of which depend nonlinearly on the nodal parameters w_{2i} and w_{3i} $(i = 1, 2, \ldots, 6)$. This column-matrix will be denoted as

$$\underset{(24\times 1)}{\{g\}} \equiv \left\{\begin{matrix} \underset{(12\times 1)}{\{g\}^{(2)}} \\ \underset{(12\times 1)}{\{g\}^{(3)}} \end{matrix}\right\}, \tag{146}$$

where $\{g\}^{(2)}$ is a column-matrix the components of which depend nonlinearly on nodal parameters w_{2i} (associated with the lower delaminated part), and $\{g\}^{(3)}$ is a column-matrix the components of which depend nonlinearly on nodal parameters w_{3i} (associated with the upper delaminated part). Components of $\{g\}^{(2)}$ and $\{g\}^{(3)}$ are not written here explicitly, because of their large size.

So, equations of motion of a finite element in the delaminated zone of the beam (Zone 2) have the form

$$\begin{bmatrix} \underset{(12\times12)}{[m]^{(2)}} & \underset{(12\times12)}{[0]} \\ \underset{(12\times12)}{[0]} & \underset{(12\times12)}{[m]^{(3)}} \end{bmatrix} \begin{Bmatrix} \underset{(12\times1)}{\{\ddot{\theta}\}^{(2)}} \\ \underset{(12\times1)}{\{\ddot{\theta}\}^{(3)}} \end{Bmatrix} + \begin{bmatrix} \underset{0(12\times12)}{[k]^{(2)}} & \underset{(12\times12)}{[0]} \\ \underset{(12\times12)}{[0]} & \underset{(12\times12)}{[k]^{(3)}} \end{bmatrix} \begin{Bmatrix} \underset{(12\times1)}{\{\theta\}^{(2)}} \\ \underset{(12\times1)}{\{\theta\}^{(3)}} \end{Bmatrix}$$
$$+\chi \begin{Bmatrix} \underset{(12\times1)}{-\{f\}} \\ \underset{(12\times1)}{\{f\}} \end{Bmatrix} + \begin{Bmatrix} \underset{(12\times1)}{\{g\}^{(2)}} \\ \underset{(12\times1)}{\{g\}^{(3)}} \end{Bmatrix} = \underset{(24\times1)}{\{0\}}\,. \tag{147}$$

In Equation (147), the nonlinear internal force vector $\chi \lfloor -\{f\}^{\mathrm{T}} \quad \{f\}^{\mathrm{T}} \rfloor^{\mathrm{T}}$ depends on nodal parameters, associated with both lower and upper delaminated parts (Part 2 and Part 3). Therefore, in the system of equations (147), the nodal parameters $\{\theta\}^{(2)}$, associated with the lower delaminated part (Part 2) are coupled to the nodal parameters $\{\theta\}^{(3)}$, associated with the upper delaminated part.

So, the derivation of the finite element matrices is completed, and an example problem will be considered next.

5 Solution of Example Problems

As an example problem, a clamped-free wooden beam with the following characteristics (Figure 2.1) is considered: length $L = 20 \times 10^{-2}m$, width $b = 2.76 \times 10^{-2}m$, thickness $h = 0.99 \times 10^{-2}m$, wood density $\rho^{(0)} = 418.02\frac{kg}{m^3}$, Young's modulus of the wood in the direction of fibers $E_1^{(0)} = 1.0897 \times 10^{10}\frac{N}{m^2}$. The piezoelectric actuator is QP10W (Active Control Experts). Thickness of the actuator is $\tau = 3.81 \times 10^{-4}m$, its length is $a = 5.08 \times 10^{-2}m$, the piezoelectric constant in the range of applied voltage (from 0 to $200V$) is $\overline{d}_{31} \approx -1.05 \times 10^{-9}\frac{m}{V}$, the Young's modulus of the actuator with its packaging is $E_1^{(p)} = 2.57 \times 10^{10}\frac{N}{m^2}$, mass density of the actuator with its packaging is $\rho^{(p)} = 6151.1\frac{kg}{m^3}$. The voltage $V(t)$, applied to the piezoelectric actuator, is distributed uniformly along the length of the actuator and varies with time as

$$V(t) = V_a \sin(\Omega t),$$

where $V_a = 200\,V,\ \Omega = 600\frac{1}{s}$. The wooden beam is cut along its fibers, so that the angle θ in the formula (6) is equal to zero, and, therefore, the elastic compliance coefficient $\overline{S}_{11}$ for the wood is equal to $\overline{S}_{11}^{(0)} = \frac{1}{E_1^{(0)}} = 9.1768 \times 10^{-11}\frac{m^2}{N}$. For the piezoelectric actuator, the material coordinate system coincides with the problem coordinate system, so that the elastic compliance coefficient $\overline{S}_{11}$ for the material of the piezo-actuator is $\overline{S}_{11}^{(p)} = \frac{1}{E_1^{(p)}} = 3.8911 \times 10^{-11}\frac{m^2}{N}$. Coordinates of the crack tips are: $\alpha = 10 \times 10^{-2}m$, $\beta = 15 \times 10^{-2}m,\ \gamma = 0.66 \times 10^{-2} - \frac{h}{2} = 1.65 \times 10^{-3}m$. Then the constants, entering into the variational formulation and the differential equations of the problem, have the following values in SI units: $A_0 = 31.463,\ B_0 = 0.178\,9,\ C_0 = 2.642\,9 \times 10^{-6},\ G_0 = 1.299\,10 \times 10^6,\ A_1 = 24.319,\ B_1 = 0.114\,22,\ C_1 = 9.328\,9 \times 10^{-7},\ G_1 = 1.190999 \times 10^6,\ A_2 = 12.61,\ B_2 = 7.614\,7 \times 10^{-2},\ C_2 = 4.837\,2 \times 10^{-7},\ G_2 = 7.93999 \times 10^5,\ A_3 = 11.709,\ B_3 = 3.807\,3 \times 10^{-2},\ C_3 = 4.491\,7 \times 10^{-7},\ G_3 = 3.969995 \times 10^5$, $A_4 = 24.319,\ B_4 = 0.114\,22,\ C_4 = 9.328\,9 \times 10^{-7},\ G_4 = 1.190999 \times 10^6,\ I_p = -3.828\,5 \times 10^{-3},\ a = 5.08 \times 10^{-2},\ V_a = 200,\ \Omega = 600,\ \alpha = 10 \times 10^{-2},\ \beta = 15 \times 10^{-2},\ \gamma = 1.65 \times 10^{-3},\ b = 2.76 \times 10^{-2},\ h = 0.99 \times 10^{-2},\ Q_2 = 1.985005666 \times 10^6,\ Q_3 = 9.925028332 \times 10^5,\ H_2 = -3275.25935,\ H_3 = 3275.25935$. The small constant ϵ and the large constant χ are chosen to be $\epsilon = 1 \times 10^{-3}$ and $\chi = 1 \times 10^6$.

5.1 Time-domain response to dynamic excitation

Time integration of a system of ordinary differential equations of a global (assembled) semi-discrete finite element model

$$[M]\{\ddot{\Theta}\} + [K]\{\Theta\} + \{R\}_{\mathrm{nonlin}} = \{F\}$$

was performed with the use of the backward-difference method [4]. In the last equation, $\{R\}_{\mathrm{nonlin}}$ is a column-matrix, which contains components that depend nonlinearly on the unknown nodal parameters Θ_i. Transverse displacements as functions of time at free ends of delaminated and undelaminated beams are shown in graphs of Figure 5.1. These graphs are noticeably different. Numerical experiments show that neglecting nonlinear terms in the strain-displacement relations (19a), and, therefore, neglecting the longitudinal force resultants $N_x^{(2)}$ and $N_x^{(3)}$ in the delaminated parts of the beam (equations (25) and (29)), produces much smaller effect on the transverse displacement of the delaminated beam than neglecting the force of contact interaction of the crack faces.

5.2 Eigenvalue analysis

For the same beam, natural frequencies and mode shapes were computed from a linear eigenvalue analysis. Results of calculation of frequencies for beams with different crack lengths are presented in tables below. For some crack lengths, comparison is made between frequencies computed on the basis of the first order shear deformation theory, presented in this paper, and the frequencies computed on the basis of the Euler-Bernoulli beam theory, presented in Reference [1]. Rotary inertia terms are taken into account in both types of solutions.

Let us consider, at first, the first seven circular frequencies of a clamped-free beam **without the delamination** and with the actuator, obtained by setting equal the x-coordinates of the crack tips. The frequencies for this case are presented below. Notation FOSDT stands for the First Order Shear Deformation Theory of the beam, notation E-B stands for the Euler-Bernoulli beam theory.

	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7
FOSDT	1395.535	8130.531	21436.8	40361.9	64915.31	9.36739×10^4	1.25461×10^5
E-B	1397.435	8217.911	21986.6	42205.0	69331.23	1.02371×10^5	1.40641×10^5

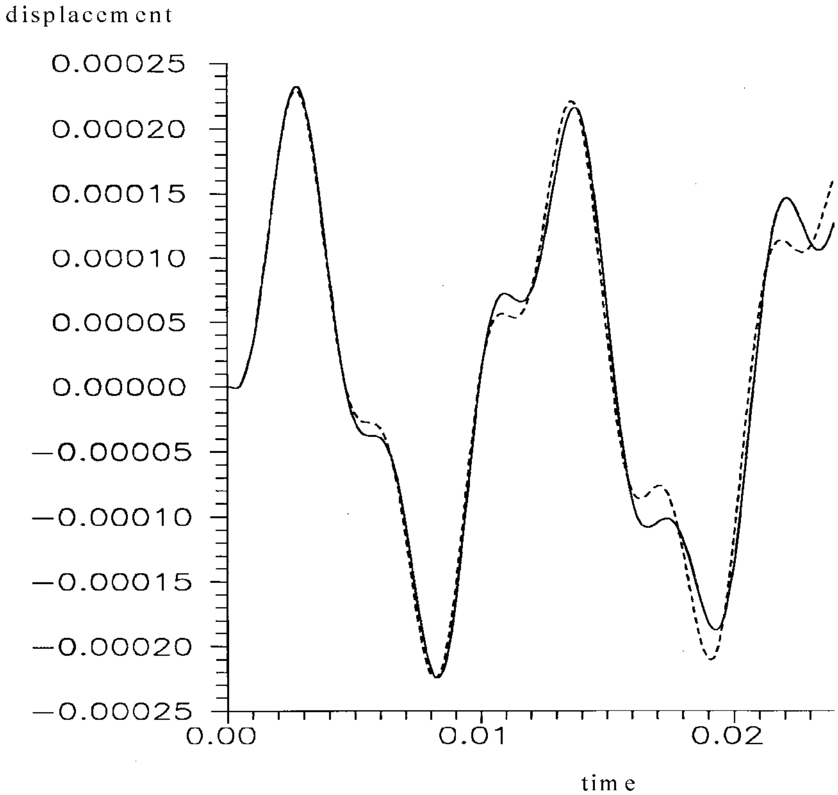

Figure 5.1. Transverse displacement of free end of delaminated beam (solid line) and undelaminated beam (dashed line). Coordinates of the crack tips of the delaminated beam are $\alpha = 0,1m$, $\beta = 0,15m$, $\gamma = 1,65 \times 10^{-3}m$.

