CONFERENCE ON DIFFERENTIAL EQUATIONS AND THEIR APPLICATIONS, BRNO, AUGUST 25 – 29, 1997

Equadiff 9

Proceedings

edited by

R. P. Agarwal, F. Neuman, J. Vosmanský

EQUADIFF 9 • MASARYK UNIVERSITY BRNO 1998

CONFERENCE ON DIFFERENTIAL EQUATIONS AND THEIR APPLICATIONS, BRNO, AUGUST 25 – 29, 1997

Equadiff 9

Proceedings

edited by

R. P. Agarwal, F. Neuman, J. Vosmanský

EQUADIFF 9 • MASARYK UNIVERSITY BRNO 1998

Preface

The tradition of the Czechoslovak EQUADIFF conferences dates back to 1962 when EQUADIFF 1 was organized in Prague. In addition to 160 local participants it was attended by 75 foreign mathematicians. Subsequent conferences held in Bratislava (1966, 1981, 1993), Brno (1972, 1985) and Prague (1977, 1989) turned EQUADIFF into the world's oldest series of comprehensive conferences on differential equations. EQUADIFF is at present one of the most important and biggest conferences of this kind.

The Conference on Differential Equations and Their Applications EQUADIFF 9 was held in Brno, August 25–29, 1997. It was organized by the Masaryk University, Brno in cooperation with Mathematical Institute of the Academy of Sciences, Technical University Brno, Union of Czech Mathematicians and Physicists, Union of Slovak Mathematicians and Physicists and other Czech scientific institutions with support of the International Mathematical Union. EQUADIFF 9 was attended by 269 participants from 32 countries and more than 50 accompanying persons and other guests.

EQUADIFF 9 was prepared by the Organizing Committee presided by F. Neuman, chairman, J. Vosmanský, executive secretary, Z. Došlá, head of the Equadiff Office, and other members J. Diblík, O. Došlý, J. Franců, J. Kalas, J. Kuben, Z. Pospíšil and J. Šimša.

The scientific program was prepared by the Scientific Committee. It consisted of the following Czech and Slovak mathematicians: P. Drábek, J. Haslinger, J. Jaroš, J. Kačur, J. Milota, J. Nečas, F. Neuman (chairman) and † M. Zlámal. The invited speakers have been proposed by the international Honorary and Advisory Board, consisted of R. P. Agarwal, I. Babuška, V. E. Barbu, P. Brunovský, W. N. Everitt, A. Friedman, J. K. Hale, W. Jäger, I. T. Kiguradze, K. Kirschgässner, J. Kurzweil (honorary chairman), V. Lakshmikantham, I. Marek, J. Mawhin, J. Nečas, M. Ráb, K. Rektorys, M. Švec, R. Temam, W. L. Wendland and † M. Zlámal.

The scientific program comprised 8 plenary lectures and 34 main lectures in the following sections:

1. Ordinary differential equations,
2. Partial differential equations,
3. Numerical methods and applications.

In addition 208 papers were presented

a) as communications in simultaneous subsections (112),
b) at the poster session (31),
c) in the form of enlarged abstracts (65).

Besides the scientific program the participants could enjoy a rich social program (e.g. Welcome Party at the Brno castle Špilberk, Glass of wine and Concert, Audience with Mayor of Brno, Trips and Farewell Party in wine cellar Queen Eliška).

This volume contains 12 survey papers mainly by the plenary speakers.

Together with this Proceedings the following EQUADIFF 9 publications have been prepared:

- EQUADIFF 9 issue of Archivum mathematicum (Tomus 34, 1998, No. 1, 232 pp.) containing 20 papers by invited speakers (published in May 1998),
- CD ROM containing, in electronic form, an Equadiff 9 issue of Archivum mathematicum, the Proceedings and 31 other papers submitted by the participants of the conference as well as other conference material (e.g. Abstracts, List of participants, and Program).

Brno, May 1998 Editors

Table of Contents

EQUADIFF 9 CD ROM, Brno 1997
Masaryk University

PROCEEDINGS
pp. 1–23

Compactness Condition for Boundary Value Problems

Ravi P. Agarwal

Department of Mathematics
National University of Singapore
10 Kent Ridge Crescent, Singapore 119260
Email: MATRAVIP@leonis.nus.sg

Dedicated to Professor Lloyd K. Jackson

Abstract. In existence and uniqueness theory of boundary value problems for ordinary differential equations *Compactness Condition* plays an important role. It has been a long standing problem whether other conditions imposed on the differential equations imply this compactness condition. In this lecture we shall survey known results on this problem, including its complete unpublished proof essentially due to L. Jackson and K. Schrader. We shall also discuss some related problems.

AMS Subject Classification. 34B15, 34C10

Keywords. Boundary value problem, Kamke convergence theorem, Banach indicatrix theorem, Helly selection theorem, total variation

1 Introduction

In this lecture we shall consider the following n (≥ 2)th order nonlinear differential equation

$$y^{(n)} = f(x, y, y', \ldots, y^{(q)}), \quad 0 \leq q \leq n-1, \quad \text{but fixed.} \tag{1.1}$$

With respect to (1.1) we shall assume that

(A) $f(x, u_0, u_1, \ldots, u_q) : (a, b) \times \mathbb{R}^{q+1} \to \mathbb{R}$ is continuous.

(B) Solutions of initial value problems for (1.1) are unique.