In the next table, results are presented for a beam **with the delamination** and with the actuator, with the following coordinates of the crack tips: $\alpha = 0.1m$, $\beta = 0.11m$, $\gamma = 1.65 \times 10^{-3}m$.

	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7
FOSDT	1395.5	8130.5	21436.0	40361.5	64909.9	9.3669×10^4	1.2545×10^5
E-B	1397.435	8217.909	21986.1	42204.9	69331.2	1.02371×10^5	1.40641×10^5

In the next table, results are presented for a beam with the delamination and with the actuator, with the following coordinates of the crack tips: $\alpha = 0.1m$, $\beta = 0.12m$, $\gamma = 1.65 \times 10^{-3}m$.

	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7
FOSDT	1395.5	8130.5	21433.6	40356.8	64900.7	9.3629×10^4	1.2544×10^5
E-B	1397.433	8217.9	21986.0	42200.0	69330.0	1.02368×10^5	1.40625×10^5

In the next table, results are presented for a beam with the delamination and with the actuator, with the following coordinates of the crack tips: $\alpha = 0.1m$, $\beta = 0.13m$, $\gamma = 1.65 \times 10^{-3}m$.

	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7
FOSDT	1395.49	8130.46	21431.655	40345.03	64894.09	9.3576×10^4	1.2535×10^5

In the next table, results are presented for a beam with the delamination and with the actuator, with the following coordinates of the crack tips: $\alpha = 0.1m$, $\beta = 0.14m$, $\gamma = 1.65 \times 10^{-3}m$.

	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7
FOSDT	1395.47	8130.28	21430.371	40330.58	64850.16	9.3541×10^4	1.2504×10^5

In the next table, results are presented for a beam with the delamination and with the actuator, with the following coordinates of the crack tips: $\alpha = 0.1m$, $\beta = 0.15m$, $\gamma = 1.65 \times 10^{-3}m$.

	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7
FOSDT	1395.456	8129.95	21427.848	40320.39	64719.1	9.31501×10^4	1.10100×10^5
E-B	1397.432	8217.62	21980.0	42201.0	69094.0	1.01932×10^5	1.33019×10^5

So, the frequencies decrease as the crack length increases. This phenomenon is more pronounced for higher frequencies.

The first four mode shapes of delaminated beams are nearly the same as the corresponding mode shapes of the undelaminated beams, so that the difference is not noticeable on graphs. But the higher mode shapes of the delaminated beams, beginning from the fifth mode shape, show the crack opening and closure during the vibration, as can be seen in Figures 5.2, 5.3 and 5.4.

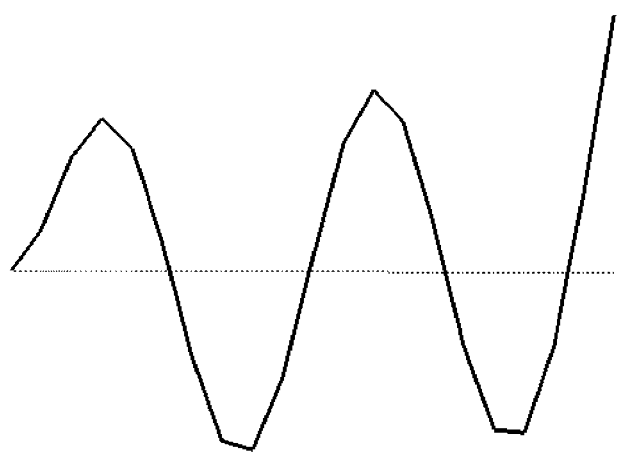

Figure 5.2a. Fifth mode shape of clamped-free beam without delamination.

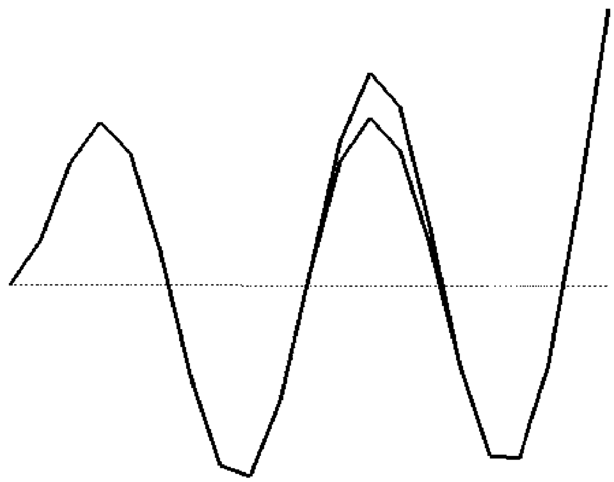

Figure 5.2b. Fifth mode shape of clamped-free beam with delamination.

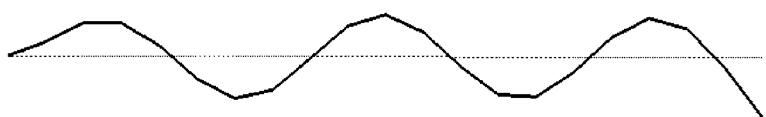

Figure 5.3a. Sixth mode shape of clamped-free beam without delamination.

Figure 5.3b. Sixth mode shape of clamped-free beam with delamination.

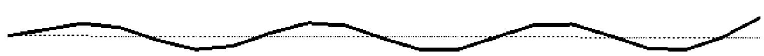

Figure 5.4a. Seventh mode shape of clamped-free beam without delamination.

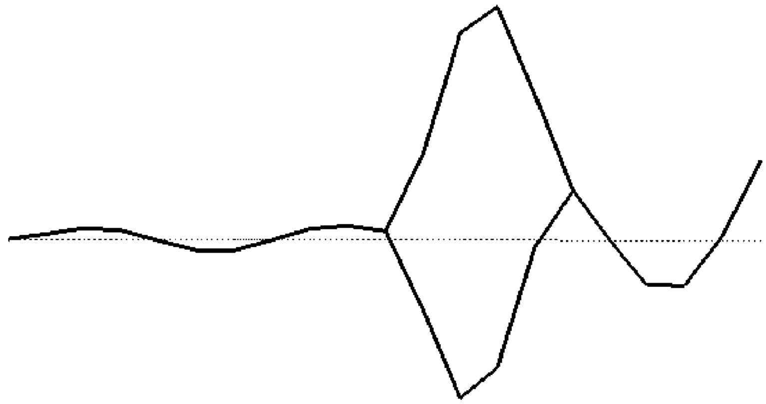

Figure 5.4b. Seventh mode shape of clamped-free beam with delamination.

Experimental verification of the developed theory and the finite element program will be presented in a subsequent publication. The theory, presented in this work, and the finite element program, based on this theory, are developed for the purpose of their subsequent use in nondestructive detection of delamination cracks in composite structures.

Appendix A
Constitutive Equations for a Piezoelectric Orthotropic Layer of a Thin Composite Beam

The constitutive equations of a generally anisotropic piezoelectric material can be written in a matrix form as follows (in these equations, the bars over characters are put to emphasize that the quantities are presented in a problem coordinate system, the coordinate planes of which do not necessarily coincide with the planes of elastic or dielectric symmetry)

$$\underset{(6\times1)}{\{\overline{\varepsilon}\}} = \underset{(6\times6)}{[\overline{S}]}\,\underset{(6\times1)}{\{\overline{\sigma}\}} + \underset{(6\times3)}{[\overline{d}]}^{\mathrm{T}}\,\underset{(3\times1)}{\{\overline{\mathcal{E}}\}}\,, \tag{A-1}$$

$$\underset{(3\times1)}{\{\overline{D}\}} = \underset{(3\times6)}{[\overline{d}]}\,\underset{(6\times1)}{\{\overline{\sigma}\}} + \underset{(3\times3)}{[\overline{\zeta}]}\,\underset{(3\times1)}{\{\overline{\mathcal{E}}\}}\,, \tag{A-2}$$

where

$$\underset{(6\times1)}{\{\overline{\varepsilon}\}} = \lfloor\, \varepsilon_{xx} \quad \varepsilon_{yy} \quad \varepsilon_{zz} \quad 2\varepsilon_{yz} \quad 2\varepsilon_{xz} \quad 2\varepsilon_{xy} \,\rfloor^{\mathrm{T}} \tag{A-3}$$

is a column-matrix of components of the strain tensor,

$$\underset{(6\times1)}{\{\overline{\sigma}\}} = \lfloor\, \sigma_{xx} \quad \sigma_{yy} \quad \sigma_{zz} \quad \sigma_{yz} \quad \sigma_{xz} \quad \sigma_{xy} \,\rfloor^{\mathrm{T}} \tag{A-4}$$

is a column-matrix of components of the stress tensor,

$$\underset{(3\times1)}{\{\overline{\mathcal{E}}\}} = \lfloor\, \mathcal{E}_{x} \quad \mathcal{E}_{y} \quad \mathcal{E}_{z} \,\rfloor^{\mathrm{T}} \tag{A-5}$$

is a column-matrix of components of the electric field intensity vector, $\underset{(6\times 6)}{[\overline{S}]}$ is a matrix of elastic coefficients (compliance coefficients) and $\underset{(3\times 6)}{[\overline{d}]}$ and $\underset{(3\times 3)}{[\overline{\zeta}]}$ are matrices of material constants that characterize electrical properties.

For an orthotropic material, in the principal material coordinate system (whose coordinate planes coincide with the planes of elastic symmetry), the matrix of compliance coefficients is denoted as $[S]$ (without a bar) and has the form

$$[S] = \begin{bmatrix} S_{11} & S_{12} & S_{13} & 0 & 0 & 0 \\ S_{12} & S_{22} & S_{23} & 0 & 0 & 0 \\ S_{13} & S_{23} & S_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & S_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & S_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & S_{66} \end{bmatrix} \tag{A-6}$$

where the compliance coefficients S_{ij} are expressed in terms of engineering constants by the formulas

$$\begin{gathered} S_{11} = \frac{1}{E_1}, \quad S_{12} = -\frac{\nu_{12}}{E_1}, \quad S_{13} = -\frac{\nu_{13}}{E_1}, \quad S_{22} = \frac{1}{E_2}, \quad S_{23} = -\frac{\nu_{23}}{E_2}, \\ S_{33} = \frac{1}{E_3}, \quad S_{44} = \frac{1}{G_{23}}, \quad S_{55} = \frac{1}{G_{13}}, \quad S_{66} = \frac{1}{G_{12}}. \end{gathered} \tag{A-7}$$

The matrices, characterizing electric properties of the material, in the principle material coordinate system, will be denoted without the bar also, i.e. as $\underset{(3\times 6)}{[d]}$ and $\underset{(3\times 3)}{[\zeta]}$.

In the laminate (problem) coordinate system, rotated clockwise by an angle θ with respect to the principle material coordinate system, the matrix of compliance coefficients and the matrices, characterizing electric properties of the material, $\underset{(3\times 6)}{[\overline{d}]}$ and $\underset{(3\times 3)}{[\overline{\zeta}]}$, have the form

$$\underset{(6\times 6)}{[\overline{S}]} = \underset{(6\times 6)}{[T]}^{\mathrm{T}} \underset{(6\times 6)}{[S]} \underset{(6\times 6)}{[T]}, \tag{A-8}$$

$$\underset{(3\times 3)}{[\overline{\zeta}]} = \underset{(3\times 3)}{[R]}^{\mathrm{T}} \underset{(3\times 3)}{[\zeta]} \underset{(3\times 3)}{[R]}, \tag{A-9}$$

$$\underset{(3\times 6)}{[\overline{d}]} = \underset{(3\times 3)}{[R]}^{\mathrm{T}} \underset{(3\times 6)}{[d]} \underset{(6\times 6)}{[T]}, \tag{A-10}$$

where the transformation matrices $[T]$ and $[R]$ are defined as follows (with the use of notation $c = \cos\theta,\ s = \sin\theta$):

$$\underset{(6\times 6)}{[T]} = \begin{bmatrix} c^2 & s^2 & 0 & 0 & 0 & 2sc \\ s^2 & c^2 & 0 & 0 & 0 & -2sc \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & c & -s & 0 \\ 0 & 0 & 0 & s & c & 0 \\ -sc & sc & 0 & 0 & 0 & c^2 - s^2 \end{bmatrix}, \tag{A-11}$$

$$\underset{(3\times 3)}{[R]} = \begin{bmatrix} c & s & 0 \\ -s & c & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{A-12}$$

For the composite piezoelectric layer of a *thin and narrow composite beam, which bends in the x-z plane*, the following assumptions can be adopted

$$\sigma_{zz} = \sigma_{xz} = \sigma_{yz} = \sigma_{yy} = 0. \tag{A-13}$$

Besides, in the problem under consideration, the electrical field is applied to the actuator only in the thickness direction (in the direction of the z-axis), i.e.

$$\mathcal{E}_x = \mathcal{E}_y = 0. \tag{A-14}$$

If equations (A-13) and (A-14) are substituted into the constitutive equations (A-1) and (A-2) with account of transformation relations (A-8), (A-9) and (A-10) and with account of equations (A-6) and (A-7) for compliance matrix in the principle material coordinate system, then the constitutive equations take the form

$$\begin{Bmatrix} \varepsilon_{xx} \\ 2\varepsilon_{xz} \end{Bmatrix} = \begin{bmatrix} \overline{S}_{11} & 0 \\ 0 & \overline{S}_{55} \end{bmatrix} \begin{Bmatrix} \sigma_{xx} \\ \sigma_{xz} \end{Bmatrix} + \begin{bmatrix} \overline{d}_{31} \\ \overline{d}_{35} \end{bmatrix} \{ \mathcal{E}_z \}, \tag{A-15}$$

$$\{ D_z \} = [\overline{d}_{31} \quad \overline{d}_{35}] \begin{Bmatrix} \sigma_{xx} \\ \sigma_{xz} \end{Bmatrix} + [\overline{\zeta}_{33}] \{ \mathcal{E}_z \}. \tag{A-16}$$

From the constitutive equations (A-15) and (A-16), one can obtain the constitutive equations in a different form:

$$\begin{Bmatrix} \sigma_{xx} \\ \sigma_{xz} \\ D_z \end{Bmatrix} = \begin{bmatrix} \dfrac{1}{\overline{S}_{11}} & 0 & -\dfrac{\overline{d}_{31}}{\overline{S}_{11}} \\ 0 & \dfrac{1}{\overline{S}_{55}} & -\dfrac{\overline{d}_{35}}{\overline{S}_{55}} \\ \dfrac{\overline{d}_{31}}{\overline{S}_{11}} & \dfrac{\overline{d}_{35}}{\overline{S}_{55}} & \left(\overline{\zeta}_{33} - \dfrac{\overline{d}_{31}^2}{\overline{S}_{11}} - \dfrac{\overline{d}_{35}^2}{\overline{S}_{55}} \right) \end{bmatrix} \begin{Bmatrix} \varepsilon_{xx} \\ 2\varepsilon_{xz} \\ \mathcal{E}_z \end{Bmatrix}, \tag{A-17a}$$

or, in view of the relationship $\mathcal{E}_z = -\frac{\partial \varphi}{\partial z}$, where φ is the electric potential,

$$\begin{Bmatrix} \sigma_{xx} \\ \sigma_{xz} \\ D_z \end{Bmatrix} = \begin{bmatrix} \dfrac{1}{\overline{S}_{11}} & 0 & \dfrac{\overline{d}_{31}}{\overline{S}_{11}} \\ 0 & \dfrac{1}{\overline{S}_{55}} & \dfrac{\overline{d}_{35}}{\overline{S}_{55}} \\ \dfrac{\overline{d}_{31}}{\overline{S}_{11}} & \dfrac{\overline{d}_{35}}{\overline{S}_{55}} & \left(-\overline{\zeta}_{33} + \dfrac{\overline{d}_{31}^2}{\overline{S}_{11}} + \dfrac{\overline{d}_{35}^2}{\overline{S}_{55}} \right) \end{bmatrix} \begin{Bmatrix} \varepsilon_{xx} \\ 2\varepsilon_{xz} \\ \frac{\partial \varphi}{\partial z} \end{Bmatrix}. \tag{A-17b}$$

According to equations (A-7) and (A-8), the compliance coefficients $\overline{S}_{11}$ and $\overline{S}_{55}$ in the problem coordinate system that enter into equations (A-17), are expressed in terms of the engineering constants by the formulas

$$\begin{aligned} \overline{S}_{55} &= \frac{1}{G_{23}} s^2 + \frac{1}{G_{13}} c^2, \\ \overline{S}_{11} &= \frac{1}{E_1} c^4 + \frac{1}{E_2} s^4 + \left(\frac{1}{G_{12}} - 2\frac{\nu_{12}}{E_1} \right) s^2 c^2. \end{aligned} \tag{A-18}$$

The material constants $\overline{d}_{31}$ and $\overline{d}_{35}$, which characterize the piezoelectric properties in the problem coordinate system, are expressed in terms of the piezoelectric constants d_{ij} of the material coordinate system by the formulas (derived from matrix transformation equations A-10)

$$\overline{d}_{31} = d_{31}c^2 + d_{32}s^2 - d_{36}sc, \tag{A-19a}$$

$$\overline{d}_{35} = -d_{34}s + d_{35}c, \tag{A-19b}$$

and, according to the transformation equation (A-9),

$$\overline{\zeta}_{33} = \zeta_{33}. \tag{A-20}$$

We consider a piezoelectric material with orthorhombic $mm2$ symmetry, such as polyvinylidene (PVDF) or lead zirconate-titanate (PZT), in which the planes of elastic symmetry are made, in the manufacturing process, the same as the planes of piezoelectric symmetry. In this case, the piezoelectric constants d_{34} and d_{35} are equal to zero (see [5] and [6]). Then, according to equation (A-19b), $\overline{d}_{35} = 0$, and equation (A-17b) takes the form

$$\begin{Bmatrix} \sigma_{xx} \\ \sigma_{xz} \\ D_z \end{Bmatrix} = \begin{bmatrix} \dfrac{1}{\overline{S}_{11}} & 0 & \dfrac{\overline{d}_{31}}{\overline{S}_{11}} \\ 0 & \dfrac{1}{\overline{S}_{55}} & 0 \\ \dfrac{\overline{d}_{31}}{\overline{S}_{11}} & 0 & \left(-\overline{\zeta}_{33} + \dfrac{\overline{d}_{31}^{\,2}}{\overline{S}_{11}} \right) \end{bmatrix} \begin{Bmatrix} \varepsilon_{xx} \\ 2\varepsilon_{xz} \\ \frac{\partial \varphi}{\partial z} \end{Bmatrix}. \tag{A-21}$$

These are the constitutive equations for a layer of orthotropic piezoelectric material with orthorhombic $mm2$ symmetry, in which the planes of elastic symmetry are the same as the planes of piezoelectric symmetry, in a narrow and thin composite beam. Obviously, for a layer of orthotropic material, in a thin narrow beam, which does not have piezoelectric properties, the constitutive equations have the form

$$\begin{Bmatrix} \sigma_{xx} \\ \sigma_{xz} \end{Bmatrix} = \begin{bmatrix} \dfrac{1}{\overline{S}_{11}} & 0 \\ 0 & \dfrac{1}{\overline{S}_{55}} \end{bmatrix} \begin{Bmatrix} \varepsilon_{xx} \\ 2\varepsilon_{xz} \end{Bmatrix}. \tag{A-22}$$

Appendix B
Properties of the Heaviside Function

It can be shown [7] that the Heaviside function (unit step-function) $\mathfrak{H}_\alpha(x)$, defined by formula

$$\mathfrak{H}_\alpha(x) = \begin{cases} 0 & \text{for } \ x < \alpha, \\ 1 & \text{for } \ x > \alpha, \end{cases} \tag{B-1}$$

has the following property

$$\frac{d\mathfrak{H}_\alpha(x)}{dx} = \delta_\alpha(x), \tag{B-2}$$

where $\delta_\alpha(x)$ is the Dirac's delta-function, defined as a function that has the following properties:

$$\delta_\alpha(x) = \begin{cases} 0 & \text{for } \; x \neq \alpha, \\ \infty & \text{for } \; x = \alpha \end{cases} \tag{B-3}$$

and

$$\int\limits_{x_1}^{x_2} f(x)\delta_\alpha(x)\,dx = \begin{cases} f(\alpha) & \text{for } \; x_1 < \alpha < x_2, \\ 0 & \text{for } \; \alpha < x_1 \text{ and for } \alpha > x_2. \end{cases} \tag{B-4}$$

The delta-function has several analytical representations, one of which has the form [8]:

$$\delta_\alpha(x) = \lim_{\epsilon \to 0} \frac{1}{\pi} \frac{\epsilon}{\epsilon^2 + (x-\alpha)^2}. \tag{B-5}$$

According to formula (B-2), the analytical representation of the Heaviside function, corresponding to the analytical representation of the delta-function (B-5) is

$$\mathfrak{H}_\alpha(x) = \lim_{\epsilon \to 0} \frac{1}{\pi} \arctan \frac{x-\alpha}{\epsilon} + \frac{1}{2} = \begin{cases} 0 & \text{for } \; x < \alpha, \\ \dfrac{1}{2} & \text{for } \; x = \alpha, \\ 1 & \text{for } \; x > \alpha. \end{cases} \tag{B-6}$$

Carrying out the Heaviside function $\mathfrak{H}_\alpha(x)$ beyond the integral sign in an indefinite integral is done with the use of the formula

$$\int \mathfrak{H}_\alpha(x)\, f(x)\,dx = \mathfrak{H}_\alpha(x) \int\limits_{\alpha}^{x} f(\eta)\,d\eta. \tag{B-7}$$

With the use of properties (B-2) and (B-4), it can be shown that

$$\int\limits_{x_1}^{x_2} f(x)\, \frac{d^2\mathfrak{H}_\alpha(x)}{dx^2}\,dx = \begin{cases} -\dfrac{df}{dx}\,(\alpha) & \text{for } \; x_1 < \alpha < x_2, \\ 0 & \text{for } \; \alpha < x_1 \; \text{ and for } \; \alpha > x_2. \end{cases} \tag{B-8}$$

References

[1] Perel, V.Y. A numerical-analytical solution for dynamics of composite delaminated beam with piezoelectric actuator, with account of nonpenetration constraint for the delamination crack faces. *Journal of Composite Materials* **39**(1) (2005) 67–103.