(C) Solutions of (1.1) extend to (a, b).

(D_n) For any $a < a_1 < a_2 < \cdots < a_n < b$ and any solutions $y(x)$ and $z(x)$ of (1.1), it follows that $y(a_i) = z(a_i)$, $1 \leq i \leq n$ implies $y(x) \equiv z(x)$, i.e., the differential equation (1.1) is n-point *disconjugate* on (a, b).

In the study of boundary value problems for the differential equation (1.1), one of the Propositions which has attracted several Mathematicians and has led

to substantially new mathematics is *whether conditions* (A) – (D_n) *imply the following* ***compactness condition:***

(E) If $[c,d]$ is a compact subinterval of (a,b) and $\{y_m(x)\}$ is a sequence of solutions of (1.1) which is uniformly bounded, i.e., $|y_m(x)| \le M$ on $[c,d]$ for some $M > 0$ and all $m = 1, 2, \ldots$, then there is a subsequence $\{y_{m(j)}(x)\}$ such that $\{y^{(i)}_{m(j)}(x)\}$ converges uniformly on $[c,d]$ for each $0 \le i \le n-1$.

In this lecture we shall survey most of the known results on this Proposition, and touch on some related topics.

2 Preliminary Results

We shall need the following version of Kamke's convergence theorem.

Theorem 2.1. ([5, p. 14]) *Assume that for the differential equation* (1.1) *the conditions* (A) *and* (C) *are satisfied. Then, if* $\{y_m(x)\}$ *is a sequence of solutions of* (1.1) *such that there exists a sequence* $\{x_m\} \subset (a,b)$ *with* $\lim_{m\to\infty} x_m = x_0 \in (a,b)$, $\lim_{m\to\infty} y^{(i)}_m(x_m) = y_i$, $0 \le i \le n-1$. *Then, there is a solution* $y(x)$ *of the differential equation* (1.1) *satisfying the initial conditions* $y^{(i)}(x_0) = y_i$, $0 \le i \le n-1$, *and a subsequence* $\{y_{m(j)}(x)\}$ *of* $\{y_m(x)\}$ *such that* $\lim_{j\to\infty} y^{(i)}_{m(j)}(x) = y^{(i)}(x)$, $0 \le i \le n-1$, *uniformly on each compact subinterval of* (a,b).

Lemma 2.2. *Let* $y(x) \in C^{(n)}[a_1, a_r]$, *satisfying*

$$\begin{aligned} &y(a_i) = y'(a_i) = \cdots = y^{(k_i)}(a_i) = 0, \quad 1 \le i \le r \ (\ge 2) \\ &a < a_1 < a_2 < \cdots < a_r < b, \quad k_i \ge 0, \quad \sum_{i=1}^{r} k_i + r = n. \end{aligned} \tag{2.1}$$

Then, there exist constants $C_{n,k}$, $0 \le k \le n-1$, *such that*

$$|y^{(k)}(x)| \le C_{n,k}(a_r - a_1)^{n-k} \max_{a_1 \le x \le a_r} |y^{(n)}(x)|. \tag{2.2}$$

The problem of finding the best possible constants $C_{n,k}$ in (2.2) is one of the most outstanding problems in polynomial interpolation theory [1,2].

Inequalities (2.2) will be used now to prove local existence of solutions of the differential equation (1.1) satisfying the *r-point conjugate boundary conditions*

$$y(a_i) = A_{1,i}, \ y'(a_i) = A_{2,i}, \ldots, y^{(k_i)}(a_i) = A_{k_i+1,i}, \quad 1 \le i \le r. \tag{2.3}$$

Theorem 2.3 ([1]). *Assume that for the differential equation* (1.1) *the condition* (A) *is satisfied. Further, assume that*

(i) $K_i > 0,\ 0 \le i \le q$ *are given real numbers and let* Q *be the maximum of* $|f(x, u_0, u_1, \ldots, u_q)|$ *on the compact set* $[a_1, a_r] \times D_0$, *where*

$$D_0 = \{(u_0, u_1, \ldots, u_q)\ :\ |u_i| \le 2K_i,\ 0 \le i \le q\},$$

(ii) $\max\limits_{a_1 \le x \le a_r} |p^{(i)}(x)| \le K_i,\ 0 \le i \le q$, *where* $p(x)$ *is the Hermite interpolating polynomial*

$$p(x) = \sum_{i=1}^{r} \sum_{j=0}^{k_i} \sum_{\ell=0}^{k_i - j} \frac{1}{j!\ell!} \left[\frac{(x-a_i)^{k_i+1}}{\Omega(x)} \right]_{x=a_i}^{(\ell)} \frac{\Omega(x)}{(x-a_i)^{k_i+1-j-\ell}} A_{j+1,i}$$

and

$$\Omega(x) = \prod_{i=1}^{r} (x - a_i)^{k_i+1},$$

(iii) $(a_r - a_1) \le \left(\dfrac{K_i}{QC_{n,i}} \right)^{1/(n-i)},\ 0 \le i \le q.$

Then, the boundary value problem (1.1), (2.3) *has a solution in* D_0.

Proof. The set

$$B[a_1, a_r] = \left\{ y(x) \in C^{(q)}[a_1, a_r]\ :\ \|y^{(i)}\| \le 2K_i,\ 0 \le i \le q \right\},$$

where

$$\|y^{(i)}\| = \max_{a_1 \le x \le a_r} |y^{(i)}(x)|$$

is a closed convex subset of the Banach space $C^{(q)}[a_1, a_r]$. Consider an operator $T : C^{(q)}[a_1, a_r] \to C^{(n)}[a_1, a_r]$ as follows

$$(Ty)(x) = p(x) + \int_{a_1}^{a_r} g(x,t) f(t, y(t), y'(t), \ldots, y^{(q)}(t)) dt, \tag{2.4}$$

where $g(x,t)$ is the Green's function of the boundary value problem $y^{(n)} = 0$, (2.1). Obviously, any fixed point of (2.4) is a solution of (1.1), (2.3).

We note that $(Ty)(x) - p(x)$ satisfies the conditions of Lemma 2.2, and

$$(Ty)^{(n)}(x) - p^{(n)}(x) = (Ty)^{(n)}(x) = f(x, y(x), y'(x), \ldots, y^{(q)}(x)).$$

Thus, for all $y(x) \in B[a_1, a_r]$, $\|(Ty)^{(n)}\| \le Q$, and

$$\|(Ty)^{(i)} - p^{(i)}\| \le C_{n,i} Q (a_r - a_1)^{n-i}, \quad 0 \le i \le q$$

which also implies that

$$\|(Ty)^{(i)}\| \le \|p^{(i)}\| + C_{n,i} Q (a_r - a_1)^{n-i} \le K_i + K_i = 2K_i, \quad 0 \le i \le q. \tag{2.5}$$

Thus, the operator T maps $B[a_1, a_r]$ into itself. Further, the inequalities (2.5) imply that the sets $\{(Ty)^{(i)}(x) \ : \ y(x) \in B[a_1, a_r]\}$, $0 \le i \le q$ are uniformly bounded and equicontinuous on $[a_1, a_r]$. Hence, $\overline{TB}[a_1, a_r]$ is compact follows from the Ascoli-Arzela theorem. The Schauder fixed point theorem is applicable and a fixed point of (2.4) in D_0 exits.

Corollary 2.4. *Assume that for the differential equation (1.1) the condition (A) is satisfied. Further, assume that there exist constants $N_i \ge 0$, $0 \le i \le q$ such that $\max_{a_1 \le x \le a_r} |p^{(i)}(x)| \le N_i$, $0 \le i \le q$. Then, there exists a $\delta = \delta(N_0, N_1, \ldots, N_q) > 0$ such that if $a_r - a_1 \le \delta$, the boundary value problem (1.1), (2.3) has a solution $y(x)$. Furthermore,*

$$|y^{(i)}(x)| \le N_i + 1, \ 0 \le i \le q \qquad on \quad [a_1, a_r].$$

Theorem 2.5. *Assume that for the differential equation (1.1) the condition (A) is satisfied. Further, assume that the conditions (i) and (iii) of Theorem 2.3 are satisfied. Then, for any $g(x) \in C^{(n-1)}[a_1, a_r]$ the differential equation (1.1) together with*

$$y^{(j)}(a_i) = g^{(j)}(a_i), \qquad 0 \le j \le k_i, \quad 1 \le i \le r \tag{2.6}$$

has a solution, if

$$\sum_{j=i}^{n-1} M_j (a_r - a_1)^{j-i} \le K_i, \qquad 0 \le i \le q$$

where

$$M_j = \max_{a_1 \le x \le a_r} |g^{(j)}(x)|, \qquad 0 \le j \le n-1.$$

Proof. We need to verify that the condition (ii) of Theorem 2.3 is satisfied. For this, in $p(x)$ we take $A_{j+1,i} = g^{(j)}(a_i)$, $0 \le j \le k_i$, $1 \le i \le r$. Then, the function $\phi(x) = g(x) - p(x)$ has n zeros in $[a_1, a_r]$. Thus, from the generalized Rolle's theorem $\phi^{(k)}(x)$, $1 \le k \le n-1$ vanishes at least $n-k$ times in (a_1, a_r). Let $x_k \in (a_1, a_r)$ be any zero of $\phi^{(k)}(x)$, then

$$\left|p^{(n-1)}(x)\right| = \left|p^{(n-1)}(x_{n-1})\right| = \left|g^{(n-1)}(x_{n-1})\right| \le \max_{a_1 \le x \le a_r} \left|g^{(n-1)}(x)\right| = M_{n-1}$$

and

$$\begin{aligned}
\left|p^{(n-2)}(x)\right| &\le \left|p^{(n-2)}(x_{n-2})\right| + \left|\int_{x_{n-2}}^{x} \left|p^{(n-1)}(t)\right| dt\right| \\
&= \left|g^{(n-2)}(x_{n-2})\right| + \left|\int_{x_{n-2}}^{x} \left|g^{(n-1)}(x_{n-1})\right| dt\right| \\
&\le M_{n-2} + M_{n-1}(a_r - a_1).
\end{aligned}$$

Using the same argument repeatedly, we obtain

$$\left|p^{(i)}(x)\right| \leq \sum_{j=i}^{n-1} M_j (a_r - a_1)^{j-i}.$$

Corollary 2.6. *Assume that for the differential equation (1.1) the condition (A) is satisfied. Then, for any $g(x) \in C^{(n-1)}[a_1, a_r]$ there exist constants $\delta > 0$, $N_i \geq 0,\ 0 \leq i \leq q$, all depending on $g(x)$ such that the boundary value problem (1.1), (2.6) has a solution $y(x)$, provided $a_r - a_1 \leq \delta$. Furthermore, $|y^{(i)}(x)| \leq N_i + 1,\ 0 \leq i \leq q$ on $[a_1, a_r]$.*