[2] Reddy, J.N. *Energy and Variational Methods in Applied Mechanics.* John Wiley & Sons, New York, 1984.

[3] Reddy, J.N. *An Introduction to the Finite Element Method.* McGraw-Hill, New York, 1993.

[4] Hairer, E., Norsett, S.P. and Wanner, G. *Solving Ordinary Differential Equations. Vol. 2: Stiff And Differential-Algebraic Problems.* Springer, Berlin, 1996.

[5] Varadan, V.V., Roh, Y.R., Varadan, Y.R. and Tancrell, R.H. Measurement of all the elastic and dielectric constants of poled PVDF films. In: *1989 Ultrasonics Symposium.* (http://www.ieee-uffc.org/archive/ul/a98.htm). IEEE, Montreal, Quebec, Canada, 1989, 727–730.

[6] Dunn, M.L. and Taya, M. Micromechanics predictions of the effective electroelastic moduli of piezoelectric composites. *Int. J. of Solids and Structures* **30**(2) (1993) 161–175.

[7] Dirac, P.A.M. *The Principles of Quantum Mechanics.* Clarendon Press, Oxford, 1978.

[8] Rubinowicz, A. *Quantum Mechanics.* Elsevier Pub. Co., 1968.

Nonlinear Dynamics and Systems Theory, **5**(1) (2005) 91–105

Tracking Control of General Nonlinear Systems by a Direct Gradient Descent Method

K. Shimizu[1], S. Ito[2] and S. Suzuki[1]

[1] *Faculty of Science and Technology, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama 223-8522, Japan*
[2] *Center for Development of Statistical Computing, The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106-8569, Japan*

Received: September 8, 2003; Revised: June 15, 2004

Abstract: This paper is concerned with tracking control of nonlinear multivariable systems whose relative degree is more than one. The control method is based on a direct steepest descent method using the gradient of a performance index. Simulation results demonstrate the usefulness of the proposed method.

Keywords: *Nonlinear control; state feedback; output feedback; tracking system; gradients; steepest descent; Lyapunov stability.*

Mathematics Subject Classification (2000): 93B52, 93C10, 90C48, 49K40, 93D05.

1 Introduction

Studies on nonlinear feedback control have been extensively made in recent years. Needless to say, stabilization and optimization are central concerns. Lyapunov stability theory and the Hamilton–Jacobi–Bellman equation for optimal control appear to be main tools for designing stabilizing feedback control laws. For affine nonlinear systems, lots of researches have been done based on feedback linearization, nonlinear optimal regulator [13, 5], the Hamilton–Jacobi–Bellman equation and inverse optimality theory [4], control Lyapunov function stabilization [20], back stepping technique [11], nonlinear H^∞ control and passivity-based control theory [22], etc. For general nonlinear systems, receding horizon control [12] is known as one of the few studies on on-line nonlinear optimal control. Besides, there are many studies on neuro-controllers [19, 15] based on the error back propagation method, but their stability and generalization ability remain unsolved questions.

In this paper the following general nonlinear system is considered as a controlled object:

$$\begin{aligned} \dot{x}(t) &= f(x(t), u(t)), \\ y(t) &= h(x(t)), \end{aligned} \tag{1}$$

where $x(t) \in R^n$ is the state vector, $u(t) \in R^r$ is the control input, $y(t) \in R^m$ is the measured output, and we study the problem of output tracking so that $y(t)$ tracks a desired output $y_d(t)$ (the reference signal). Output tracking control (output regulation or servo mechanism) for nonlinear systems has intensively been investigated [23, 9, 21, 7, 10, 6, 1]. Among them, Vidyasagar [23] and Tsinias [21] showed that if the system (1) is stabilizable and weakly detectable by means of a continuous state feedback $u(t) = \alpha(x(t))$, then the system is also stabilized by $\alpha(z(t))$, where $z(t)$ is the output of a weak detector for the state $x(t)$. More precisely, when

$$\dot{z}(t) = g(z(t), y(t), u(t))$$

is an observer (i.e., $z(t) - x(t) \to 0$ as $t \to \infty$ for every $x(0)$ and $z(0)$) and $u(t) = \alpha(z(t))$ is an asymptotically stabilizing control law, then the closed-loop system

$$\begin{aligned} \dot{x}(t) &= f(x(t), \alpha(z(t))), \qquad y(t) = h(x(t)), \\ \dot{z}(t) &= g(z(t), y(t), \alpha(z(t))) \end{aligned}$$

is asymptotically stable in a neighborhood of $(x, z) = (0, 0)$. Note that, however, it is another hard task to obtain the state feedback law $u = \alpha(x)$.

In the pioneering work of Isidori and Byrnes [9], nonlinear output regulation problem has been formulated and solved, in which the objective is to design a dynamic controller such that the closed-loop system is stable and the error approaches zero asymptotically. Supposing the reference signal $y_d(t)$ to be generated by the exosystem

$$\dot{w}(t) = s(w(t)), \quad y_d(t) = q(w(t))$$

and considering the extended system

$$\begin{aligned} \dot{x}(t) &= f(x(t), u(t)), \qquad & y(t) &= h(x(t)), \\ \dot{w}(t) &= s(w(t)), & y_d(t) &= q(w(t)), \\ e(t) &= y_d(t) - y(t) \end{aligned} \tag{2}$$

they solved the output regulation problem by means of an error feedback controller (a dynamic controller)

$$\begin{aligned} \dot{z}(t) &= \eta(z(t), e(t)), \\ u(t) &= \alpha(z(t)). \end{aligned} \tag{3}$$

More precisely, the output regulation means that the unforced closed-loop system with $w = 0$ is exponentially stable and that the forced closed-loop system (2) − (3) satisfies $\lim_{t \to \infty} e(t) \to 0$ for any initial condition $(x(0), z(0), w(0))$ in a neighborhood of the origin $(0, 0, 0)$. Isidori and Byrnes [9] derived a necessary condition for the output regulation, called the nonlinear regulator equation, using the center manifold theorem. Though the Isidori–Byrnes theory is precise and sophisticated, it requires many assumptions and, in order to synthesize a solution numerically, one has to solve the nonlinear regulator equation described by a system of nonlinear partial differential equations, which is difficult to solve as in the Hamilton–Jacobi–Bellman equation. The nonlinear output regulation can achieve asymptotic disturbance rejection based on the exosystem as well as asymptotic output tracking. Actually, in the work of Isidori and Byrnes [9] trajectory tracking and/or disturbance rejection are unificatively formulated as the problem of output regulation. Furthermore, structurally stable and robust output regulation under parametric uncertainties has been investigated by Khalil [10], Huang [6] and Byrnes, *et al.* [1]. Huang and Rugh [7] also proposed an approximation method of finding a power series expansion of the solution to the nonlinear regulator equation.

Direct gradient descent control was proposed by Shimizu, *et al.* [18], which directly manipulates control inputs so as to decrease a performance index such as the squared error from a desired equilibrium state based on the gradient of the performance index with respect to the control inputs. The gradient is derived from sensitivity equations. A similar method called "speed gradient control" was also proposed by Fradkov, *et al.* [2, 3]. In their method, however, the performance function F contains only x (not both x and u). $F(x)$ and $F(x, u)$ makes a big difference in application. Further, their derivation is not based on the sensitivity equations but on the Lyapunov direct method.

In this paper we investigate output tracking control of nonlinear multivariable systems by use of the direct gradient descent method. Our main concern is the control of plants with relative degrees of more than one. The proposed method is an on-line implementation and can be executed in a very simple and practical manner. Our simulation results for various plants showed remarkably good performance, one of which will be demonstrated in the last section.

2 Direct Gradient Descent Control of Nonlinear Systems

The aim of our control is to modify $u(t)$ so that a performance index $F(y(t), u(t))$ decreases. The problem is written as

$$\underset{u(t)}{\text{decrease}}\, F(y(t), u(t)), \tag{4a}$$

$$\text{subj. to} \quad \dot{x}(t) = f(x(t), u(t)), \quad x(t_0) = x_0 \tag{4b}$$

$$y(t) = h(x(t)) \tag{4c}$$

where we make the following assumption:

Assumption 1 *Plant* (4b), (4c) is *locally controllable and observable*.

To solve this problem, we prepare some fundamental results concerning the gradient of the performance index. We confine our attention in this section to the basic state feedback regulation case:

$$\underset{u(t)}{\text{decrease}}\, F(x(t), u(t)), \tag{5a}$$

$$\text{subj. to} \quad \dot{x}(t) = f(x(t), u(t)), \quad x(t_0) = x_0, \tag{5b}$$

where we assume the following:

Assumption 2 *Function* f is *continuously differentiable*; f_u, F_x and F_u are *Lipschitz continuous.*

For any continuous $u\colon\ u(t),\ t \geq t_0$, system (4b) has a unique smooth solution $x\colon\ x(t)$, $t \geq t_0$. We denote the state trajectory x associated with a given u by $x(u)$, whose value at t will be denoted by $x(t;u)$. Then, for an arbitrarily fixed t, let us define a functional ϕ^t by

$$\phi^t[u] \triangleq F(x(t;u), u(t)). \tag{6}$$

The derivative of the objective $F(x(t;u), u(t))$ with respect to $u(t)$ can be conceptually given as

$$F_x(x(t;u), u(t))\frac{dx(t;u)}{du(t)} + F_u(x(t;u), u(t)). \tag{7}$$

Here the notion $dx(t;u)/du(t)$ denotes the effect on $x(t;u)$ caused by the change of $u(t)$, but it is impossible and impractical to change $u(t)$ freely without any reference to the past trajectory of u. So we consider a time interval $[t', t]$, where t' is an arbitrarily given time such that $t_0 \leq t' < t$, and see the effect on the state at time t caused by the change of u as a function on the interval.

As a class of admissible control for the fixed interval $[t', t]$, we consider the space $U_{[t',t]}$ consisting of r-dimensional vector-valued continuous functions and define the inner product:

$$\langle u, v\rangle_{U_{[t',t]}} \triangleq \int_{t'}^{t} u(\tau)^{\mathrm{T}} v(\tau)\, d\tau. \tag{8}$$

Then the following theorem holds.

Theorem 1 *The operator* $x(t;\cdot)\colon U_{[t',t]} \to R^n$ *is Gâteaux differentiable, and its Jacobian is given, at time* t, *as follows:*

$$\nabla x(t;u)(t) = f_u(x(t;u), u(t))^{\mathrm{T}}. \tag{9}$$

Proof We show that the functional $x(t;\cdot)\colon U_{[t',t]} \to R^n$ is Gâteaux differentiable, and calculate the Gâteaux differential

$$\delta x(t;u;s) \triangleq \frac{d}{d\varepsilon} x(t; u + \varepsilon s)\Big|_{\varepsilon=0}.$$

Integrating (5b) from t' to t with $u + \varepsilon s$, we have

$$x(t;\, u+\varepsilon s) = x(t') + \int_{t'}^{t} f(x(\tau;\, u+\varepsilon s),\, u(\tau) + \varepsilon s(\tau))\, d\tau. \tag{10}$$

Differentiating (10) w.r.t. ε, letting $\varepsilon = 0$, and differentiating it w.r.t. t, we finally obtain

$$\frac{d}{dt}\frac{d}{d\varepsilon} x(t; u+\varepsilon s)\Big|_{\varepsilon=0} = f_x(x(t;u), u(t))\, \frac{d}{d\varepsilon} x(t; u+\varepsilon s)\Big|_{\varepsilon=0} + f_u(x(t;u), u(t))s(t)$$

with

$$\frac{d}{d\varepsilon}\, x(t'; u+\varepsilon s)\Big|_{\varepsilon=0} = 0.$$

Since this is a time-variant linear differential equation w.r.t.

$$\delta x(t;u;s) = \frac{d}{d\varepsilon} x(t;u+\varepsilon s)\bigg|_{\varepsilon=0},$$

its solution exists and is given by

$$\delta x(t;u;s) = \int_{t'}^{t} \Phi(t,\tau) f_u(x(\tau;u),u(\tau))s(\tau)\,d\tau$$

where Φ is a continuous transition-matrix function defined on $\{(t,\tau)\colon\ t' \le \tau \le t\}$ by

$$\frac{\partial}{\partial t}\Phi(t,\tau) = f_x(x(t;u),u(t))\Phi(t,\tau), \quad \Phi(\tau,\tau) = I \tag{11}$$

(see, e.g., Pontryagin [14]). The Gâteaux differential of each component $x_i(t;\cdot)\colon\ U_{[t',t]} \to R$ is then expressed as

$$\delta x_i(t;u;s) = \int_{t'}^{t} \Phi_i(t,\tau) f_u(x(\tau;u),u(\tau))s(\tau)\,d\tau,$$

where $\Phi_i(t,\tau)$ denotes the i-th row of $\Phi(t,\tau)$. Comparing this with definition (8), we can see that there exists $\nabla x_i(t;u) \in U_{[t',t]}$ satisfying

$$\delta x_i(t;u;s) = \langle \nabla x_i(t;u),\ s\rangle_{U_{[t',t]}} \quad \forall\, s \in U_{[t',t]}$$

and it is given by

$$\nabla x_i(t;u)(\tau) = f_u(x(\tau;u),u(\tau))^{\mathrm{T}}\Phi_i(t,\tau)^{\mathrm{T}}, \quad \tau \in [t',t].$$

Each $\nabla x_i(t;u)$ is an r-dimensional vector-valued function, and here we define an $(r\times n)$-matrix-valued function $\nabla x(t;u)$ by

$$\nabla x(t;u)(\tau) \triangleq (\nabla x_1(t;u)(\tau), \dots, \nabla x_n(t;u)(\tau)).$$

In other words, $\nabla x(t;u)$ is given by

$$\nabla x(t;u)(\tau) = f_u(x(\tau;u),u(\tau))^{\mathrm{T}}\Phi(t,\tau)^{\mathrm{T}}, \quad \tau \in [t',t]. \tag{12}$$

It follows from (11) and (12) that

$$\begin{gathered}\frac{d}{dt}\nabla x(t;u)(\tau) = \nabla x(t;u)(\tau) f_x(x(t;u),u(t))^{\mathrm{T}},\\ \nabla x(\tau;u)(\tau) = f_u(x(\tau;u),u(\tau))^{\mathrm{T}}\end{gathered} \tag{13}$$

on the region $\{(t,\tau)\colon\ t' \le \tau \le t\}$, from which we obtain (9). It is noted that equation (13) represents the sensitivity equation of the state x with respect to the input u.

Since (9) does not depend on t', we regard it as the effect $dx(t;u)/du(t)$ in (7) and consider the transpose of (7), i.e.,

$$f_u(x(t;u),u(t))^{\mathrm{T}}F_x(x(t;u),u(t))^{\mathrm{T}}+F_u(x(t;u),u(t))^{\mathrm{T}},$$

as the *gradient* of the objective $F(x(t;u),u(t))$ with respect to $u(t)$. We denote it by $\nabla\phi^t[u](t)$, i.e.,

$$\nabla\phi^t[u](t)=f_u(x(t;u),u(t))^{\mathrm{T}}F_x(x(t;u),u(t))^{\mathrm{T}}+F_u(x(t;u),u(t))^{\mathrm{T}}. \tag{14}$$

As an on-line control law for problem (5), we apply the steepest descent method at each time $t\in[t_0,\infty)$ by using $\nabla\phi^t[u](t)$. Namely, $u(t)$ is modified by the direct gradient descent control algorithm

$$\dot{u}(t)=-\mathcal{L}\nabla\phi^t[u](t) \tag{15}$$

where $\mathcal{L}=\operatorname{diag}[\alpha_1,\alpha_2,\ldots,\alpha_r]$, $\alpha_i>0$, is a proportional constant. Substituting (14) into (15) yields

$$\dot{u}(t)=-\mathcal{L}\big\{f_u(x(t;u),u(t))^{\mathrm{T}}F_x(x(t;u),u(t))^{\mathrm{T}}+F_u(x(t;u),u(t))^{\mathrm{T}}\big\}. \tag{16}$$

Assumption 2 is a sufficient condition for systems (5b) and (16) to be solvable for a unique smooth pair (x,u). Furthermore, in order to realize this control, we set F to satisfy the following assumption:

Assumption 3 For every i,

$$F_{x_i}(x,u)\neq 0 \qquad \forall\,(x,u)\neq(x_d,u_d)$$

where x_d is a desired stationary state and u_d is the corresponding control.

Let us set the performance index F in problem (5) as a quadratic form. Our purpose of control is then to transfer the state $x(t)$ to a desired stationary state x_d. At the stationary state, it must hold that $0=f(x_d,u_d)$. In general, we can arbitrarily specify r of n components of x_d, but the remaining $(n-r)$ components and u_d are dependently determined. We consider

$$F(x(t),u(t))\triangleq(x_d-x(t))^{\mathrm{T}}Q(x_d-x(t))+(u_d-u(t))^{\mathrm{T}}R(u_d-u(t)) \tag{17}$$

as a performance index to be decreased, where Q and R are (normally diagonal) positive definite matrices. Then the gradient is written as

$$\nabla\phi^t[u](t)=-2f_u(x(t;u),u(t))^{\mathrm{T}}Q(x_d-x(t;u))-2R(u_d-u(t)) \tag{18}$$

and hence the direct gradient descent control formula (16) is given by

$$\dot{u}(t)=2\mathcal{L}\big\{f_u(x(t;u),u(t))^{\mathrm{T}}Q(x_d-x(t;u))+R(u_d-u(t))\big\}. \tag{19}$$

The stability of direct gradient descent control is proved in Appendix by use of Lyapunov's direct method.

3 Output Tracking via Direct Gradient Descent Control

We define the inverse dynamics of nonlinear systems using the concept of relative degree of nonlinear dynamical systems (see, e.g., [8]). Let us consider each component $y_i(t)$ of $y(t)$, and denote by $y_i^{(j)}(t)$ the j-th order derivative of $y_i(t)$ with respect to t, which generally represents a function of x, u, $\dot{u}$, $\ddot{u},\ldots,u^{(j-1)}$. Then the relative degree of $y_i(t)$ is defined as follows.

Definition 1 The integer q_i satisfying

$$\frac{\partial y_i^{(j)}(t)}{\partial u} = 0, \quad j = 1, 2, \dots, q_i - 1, \tag{20a}$$

$$\frac{\partial y_i^{(q_i)}(t)}{\partial u} \neq 0 \tag{20b}$$

is called the *relative degree* of component $y_i(t)$.

Let us denote by $\alpha_i^j(x(t))$ and $\beta_i^{q_i}(x(t), u(t))$ the j-th derivative of $y_i(t)$ as $j = 1, 2, \dots, q_i - 1$ and as $j = q_i$, respectively. Then we have the following system of equations:

$$\begin{aligned} y_1^{(q_1)}(t) &= \beta_1^{q_1}(x(t), u(t)) \\ \cdots&\cdots\cdots\cdots \\ y_m^{(q_m)}(t) &= \beta_m^{q_m}(x(t), u(t)). \end{aligned} \tag{21}$$

Assumption 4

$$\operatorname{rank} \begin{bmatrix} \dfrac{\partial \beta_1^{q_1}(x(t), u(t))}{\partial u} \\ \cdots\cdots\cdots\cdots \\ \dfrac{\partial \beta_m^{q_m}(x(t), u(t))}{\partial u} \end{bmatrix} = r.$$

Then, by the implicit function theorem, there exists an inverse mapping of (21) in regard to $u(t)$. Hence $u(t)$ can be expressed as

$$u(t) = \eta\big(x(t), y_1^{(q_1)}(t), \dots, y_i^{(q_i)}(t), \dots, y_m^{(q_m)}(t)\big) \tag{22}$$

where q_i denotes the relative degree of $y_i(t)$. Let us call system (22) the inverse dynamics or the inverse system. The control $u(t)$ represented by (22) can be regarded as an input by which the q_i-th order derivative of $y_i(t)$ becomes equal to $y_i^{(q_i)}(t)$ when $x(t)$ is the present state.