Corollary 2.7. *Assume that for the differential equation (1.1) the condition (A) is satisfied. Further, assume that there exist constants $N_i \geq 0,\ 0 \leq i \leq q$ such that $\max_{a_1 \leq x \leq a_r} |p^{(i)}(x)| \leq N_i,\ 0 \leq i \leq q$. Then, there exist a $\delta = \delta(N_0, N_1, \ldots, N_q) > 0$, and an $\varepsilon = \varepsilon(a_1, \ldots, a_r)$ such that for $a_r - a_1 \leq \delta$, the boundary value problem (1.1),*

$$y(a_i) = A_{1,i} + \varepsilon_{1,i},\ y'(a_i) = A_{2,i} + \varepsilon_{2,i}, \ldots, y^{(k_i)}(a_i) = A_{k_i+1,i} + \varepsilon_{k_i+1,i},\ 1 \leq i \leq r$$

has a solution $y_\varepsilon(x)$, provided $|\varepsilon_{j,i}| \leq \varepsilon,\ 0 \leq j \leq k_i,\ 1 \leq i \leq r$. Furthermore, $|y_\varepsilon^{(i)}(x)| \leq N_i + 1,\ 0 \leq i \leq q$ on $[a_1, a_r]$.

3 The Case $n = 2$

For the second order differential equation (1.1) only conditions (A) and (C) imply (E). We shall prove this in the following:

Theorem 3.1. *If the differential equation (1.1) is of second order and satisfies conditions (A) and (C), then (1.1) also satisfies condition (E).*

Proof. If $\{y_m(x)\}$ is a sequence of solutions of (1.1) with $|y_m(x)| \leq M$ on $[c, d] \subset (a, b)$ for some $M > 0$, and each $m \geq 1$, then for each m there is a $x_m \in (c, d)$ such that

$$|y'_m(x_m)| = \frac{|y_m(d) - y_m(c)|}{d - c} \leq \frac{2M}{d - c}.$$

Consequently, $\{x_m\},\ \{y_m(x_m)\}$ and $\{y'_m(x_m)\}$ are bounded sequences. By taking subsequences in succession which converge, we conclude that there exist values $x_0,\ y_0,\ y'_0$ such that $x_{m(1)} \to x_0,\ y_{m(1)}(x_{m(1)}) \to y_0\ y'_{m(1)}(x_{m(1)}) \to y'_0$, where $\{m(1)\}$ is some subsequence of $\{m\}$. Thus, by Theorem 2.1 there is a subsequence $\{y_{m(2)}(x)\}$ of $\{y_{m(1)}(x)\}$ and a solution $y(x)$ of (1.1) satisfying $y(x_0) = y_0,\ y'(x_0) = y'_0$ such that $\lim_{m\to\infty} y^{(i)}_{m(2)}(x) = y^{(i)}(x),\ i = 0, 1$, uniformly on $[c, d]$.

4 Case $n = 3$

If the differential equation (1.1) is of third order, then conditions (A) and (C) are not enough for (E). In fact, the equation $y''' = -[y']^3$ satisfies (A) and (C) on $\mathbb{R}^4$, but the sequence $\{y_m(x)\}$ of solutions of the initial value problem

$$y''' = -[y']^3, \quad y(0) = y'(0) = 0, \quad y''(0) = m$$

for $m = 1, 2, \ldots$, is uniformly bounded on $\mathbb{R}$ and does not contain a subsequence satisfying (E) on any compact subinterval of $\mathbb{R}$. Here, we shall show that conditions (A), (C) and (D_3) do imply (E). However, for this an immediate appeal to Theorem 2.1 is not possible and we shall need the following lemmas.

Lemma 4.1. *Assume that the differential equation (1.1) is of third order and satisfies the condition (A). Then, given any compact subinterval $[c, d] \subset (a, b)$ and any fixed $M > 0$, there is a $\delta(M) > 0$ such that for any $[a_1, a_2] \subset [c, d]$ with $a_2 - a_1 \le \delta(M)$, and any real α with $|\alpha| \le M$, (1.1) has solutions satisfying each of the boundary conditions*

$$y(a_1) = y(a_2) = \alpha, \ y'(a_1) = 0 \quad \text{and} \quad y(a_1) = y(a_2) = \alpha, \ y'(a_2) = 0.$$

Furthermore, for any such solution $|y'(x)| \le 1$ and $|y''(x)| \le 1$ on $[a_1, a_2]$.

Proof. This is a particular case of Corollary 2.4.