Now we investigate an on-line tracking control for problem (4) based on the preliminary knowledge on state feedback regulation. Let us consider the case where the performance index is given in the quadratic form

$$F(y(t), u(t)) \triangleq (y_d(t) - y(t))^{\mathrm{T}} Q (y_d(t) - y(t)) + (u_d(t) - u(t))^{\mathrm{T}} R (u_d(t) - u(t)), \tag{23}$$

where $y_d(t)$ is a desired output, $u_d(t)$ is the corresponding control input, and Q, R are diagonal positive definite matrices. In what follows, we consider the functional

$$\phi^t[u] \triangleq F(y(t; u), u(t)) \tag{24}$$

where $y(t; u) \triangleq h(x(t; u))$. We assume sufficiently higher order continuous differentiability of f and h for a while. Precise description of required assumptions will be given at the end of the next section. Applying Theorem 1, we obtain an expression of the gradient needed for the gradient descent tracking control as follows:

$$\begin{aligned}\nabla\phi^t[u](t) &= \nabla x(t;u)(t)\,\frac{\partial y(t;u)}{\partial x}^{\mathrm{T}} F_y(y(t;u),u(t))^{\mathrm{T}} + F_u(y(t;u),u(t))^{\mathrm{T}} \\ &= -2f_u(x(t;u),u(t))^{\mathrm{T}}\,\frac{\partial h(x(t;u))}{\partial x}^{\mathrm{T}} Q(y_d(t)-y(t;u)) \\ &\quad - 2R(u_d(t)-u(t)).\end{aligned} \tag{25}$$

From this we realize that, for $y_i(t)$ with relative degree of more than 1, the error $(y_{id}(t)-y_i(t))$ cannot be evaluated at all in the calculation of $\nabla\phi^t[u](t)$ since

$$f_u(x(t),u(t))^{\mathrm{T}}\,\frac{\partial h_i(x(t))}{\partial x}^{\mathrm{T}} = 0.$$

Therefore the error information on $y_i(t)$ is not used in modifying $u(t)$ by the direct gradient descent control with (25), which implies that it is not always possible to accomplish the tracking control.

In order to control those plants with higher relative degrees, it is essential to incorporate some information on higher order derivatives into the algorithm, and we consider the following performance index:

$$\begin{aligned}F\big(y_1^{(q_1-1)}(t),\dots,y_m^{(q_m-1)}(t),u(t)\big) \triangleq \sum_{i=1}^{m}\omega_i\big(y_{id}^{(q_i-1)}(t)-y_i^{(q_i-1)}(t)\big)^2 \\ + (u_d(t)-u(t))^{\mathrm{T}}R(u_d(t)-u(t)),\end{aligned} \tag{26}$$

where $y_{id}^{(q_i-1)}(t)$ denotes the (q_i-1)-th order derivative of the i-th component $y_{id}(t)$ of the desired output (q_i is the relative degree of $y_i(t)$). Taking account of the inverse dynamics given by (22), it seems that we need the q_i-th order derivative for each output component, but, actually, the (q_i-1)-th order derivative turns out to be enough by the nature of the direct gradient descent control and by the definition of relative degree. We again use the same notation

$$\phi^t[u] \triangleq F\big(y_1^{(q_1-1)}(t;u),\dots,y_m^{(q_m-1)}(t;u),u(t)\big). \tag{27}$$

The gradient $\nabla\phi^t[u](t)$ is then given by

$$\begin{aligned}\nabla\phi^t[u](t) &= \sum_{i=1}^{m}\nabla x(t;u)(t)\,\frac{\partial y_i^{(q_i-1)}(t;u)}{\partial x}^{\mathrm{T}} F_{y_i^{(q_i-1)}} + F_u^{\mathrm{T}} \\ &= -2\sum_{i=1}^{m}\omega_i f_u(x(t;u),u(t))^{\mathrm{T}}\,\frac{\partial\alpha_i^{q_i-1}(x(t;u))}{\partial x}^{\mathrm{T}}\big(y_{id}^{(q_i-1)}(t)-y_i^{(q_i-1)}(t;u)\big) \\ &\quad - 2R(u_d(t)-u(t))\end{aligned} \tag{28}$$

where we eliminated the arguments of F for simplicity. The direct gradient descent control is given as follows:

$$\begin{aligned}\dot u(t) &= -\alpha \nabla \phi^t[u](t)\\ &= 2\alpha\biggl[\sum_{i=1}^{m} \omega_i f_u(x(t;u),u(t))^{\mathrm T}\,\frac{\partial \alpha_i^{q_i-1}(x(t;u))}{\partial x}^{\mathrm T}\bigl(y_{id}^{(q_i-1)}(t) - y_i^{(q_i-1)}(t;u)\bigr)\\ &\quad + R(u_d(t)-u(t))\biggr].\end{aligned} \tag{29}$$

4 Convergence of the Output Error

Execution of (29) can enforce $y_i^{(q_i-1)}(t) \to y_{id}^{(q_i-1)}(t)$ for each i, but this does not guarantee that $y_i(t) \to y_{id}(t)$ when $q_i > 1$. In this section we shall utilize some device so that $y(t)$ can asymptotically converge to $y_d(t)$ whenever $y_i^{(q_i-1)}(t) \to y_{id}^{(q_i-1)}(t)$ for all i's.

Let us first consider a component $y_i(t)$ whose relative degree is 2. If we use $\dot{\tilde y}_{id}(t) \triangleq \dot y_{id}(t) + a_{i,0}(y_{id}(t) - y_i(t))$ instead of $\dot y_{id}(t)$ in (26) or (29), we obtain $\dot y_i(t) \to \dot{\tilde y}_{id}(t)$ and hence $\dot y_i(t) = \dot y_{id}(t) + a_{i,0}(y_{id}(t) - y_i(t))$, i.e., $\dot y_{id}(t) - \dot y_i(t) = -a_{i,0}(y_{id}(t) - y_i(t))$ for sufficiently large t. The tracking error $e_i(t) = y_{id}(t) - y_i(t)$ then satisfies $\dot e_i(t) = -a_{i,0}\, e_i(t)$, and hence, if $a_{i,0} > 0$, we can expect that $e_i(t) \to 0$ (i.e., $y_i(t) \to y_{id}(t)$) as $t \to \infty$.

In a similar manner, let us consider the general case where the relative degree is q_i. If we use

$$\begin{aligned}\tilde y_{id}^{(q_i-1)}(t) \triangleq y_{id}^{(q_i-1)}(t) &+ a_{i,q_i-2}\bigl(y_{id}^{(q_i-2)}(t) - y_i^{(q_i-2)}(t)\bigr) + \cdots\\ &+ a_{i,1}(\dot y_{id}(t) - \dot y_i(t)) + a_{i,0}(y_{id}(t) - y_i(t))\end{aligned} \tag{30}$$

instead of $y_{id}^{(q_i-1)}(t)$, we can expect $y_i^{(q_i-1)}(t) \to \tilde y_{id}^{(q_i-1)}(t)$, and hence the tracking error $e_i(t) = y_{id}(t) - y_i(t)$ asymptotically satisfies

$$e_i^{(q_i-1)}(t) + a_{i,q_i-2} e_i^{(q_i-2)}(t) + \cdots + a_{i,1}\dot e_i(t) + a_{i,0} e_i(t) = 0.$$

If $a_{i,j}$, $j = 0, 1, \dots, q_i - 2$, are chosen so that every root of the characteristic equation

$$\lambda^{q_i-1} + a_{i,q_i-2}\lambda^{q_i-2} + \cdots + a_{i,1}\lambda + a_{i,0} = 0$$

is real negative, then we have $e_i(t) \to 0$ (i.e., $y_i(t) \to y_{id}(t)$) as $t \to \infty$. The output $y(t)$ can thus track the desired output $y_d(t)$ asymptotically. (Such an idea was also suggested in [24] and [15] for the case with relative degree 1.) If we consider $y_d(t)$ as the output of a reference model, this method can also be regarded as a model reference tracking control in which $y(t)$ asymptotically follows $y_d(t)$.

Hence we modify the performance index as follows:

$$\begin{aligned}&F\bigl(y_1(t), \dot y_1(t), \dots, y_1^{(q_1-1)}(t), \dots, y_m(t), \dot y_m(t), \dots, y_m^{(q_m-1)}(t), u(t)\bigr)\\ &\qquad \triangleq \sum_{i=1}^{m} w_i\bigl(\tilde y_{id}^{(q_i-1)}(t) - y_i^{(q_i-1)}(t)\bigr)^2 + (u_d(t)-u(t))^{\mathrm T} R(u_d(t)-u(t))\end{aligned} \tag{31}$$

where $\tilde{y}_{id}^{(q_i-1)}(t)$ is defined by (30). Letting $\phi^t[u]$ denote the performance index (31), we have

$$\nabla\phi^t[u](t) = \sum_{i=1}^{m}\sum_{j=0}^{q_i-1} \nabla x(t;u)(t)\,\frac{\partial y_i^{(j)}(t;u)}{\partial x}^{\mathrm{T}} F_{y_i^{(j)}} + F_u^{\mathrm{T}}.$$

Noting (9) and (20a) and substituting (31), we obtain the gradient for the quadratic case:

$$\begin{aligned}\nabla\phi^t[u](t) = & -2\sum_{i=1}^{m} w_i f_u(x(t;u),u(t))^{\mathrm{T}}\,\frac{\partial\alpha_i^{q_i-1}(x(t;u))}{\partial x}^{\mathrm{T}} \\ & \times\left\{\sum_{k=0}^{q_i-1} a_{i,k}\big(y_{id}^{(k)}(t) - y_i^{(k)}(t;u)\big)\right\} - 2R(u_d(t)-u(t)).\end{aligned} \tag{32}$$

Finally from (15) and (32) we have the following direct gradient descent control for output tracking:

$$\begin{aligned}\dot u(t) = 2\alpha\Bigg[& \sum_{i=1}^{m} w_i f_u(x(t;u),u(t))^{\mathrm{T}}\,\frac{\partial\alpha_i^{q_i-1}(x(t;u))}{\partial x}^{\mathrm{T}} \\ & \times\left\{\sum_{k=0}^{q_i-1} a_{i,k}\big(y_{id}^{(k)}(t) - y_i^{(k)}(t;u)\big)\right\} + R(u_d(t)-u(t))\Bigg].\end{aligned} \tag{33}$$

Remark Let f and f_u be Lipschitz continuous; let f be $(\max_i q_i - 1)$-times continuously differentiable in x with Lipschitz continuous derivatives; let h_i, $i = 1, 2, \dots, m$, be q_i-times continuously differentiable with Lipschitz continuous derivatives; let y_{id}, $i = 1, 2, \dots, m$, be q_i-times continuously differentiable and u_d be continuously differentiable. Then a simultaneous system of (1) and (33) has a unique smooth solution for arbitrarily given initial condition.