Lemma 4.2 ([11]). *Assume that the differential equation (1.1) is of third order and satisfies the condition (A). Let $\phi(x)$, $\psi(x)$ be of class $C^{(2)}$ on $[a_1 - \tau, a_1 + \tau] \subset (a, b)$ with $\phi(a_1) = \psi(a_1)$, $\phi'(a_1) = \psi'(a_1)$ and $\phi''(a_1) < \psi''(a_1)$. Then, there is a δ, $0 < \delta \le \tau$, such that all solutions $y(x)$ of (1.1) with the initial conditions $y(a_1) = y_0 = \phi(a_1)$, $y'(a_1) = y_1 = \phi'(a_1)$, and $y''(a_1) = y_2 = \frac{1}{2}[\phi''(a_1) + \psi''(a_1)]$ exist on $[a_1 - \delta, a_1 + \delta]$ and satisfy $\phi(x) < y(x) < \psi(x)$ for $0 < |x - a_1| \le \delta$.*

Proof. Let $8\rho = \psi''(a_1) - \phi''(a_1)$ and choose δ_0, $0 < \delta_0 \le \tau$ such that $|\phi''(x) - \phi''(a_1)| \le \rho$ and $|\psi''(x) - \psi''(a_1)| \le \rho$ for $|x - a_1| \le \delta_0$. Let $M > 0$ be a bound for $f(x, y, y', y'')$ on the compact set

$$\{(x, y, y', y'') \ : \ |x - a_1| \ \le \ \delta_0, \ |y - y_0| \le 1, \ |y' - y_1| \le 1, \ |y'' - y_2| \le 1\}.$$

Then, it follows from the relations

$$y(x) = y_0 + \int_{a_1}^{x} y'(t)dt, \qquad y'(x) = y_1 + \int_{a_1}^{x} y''(t)dt$$

$$y''(x) = y_2 + \int_{a_1}^{x} f(t, y(t), y'(t), y''(t))dt$$

that all solutions of the stated initial value problem exist on the closed interval $[a_1 - \delta_1, a_1 + \delta_1]$, where

$$\delta_1 = \min\left\{\delta_0, \frac{1}{M}, \frac{1}{1+|y_2|}, \frac{1}{1+|y_1|}\right\}.$$

Thus, if $\delta = \min\{\delta_1, \rho/M\}$, it follows that for all solutions $y(x)$ of the initial value problem $|y''(x) - y_2| \le \rho$ for $|x - a_1| \le \delta$. Hence, for all solutions $y(x)$ of the initial value problem

$$y''(x) - \phi''(x) \ge 2\rho \quad \text{and} \quad \psi''(x) - y''(x) \ge 2\rho$$

on $[a_1 - \delta, a_1 + \delta]$.

Lemma 4.3 ([11]). *Assume that the differential equation (1.1) is of third order and satisfies the conditions (A), (C) and (D$_3$). Then, solutions of two point boundary value problems for (1.1) are unique, i.e., the following condition (D$_2$) holds:*

(D$_2$) *If $a < a_1 < a_2 < b$, and $y(x)$, $z(x)$ are both solutions of (1.1) satisfying $y(a_1) = z(a_1)$, $y'(a_1) = z'(a_1)$, $y(a_2) = z(a_2)$, or $y(a_1) = z(a_1)$, $y(a_2) = z(a_2)$, $y'(a_2) = z'(a_2)$, then it follows that $y(x) \equiv z(x)$ on $[a_1, a_2]$.*

Proof. We shall consider only the case where $y(a_1) = z(a_1)$, $y'(a_1) = z'(a_1)$, $y(a_2) = z(a_2)$. We first assume that $y''(a_1) \ne z''(a_1)$, and to be specific assume that $y''(a_1) > z''(a_1)$. Then, by Lemma 4.2 there is a $\delta > 0$ with $a < a_1 - \delta < a_1 + \delta < a_2$ such that all solutions $w(x)$ of the initial value problem for (1.1) with the initial conditions

$$w(a_1) = y_0 = z(a_1), \quad w'(a_1) = y_1 = z'(a_1), \quad w''(a_1) = y_2 = \frac{1}{2}\left[y''(a_1) + z''(a_1)\right] \tag{4.1}$$

satisfy $z(x) < w(x) < y(x)$ for $0 < |x - a_1| \le \delta$. Let $\{\varepsilon_m\}$ be a monotone decreasing sequence of positive numbers converging to zero, and let $z_m(x)$ be a solution of (1.1) with the initial conditions

$$z_m(a_1) = y_0, \quad z_m'(a_1) = y_1 + \varepsilon_m, \quad z_m''(a_1) = y_2. \tag{4.2}$$

Then, $\{z_m(x)\}$ contains a subsequence converging uniformly on $[a_1 - \delta, a_1 + \delta]$ to a solution of (1.1), (4.1). Hence, for sufficiently large m, there is a solution $z_m(x)$ of (1.1), (4.2) such that

$$z(a_1 - \delta) < z_m(a_1 - \delta) < y(a_1 - \delta) \quad \text{and} \quad z(a_1 + \delta) < z_m(a_1 + \delta) < y(a_1 + \delta).$$

Since $z_m(a_1) = y(a_1) = z(a_1)$ and $z_m'(a_1) = y_1 + \varepsilon_m > y'(a_1) = z'(a_1)$, it follows that there are x_1, x_2 with $a_1 - \delta < x_1 < a_1 < x_2 < a_1 + \delta$ such that $z_m(x_1) = z(x_1)$, and $z_m(x_2) = y(x_2)$. Since $y(a_1) = z(a_1)$ at $a_2 > a_1 + \delta$, it follows that any extension of $z_m(x)$ intersects either $y(x)$ or $z(x)$ again on

$[a_1+\delta, b)$. Since $z_m(x) \not\equiv z(x)$ on $[x_1, a_1]$ and $z_m(x) \not\equiv y(x)$ on $[a_1, x_2]$, this contradicts condition (D_3).

Thus, if $y(x) \not\equiv z(x)$ on $[a_1, a_2]$, then $y^{(i)}(a_1) = z^{(i)}(a_1)$ for $i = 0, 1, 2$. However, if $y^{(i)}(a_1) = z^{(i)}(a_1)$ for $i = 0, 1, 2$ and $y(x) \not\equiv z(x)$ on $[a_1, a_2]$, then for $a < x_3 < a_1$, $u(x)$ and $v(x)$ defined by

$$u(x) = v(x) = y(x) \quad \text{on} \quad [x_3, a_1], \quad u(x) = y(x) \quad \text{on} \quad (a_1, a_2],$$
$$\text{and} \quad v(x) = z(x) \quad \text{on} \quad (a_1, a_2]$$

will be solutions on $[x_3, a_2]$ which again contradicts condition (D_3). Thus, we conclude that $y(x) \equiv z(x)$ on $[a_1, a_2]$.

Lemma 4.4 ([12]). *Let $y(x) \in C^{(2)}[\alpha, \beta]$ and assume that $|y(x)| \le M$ on $[\alpha, \beta]$. There is a $K > 0$ depending on M and $\beta - \alpha$ such that if, $\max\{|y'(x)|, |y''(x)|\} > K$ for all $\alpha \le x \le \beta$, then $y'(x_0) = 0$ for some x_0 with $\alpha < x_0 < \beta$.*

Proof. Assume that the conclusion is false. We shall determine $N > 0$ so that the following inequality holds

$$|y'(x)| + |y''(x)| \ge N + \frac{2M}{\beta - \alpha} + 1 \quad \text{on} \quad [\alpha, \beta]. \tag{4.3}$$

For this, by the Mean value theorem there exists a $x_1 \in (\alpha, \beta)$ such that

$$|y'(x_1)| = \left| \frac{y(\beta) - y(\alpha)}{\beta - \alpha} \right| \le \frac{2M}{\beta - \alpha}.$$

There are now two possible cases, however, since both are similar, we shall consider only the case

$$0 < y'(x_1) \le \frac{2M}{\beta - \alpha} \quad \text{and} \quad \alpha < x_1 \le \frac{\alpha + \beta}{2}.$$

If $y(x_1) = M$, then $y'(x_1) = 0$ and the proof is finished. So, we assume that $y(x_1) \ne M$. (It is clear that we are assuming $y'(x_1) \ne 0$.) We define $\eta = (\beta - \alpha)/8$. Now to complete the proof we need to consider the following two subcases:

Case (i). Assume that $y''(x_1) \le 0$. Then, in order (4.3) holds, it is necessary that $y''(x_1) \le -N$. Thus, $y'(x)$ is decreasing on a right neighborhood of $x = x_1$. In fact, if $0 \le y'(x) \le (2M/(\beta - \alpha))$, then it will follow that $y''(x) \le -N$ on $[x_1, \beta]$. However, then by Taylor's formula, we have

$$y(\beta) = y(x_1) + (x_1 - \beta)y'(x_1) + \frac{(x_1 - \beta)^2}{2} y''(\xi), \qquad \xi \in (x_1, \beta)$$

$$< M + \frac{2M(\beta - \alpha)}{\beta - \alpha} - \frac{(\beta - \alpha)^2}{4} N$$

$$\le -M$$

provided

$$N \geq \frac{32M}{(\beta-\alpha)^2} = \frac{M}{2\eta^2}.$$

But, this implies that $|y(\beta)| > M$, which is a contradiction. Thus, there exists a point $x_1 \leq x_0 < \beta$ such that $y'(x_0) = 0$.

Case (ii). Assume that $y''(x_1) > 0$, (we shall work across $[\alpha, \beta]$ on subintervals of length η.) Then, from (4.3) it follows that $y''(x_1) > N$. We assume that $y''(x) \geq N/2$ on $[x_1, x_1 + \eta]$. Again, by Taylor's formula, we have

$$y(x_1+\eta) = y(x_1) + \eta y'(x_1) + \frac{1}{2}\eta^2 y''(\xi), \quad \xi \in (x_1, x_1+\eta).$$

As in Case (i), we find $y(x_1 + \eta) > -M + N\eta^2/4 \geq M$, provided $N \geq (8M/\eta^2)$, which is a contradiction.

From this contradiction, we conclude that there exists a $x_1 < x_2 < x_1 + \eta$ such that $y''(x_2) = N/2$. Since, $y''(x)$ is positive up to x_2, we can assume that x_2 is the first point such that $y''(x_2) = N/2$. Thus,

$$y'(x_2) > \frac{1}{2}N + \frac{2M}{\beta-\alpha} \quad \text{on} \quad [x_1, x_2).$$

Now assume that

$$y'(x) \geq \frac{1}{2}N + \frac{2M}{\beta-\alpha} \quad \text{on} \quad [x_2, x_2+\eta).$$

Then, it follows that

$$\begin{aligned} y(x_2+\eta) &= y(x_2) + \eta y'(\xi), \quad \xi \in (x_2, x_2+\eta) \\ &> -M + \frac{1}{2}\eta N + \eta\frac{2M}{\beta-\alpha} \\ &= -M + \frac{1}{2}\eta N + \frac{1}{4}M \\ &\geq M, \end{aligned}$$

provided $N \geq (7M/2\eta)$. But, this implies $y(x_2 + \eta) > M$, which is a contradiction.