5 Simulation Results

Let us consider a link of length $2l$ and weight m, at one end of which a torque $\tau(t)$ is added as a control input. The single-link manipulator system is then described by $I\ddot\theta(t) + D\dot\theta(t) - mlg\sin\theta(t) = \tau(t)$, where θ is the angle of rotation, I is the moment of inertia of the link, and D is the viscous friction coefficient at the other end of the link. Letting $\theta(t) = x_1(t)$, $\dot\theta(t) = x_2(t)$, $\tau(t) = u(t)$, we have

$$\begin{aligned}\dot x_1(t) &= x_2(t),\\ \dot x_2(t) &= -\frac{D}{I}\,x_2(t) + \frac{mlg}{I}\,\sin x_1(t) + \frac{1}{I}\,u(t).\end{aligned}$$

We consider this nonlinear plant with output $y(t) = x_1(t)$, whose relative degree is 2. The gradient descent control formula is then given by

$$\begin{aligned}\dot u(t) &= 2\alpha\left[\frac{w}{I}\big\{(\dot y_d(t) - \dot y(t)) + a_0(y_d(t) - y(t))\big\} + R(u_d(t) - u(t))\right]\\ &= 2\alpha\left[\frac{w}{I}\big\{(x_{2d}(t) - x_2(t)) + a_0(x_{1d}(t) - x_1(t))\big\} + R(u_d(t) - u(t))\right].\end{aligned}$$

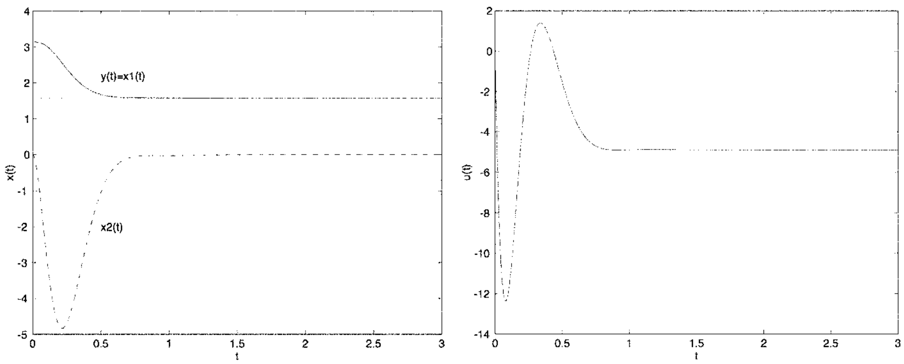

Figure 5.1. $y_d(t) = \pi/2$.

Case 1: $y_d(t) = \pi/2$.

Any equilibrium point (x_{1d}, x_{2d}, u_d) must satisfy $0 = x_{2d}$ and $0 = mlg \sin x_{1d} + u_d$. We set the system parameters as $l = 0.5$, $m = 1$, $I = 1/3$, $D = 0.00198$, and applied the direct steepest descent control with $\alpha = 10$, $a_0 = 3$, $R = w = 1$. The result is shown in Figure 5.1 for initial values $x(0) = (\pi, 0)^{\mathrm{T}}$, $u(0) = 0$, and desired values $(x_{1d}(t), x_{2d}(t), u_d(t)) = (\pi/2, 0, -mlg)$.

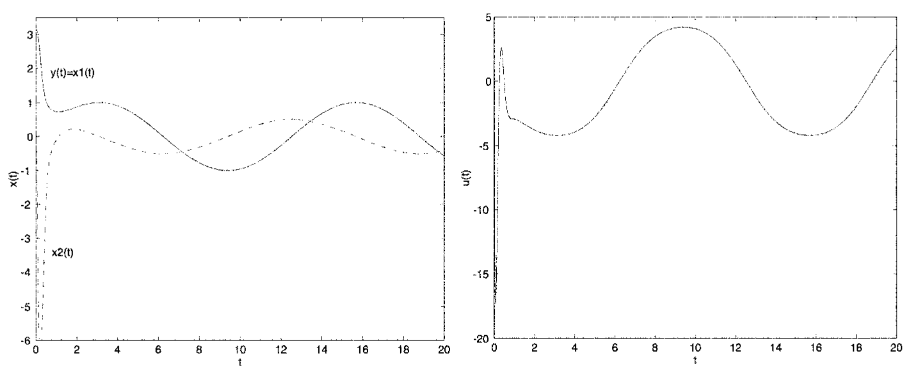

Figure 5.2. $y_d(t) = \sin 0.5t$.

Case 2: $y_d(t) = \sin 0.5t$.

The corresponding desired states $(x_{1d}(t), x_{2d}(t))$ and control $u_d(t)$ must satisfy

$$\dot{x}_{1d}(t) = x_{2d}(t),$$

$$\dot{x}_{2d}(t) = -\frac{D}{I}\, x_{2d}(t) + \frac{mlg}{I}\, \sin x_{1d}(t) + \frac{1}{I}\, u_d(t).$$

By substituting $x_{1d}(t) = \sin 0.5t$ here, we obtain

$$\begin{aligned}(x_{1d}(t), x_{2d}(t), u_d(t)) = (&\sin 0.5\,t,\ 0.5 \cos 0.5\,t,\\ &- 0.25 I \sin 0.5\,t + 0.5 D \cos 0.5\,t - mlg \sin(\sin 0.5\,t)).\end{aligned}$$

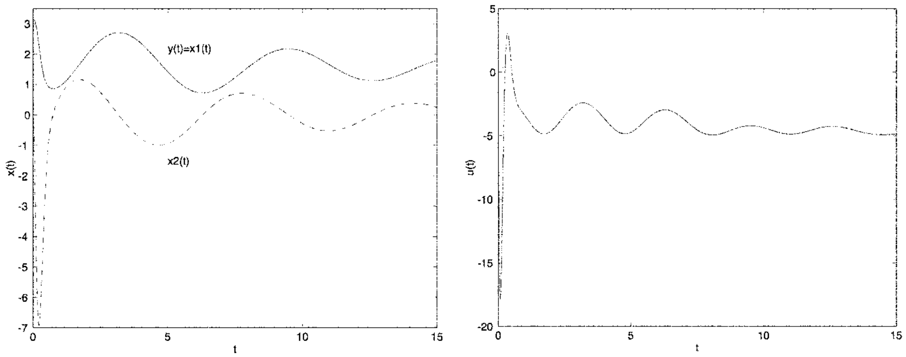

Figure 5.3. $y_d(t) = \dfrac{\pi}{2}\left\{1 - \dfrac{1}{\omega}\,e^{-\zeta t}(\zeta \sin \omega t + \omega \cos \omega t)\right\}$, where $\omega = \sqrt{1-\zeta^2}$, $\zeta = 0.1$.

Figure 5.2 shows the result for the same initial values by the same control parameters as in the previous case except $\alpha = 20$.

Case 3: $y_d(t) = \dfrac{\pi}{2}\left\{1 - \dfrac{1}{\omega}\,e^{-\zeta t}(\zeta \sin \omega t + \omega \cos \omega t)\right\}$**, where** $\omega = \sqrt{1-\zeta^2}$**,** $\zeta = 0.1$**.**

This reference $y_d(t)$ corresponds to the response of a second-order linear system with damping ratio ζ, zero initial states, and forced input $\pi/2$. For $0 < \zeta < 1$, each $y_d(t)$ generates an oscillating signal converging to $\pi/2$. A result for $\zeta = 0.1$ is shown in Figure 5.3 for the same initial states and control parameters as in Case 1.

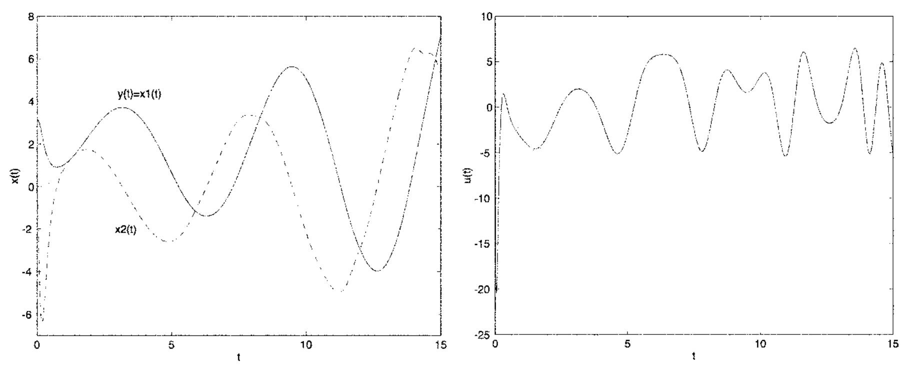

Figure 5.4. $y_d(t) = \dfrac{\pi}{2}\left\{1 - \dfrac{1}{\omega}\,e^{-\zeta t}(\zeta \sin \omega t + \omega \cos \omega t)\right\}$, where $\omega = \sqrt{1-\zeta^2}$, $\zeta = -0.1$.

Case 4: $y_d(t) = \dfrac{\pi}{2}\left\{1 - \dfrac{1}{\omega}\,e^{-\zeta t}(\zeta \sin \omega t + \omega \cos \omega t)\right\}$**, where** $\omega = \sqrt{1-\zeta^2}$**,** $\zeta = -0.1$**.**

For $\zeta < 0$, the reference $y_d(t)$ gives a divergent signal oscillating around $\pi/2$. Figure 5.4 shows a result for $\zeta = -0.1$ when the same initial states and control parameters are applied except $\alpha = 20$.

6 Concluding Remarks

We proposed the direct gradient descent control for tracking of general nonlinear systems with relative degrees of more than one. The effectiveness of the control method was confirmed by computer simulation, and very good performance was observed with various examples. Results for a Rayleigh model are given in [17]. In regard to stability, the direct gradient descent control is considered fairly stable since the resultant control inputs are always manipulated so as to decrease the squared error of outputs. We also observed that the choice of the proportional coefficient $\mathcal{L}$ did not seriously affect the stability, but the direct gradient descent control does not guarantee the monotone decrease of performance index. It is difficult to theoretically verify the stability of the proposed method in general. For individual plants, however, we can find some asymptotically stable region in a neighborhood of the desired equilibrium by constructing a Lyapunov function via Zubov's successive approximation method [25] as shown in [18]. Stability is guaranteed as long as the plant is controlled within that region.