From this contradiction, there exists a $x_2 < x_3 < x_2 + \eta$ such that

$$y'(x_3) = \frac{1}{2}N + \frac{2M}{\beta-\alpha},$$

and we take x_3 to be the first such point, i.e.,

$$y'(x) > \frac{1}{2}N + \frac{2M}{\beta-\alpha}, \quad \text{on} \quad [x_2, x_3).$$

So, $y'(x)$ is decreasing on $[x_2, x_3)$. Thus, $y''(x_3) < -N/2$. This implies that in a right neighborhood of x_3, $y''(x) < -N/2$ and $y'(x)$ is decreasing. Assume that

$$0 < y'(x) \le \frac{1}{2}N + \frac{2M}{\beta - \alpha} \quad \text{on} \quad [x_3, \beta).$$

Then, by Taylor's formula

$$y(x_3) = y(\beta) + (x_3 - \beta)y'(\beta) + \frac{(x_3 - \beta)^2}{2}y''(\xi), \qquad \xi \in (x_3, \beta).$$

Since in the above relation the second term in the right side is nonpositive, in view of $(\beta - x_3) \ge (\beta - \alpha)/4$, it follows that

$$y(x_3) < M - \frac{1}{4}(\beta - \alpha)^2 N \le -M$$

provided $N \ge M/\eta^2$. But, this leads to the contradiction that $|y(x_3)| > M$.

From this construction we conclude that

$$0 < y'(x) \le \frac{1}{2}N + \frac{2M}{\beta - \alpha} \quad \text{on} \quad [x_3, \beta)$$

is false. Thus, in conclusion $y'(x_0) = 0$, for some $x_3 < x_0 < \beta$.

Theorem 4.5 ([12]). *If the differential equation (1.1) is of third order and satisfies conditions (A), (C) and (D$_3$), then (1.1) also satisfies condition (E).*

Proof. Suppose that the result is not true. Then, since $|y_m(x)| \le M$ on $[c, d]$ for $n \ge 1$, it follows from Theorem 2.1 that $|y'_m(x)| + |y''_m(x)| \to \infty$ uniformly on $[c, d]$. Let $c \le a_1 < a_2 < a_3 < a_4 \le d$ be such that $a_4 - a_1 \le \delta(M)$, where $\delta(M)$ is as defined in Lemma 4.1. By Lemma 4.4 there is a $K > 0$ such that, if $\max\{|y'_m(x)|, |y''_m(x)|\} > K$ for each $x \in [c, d]$, then $y'_m(x)$ has a zero on (a_1, a_2), on (a_2, a_3) and on (a_3, a_4). Furthermore, we can assume that $K > 1$. Now from the fact that $|y'_m(x)| + |y''_m(x)| \to \infty$ uniformly on $[c, d]$ we can conclude that there is a positive integer m_0 such that $\max\{|y'_{m_0}(x)|, |y''_{m_0}(x)|\} > K$ on $[c, d]$. Let $a_1 < x_1 < a_2 < x_2 < a_3 < x_3 < a_4$ be such that $y'_{m_0}(x_i) = 0$ for $i = 1, 2, 3$. Then, $|y''_{m_0}(x_i)| > K > 1$ for $i = 1, 2, 3$. Now we need to consider the following two cases:

If $y_{m_0}(x_i) = y_{m_0}(x_j)$ with $x_i < x_j$, then $y_{m_0}(x)$ is the solution of the differential equation (1.1) together with the two-point boundary conditions $y(x_i) = y(x_j) = y_{m_0}(x_i)$, $y'(x_i) = 0$. However, since $x_j - x_i < \delta(M)$, it follows from Lemma 4.1 that $|y'_{m_0}(x)| \le 1$ and $|y''_{m_0}(x)| \le 1$ on $[x_i, x_j]$, which is a contradiction to $|y''_{m_0}(x_i)| > K > 1$.

If $y_{m_0}(x_i) \ne y_{m_0}(x_j)$ for $x_i \ne x_j$, then it suffices to assume that $y_{m_0}(x_1) < y_{m_0}(x_2) < y_{m_0}(x_3)$. In fact, the same argument applies to the other orderings of the values of $y_{m_0}(x_i)$, $i = 1, 2, 3$. If $y''_{m_0}(x_2) > K$, there is a t_1, $x_1 < t_1 < x_2$, such that $y_{m_0}(t_1) = y_{m_0}(x_2)$. If $y''_{m_0}(x_2) < -K$, there is a t_2, $x_2 < t_2 < x_3$, such that $y_{m_0}(t_2) = y_{m_0}(x_2)$. In either case Lemma 4.1 is again applied to obtain a contradiction.

Hence, the sequence $\{y_m(x)\}$ contains a subsequence converging uniformly on $[c, d]$ along with its first and second order derivative sequences.

5 Weak Compactness Condition

It seems very difficult, if not impossible, to extend the method of Theorem 4.5 to equations of higher orders. Here, we shall show that for equation (1.1) of arbitrary order n, conditions (A), (C) and (D_n) do imply a weaker type of compactness condition for the solutions of (1.1). For this we shall need the following:

Theorem 5.1. (Banach Indicatrix Theorem, [9, p. 271]) *If $h \in C[c,d] \cap BV[c,d]$, then*

$$V_c^d(h) = \int_{-\infty}^{\infty} N_h(\alpha) d\alpha,$$

where

$$N_h(\alpha) = \begin{cases} Card\{x \in [c,d] \ : \ h(x) = \alpha\}, & \text{if this set is finite,} \\ \\ +\infty, & \text{if the above set is infinite,} \end{cases}$$

and where the above integral is in the Lebesgue sense.