Appendix: Stability

In this appendix, we establish the stability of the direct gradient descent control for the state feedback case, in which the control law is given by

$$\dot{x}(t) = f(x(t), u(t)), \tag{34}$$

$$\dot{u}(t) = -\alpha\big\{f_u(x(t), u(t))^{\mathrm{T}} F_x(x(t), u(t))^{\mathrm{T}} + F_u(x(t), u(t))^{\mathrm{T}}\big\}, \quad \alpha > 0. \tag{35}$$

As a performance index F, we consider the most practical quadratic error function

$$F(x(t), u(t)) = \frac{1}{2}(x_d - x(t))^{\mathrm{T}} Q(x_d - x(t)) + \frac{1}{2}(u_d - u(t))^{\mathrm{T}}(u_d - u(t)), \tag{36}$$

where $Q > 0$, and (x_d, u_d) is a desired equilibrium point, and assume:

Assumption 5 Plant (34) is Lyapunov asymptotically stable for the fixed u_d. That is, for a Lyapunov function

$$V_1(x) = \frac{1}{2}(x_d - x)^{\mathrm{T}} Q(x_d - x),$$

there exists a positive definite function σ such that

$$V_{1x}(x) f(x, u_d) = -(x_d - x)^{\mathrm{T}} Q f(x, u_d) \le -\sigma(\|x_d - x\|).$$

Assumption 6 The function $V_{1x}(x) f(x, u) = -(x_d - x)^{\mathrm{T}} Q f(x, u)$ is convex with respect to u. (This always holds for affine nonlinear systems.)

Assumption 5 is a sufficient condition for the internal stability of plant (34) when the input is fixed to u_d, and this implies that the equilibrium point x_d of the plant $\dot{x}(t) = f(x(t), u_d)$ is asymptotically stable.

Under these assumptions, one can show the asymptotical stability of extended system (34) and (35) by means of Lyapunov's direct method in a similar way to [2]. Let us consider a Lyapunov function candidate

$$V(x,u) = \frac{1}{2}\,\alpha(x_d - x)^{\mathrm{T}}Q(x_d - x) + \frac{1}{2}(u_d - u)^{\mathrm{T}}(u_d - u) > 0 \qquad \forall\,(x,u) \neq (x_d, u_d). \quad (37)$$

For (34) and (35), the time derivative of $V(x,u)$ is given by

$$\begin{aligned}\frac{dV(x,u)}{dt} &= V_x(x,u)f(x,u) - V_u(x,u)\alpha\{f_u(x,u)^{\mathrm{T}}F_x(x,u)^{\mathrm{T}} + F_u(x,u)^{\mathrm{T}}\}\\ &= -\alpha(x_d - x)^{\mathrm{T}}Qf(x,u) - \alpha(x_d - x)^{\mathrm{T}}Qf_u(x,u)(u_d - u) - \alpha(u_d - u)^{\mathrm{T}}(u_d - u).\end{aligned}$$

On the other hand, since, by Assumption 6, $V_x(x,u)f(x,u) = \alpha V_{1x}(x)f(x,u) = -\alpha(x_d - x)^{\mathrm{T}}Qf(x,u)$ is convex with respect to u, we have

$$-\alpha(x_d - x)^{\mathrm{T}}Qf(x,u_d) \geq -\alpha(x_d - x)^{\mathrm{T}}Qf(x,u) - \alpha(x_d - x)^{\mathrm{T}}Qf_u(x,u)(u_d - u).$$

We thus obtain

$$\frac{dV(x,u)}{dt} \leq -\alpha(x_d - x)^{\mathrm{T}}Qf(x,u_d) - \alpha(u_d - u)^{\mathrm{T}}(u_d - u).$$

Since the first term of the right-hand side is negative definite by Assumption 5, we have $dV(x,u)/dt < 0$ for all $(x,u) \neq (x_d, u_d)$. The system (34) and (35) is hence asymptotically stable by the Lyapunov's theorem, i.e., $x(t) \to x_d$ and $u(t) \to u_d$ as $t \to \infty$.

References

[1] Byrnes, C.I., Priscoli, F.Delli, Isidori, A. and Kang, W. Structurally stable output regulation of nonlinear systems. *Automatica* **33**(3) (1997) 369–385.

[2] Fradkov, A.L. Swing control of nonlinear oscillations. *Int. J. of Control* **64**(6) (1996) 1189–1202.

[3] Fradkov, A.L. and Pogromsky, A.Yu. Speed gradient control of chaotic continuous-time systems. *IEEE Trans. on Circuits and Systems I: Fundamental Theory and Appl.* **43**(11) (1996) 907–913.

[4] Freeman, R.A. and Kokotović, P.V. *Robust Nonlinear Control Design: State-Space and Lyapunov Techniques*. Birkhäuser, Boston, 1996.

[5] Goh, C.J. On the nonlinear optimal regulator problem. *Automatica* **29**(3) (1993) 751–756.

[6] Huang, J. Asymptotic tracking and disturbance rejection in uncertain nonlinear systems. *IEEE Trans. on Automatic Control* **40**(6) (1995) 1118–1122.

[7] Huang, J. and Rugh, W.J. An approximation method for the nonlinear servomechanism problem. *IEEE Trans. on Automatic Control* **37**(9) (1992) 1395–1398.

[8] Isidori, A. *Nonlinear Control Systems*. Springer-Verlag, London, 1995.

[9] Isidori, A. and Byrnes, C.I. Output regulation of nonlinear systems. *IEEE Trans. on Automatic Control* **35**(2) (1990) 131–140.

[10] Khalil, H.K. Robust servomechanism output feedback controllers for feedback linearizable systems. *Automatica* **30**(10) (1994) 1587–1599.

[11] Kristic, M., Kanellakopoulos, I. and Kokotović, P. *Nonlinear and Adaptive Control Design.* Wiley, New York, 1995.

[12] Mayne, D.Q. and Michalska, H. Receding horizon control of nonlinear systems. *IEEE Trans. on Automatic Control* **35**(7) (1990) 814– 824.

[13] Moylan, P.J. and Anderson, B.D.O. Nonlinear regulator theory and an inverse optimal control problem. *IEEE Trans. on Automatic Control* **18**(5) (1973) 460–465.

[14] Pontryagin, L.S. *Ordinary Differential Equations.* Addison-Wesley, 1962.

[15] Saerens, M., Renders, J.M. and Bersini, N. Neuro control based on the backpropagation algorithm. In: *Intelligent Control Systems.* (Eds.: M.M. Gupta and N.K. Sinha). IEEE Press, 1996, ch. 12, P.118–128.

[16] Shimizu, K. and Ito, S. Constrained optimization in Hilbert space and a generalized dual quasi-Newton algorithm for state-constrained optimal control problems. *IEEE Trans. on Automatic Control* **39**(5) (1994) 982–986.

[17] Shimizu, K., Ito, S. and Suzuki, S. Tracking control of general nonlinear systems by direct gradient descent method. In: *Proc. of the IFAC Nonlinear Control Systems Design Symposium 1998.* (Eds.: H.J.C. Huijberts, H. Nijmeijer, A.J. van der Schaft and J.M.A. Scherpen). Int. Federation of Automatic Control, 1998, Vol. 1, P.185,–190.

[18] Shimizu, K., Nukumi, H. and Ito, S. Direct steepest descent control of nonlinear dynamical systems. In: *Nonlinear Control Systems Design 1995.* (Eds.: A.J. Krener and D.Q. Mayne). Pergamon, 1996, P.801–806.

[19] Shimizu, K. and Ohtani, M. Optimal control for nonlinear systems by a neural controller of the state feedback type. In: *Proc. of the 35th IEEE Conf. on Decision and Control.* Institute of Electrical and Electronics Engineers, 1996, Vol. 3, P.3300–3303.

[20] Sontag, E.D. *Mathematical Control Theory: Deterministic Finite Dimensional Systems.* Springer-Verlag, New York, 1990.

[21] Tsinias, J. A generalization of Vidyasagar's theorem on stabilizability using state detection. *Systems and Control Letters* **17** (1991) 37–42.

[22] van der Schaft, A. *L_2-Gain and Passivity Techniques in Nonlinear Control.* Springer-Verlag, London, 1996.

[23] Vidyasagar, M. On the stabilization of nonlinear systems using state detection. *IEEE Trans. on Automatic Control* **25**(3) (1980) 504–509.

[24] Yamashita, Y., Shima, M. and Isurugi, Y. Learning and adaptive control using neural networks. *Trans. of the Society of Instrument and Control Engineers* **30**(4) (1991) 302–308. [Japanese].

[25] Zubov, V.I. *Methods of A.M. Lyapunov and Their Applications.* P. Noordhoff, Groningen, 1994.

NONLINEAR DYNAMICS AND SYSTEMS THEORY

An International Journal of Research and Surveys

INSTRUCTIONS FOR CONTRIBUTORS

(1) General. The Journal will publish original carefully refereed papers, brief notes and reviews on a wide range of nonlinear dynamics and systems theory problems. Contributions will be considered for publication in ND&ST if they have not been published previously. Before preparing your submission, it is essential that you consult our style guide; please visit our website: http://www.sciencearea.com.ua

(2) Manuscript and Correspondence. Contributions are welcome from all countries and should be written in English. Two copies of the manuscript, double spaced one column format, and the electronic version by AMSTEX, TEX or LATEX program (on diskette) should be sent directly to

Professor A.A. Martynyuk
Institute of Mechanics,
Nesterov str.3, 03057, MSP 680
Kiev-57, Ukraine
(e-mail: anmart@stability.kiev.ua).

or to one of the Editors or to a member of the Editorial Board.

The title of the article must include: author(s) name, name of institution, department, address, FAX, and e-mail; an Abstract of 50-100 words should not include any formulas and citations; key words, and AMS subject classifications number(s). The size for regular paper 10-14 pages, survey (up to 24 pages), short papers, letter to the editor and book reviews (2-3 pages).

(3) Tables, Graphs and Illustrations. All figures must be suitable for reproduction without being retouched or redrawn and must include a title. Line drawings should include all relevant details and should be drawn in black ink on plain white drawing paper. In addition to a hard copy of the artwork, it is necessary to attach a PC diskette with files of the artwork (preferably in PCX format).

(4) References. Each entry must be cited in the text by author(s) and number or by number alone. All references should be listed in their alphabetic order. Use please the following style:

Journal: [1] Poincaré, H. Title of the article. *Title of the Journal* **Vol.1**(No.1) (year) pages. [Language].

Book: [2] Liapunov, A.M. *Title of the book.* Name of the Publishers, Town, year.

Proceeding: [3] Bellman, R. Title of the article. In: *Title of the book.* (Eds.). Name of the Publishers, Town, year, pages. [Language].

(5) Proofs and Reprints. Proofs sent to authors should be returned to the Editor with corrections within three days after receipt. Acceptance of the paper entitles the author to 10 free reprints.

(6) Editorial Policy. Every paper is reviewed by the regional editor, and/or a referee, and it may be returned for revision or rejected if considered unsuitable for publication.

(7) Copyright Assignment. When a paper is accepted for publication, author(s) will be requested to sign a form assigning copyright to Informath Publishing Group. Failure to do it promptly may delay the publication.

NONLINEAR DYNAMICS AND SYSTEMS THEORY

An International Journal of Research and Surveys

Volume 5 Number 1 2005

CONTENTS

 ISSN 1562-8353 Printed in Ukraine
To receive contents pages by e-mail, visit our Website at: **http://www.e-ndst.kiev.ua**