Theorem 5.2 ([16]). *Assume that the differential equation (1.1) satisfies the conditions (A), (C), and (D_n). Further, assume that $[c,d]$ is a compact subinterval of (a,b) and $\{y_m(x)\}$ is a sequence of solutions of (1.1) which is uniformly bounded on $[c,d]$. Then, the sequence $\left\{V_c^d(y_m)\right\}$ of total variations of the functions $y_m(x)$ on $[c,d]$ is bounded, i.e., there exists an $N > 0$ such that $V_c^d(y_m) \leq N$ for all m.*

Proof. Assume the assertion is false. Then there is a sequence of solutions $\{y_m(x)\}$ of (1.1), a compact interval $[c,d] \subset (a,b)$, and an $M > 0$ such that $|y_m(x)| \leq M$ on $[c,d]$ for all m, but such that $V_c^d(y_m) \to \infty$, as $m \to \infty$.

We claim that $\sum_{i=0}^{n-1} |y_m^{(i)}(x)| \to \infty$ on $[c,d]$ as $m \to \infty$, i.e., given $R > 0$, there exists a $L > 0$ such that $\sum_{i=0}^{n-1} |y_m^{(i)}(x)| > R$, on $[c,d]$ for all $m \geq L$. If the claim is false, then there exists a $\beta > 0$ and a subsequence $\{y_{m(j)}(x)\}$ such that $\sum_{i=0}^{n-1} |y_{m(j)}^{(i)}(x_j)| \leq \beta$, for all $j \geq 1$, and where $\{x_j\} \subset [c,d]$. Now by choosing successive subsequences and relabeling, we obtain points $\{x_p\}$ and solutions $\{y_p(x)\}$ such that $\{x_p\}$ and $\{y_p^{(i)}(x_p)\}$, $0 \leq i \leq n-1$ all converge. Thus, by Theorem 2.1 there exists a further subsequence $\{y_{p(j)}(x)\}$ such that $\{y_{p(j)}^{(i)}(x)\}$ converges uniformly on $[c,d]$, $0 \leq i \leq n-1$. This implies that $\{y'_{p(j)}(x)\}$ is a uniformly bounded sequence on $[c,d]$. Now since each $y'_{p(j)}(x)$ is absolutely continuous, it follows that

$$V_c^d(y_{p(j)}) = \int_c^d |y'_{p(j)}(x)| dx.$$

Hence, the sequence $\{V_c^d(y_{p(j)})\}$ is a bounded sequence, which is a contradiction. Thus, $\sum_{i=0}^{n-1} |y_m^{(i)}(x)| \to \infty$ on $[c,d]$ as $m \to \infty$.

We shall now apply Corollary 2.4 with $N_0 = M$ and $N_0 = 0,\ 1 \le i \le n-1$, so that there exists a $\delta(M) > 0$ such that for any α with $|\alpha| \le M$ and any points $c \le a_1 < \cdots < a_n \le d$ with $a_n - a_1 \le \delta$, and with $w(x) \equiv \alpha$ so that $w^{(i)}(x) \equiv 0,\ 1 \le i \le n-1$, the boundary value problem for (1.1) satisfying $y(a_i) = \alpha,\ 1 \le i \le n$ has a solution $y(x)$ with $|y^{(i)}(x)| \le N_i + 1$ on $[a_1, a_n]$ for $0 \le i \le n-1$. In particular, the boundary value problem has a solution $y(x)$ with $|y(x)| \le M + 1$, and $|y^{(i)}(x)| \le 1,\ 1 \le i \le n-1$ on $[a_1, a_n]$. Further, for such a solution, we have $\sum_{i=0}^{n-1} |y^{(i)}(x)| \le M + n$ on $[a_1, a_n]$.

Now let L_0 be such that $\sum_{i=0}^{n-1} |y_m^{(i)}(x)| > M + n$ on $[c, d]$ for all $m \ge L_0$. It follows that given $m \ge L_0$ and α, with $|\alpha| \le M$ the solution $y_m(x)$ intersects the line $y = \alpha$ at most $n-1$ times in any closed subinterval of $[c, d]$ of length less than δ. For if, there are points $c \le x_1 < \cdots < x_n \le d,\ x_n - x_1 \le \delta$ such that $y_m(x_i) = \alpha,\ 1 \le i \le n$ where $|\alpha| \le M$ and $m \ge L_0$, there is also the above mentioned solution $y(x)$ by Corollary 2.4 satisfying $y(x_i) = \alpha,\ 1 \le i \le n$. By (D_n), $y_m(x) \equiv y(x)$ on $[x_1, x_n]$. But, then $\sum_{i=0}^{n-1} |y^{(i)}(x)| \le M+n$ on $[a_1, a_n]$ and $\sum_{i=0}^{n-1} |y_m^{(i)}(x)| > M+n$ on $[a_1, a_n]$ is not possible. Thus, for every $m \ge L_0,\ y_m(x)$ intersects each line $y = \alpha,\ |\alpha| \le M$ at most $n-1$ times in any closed subinterval of $[c, d]$ of length less than δ. So, for all $m \ge L_0$,

$$N_{y_m}(\alpha) \le (n-1)\left(\left[\frac{d-c}{\delta}\right] + 1\right), \quad \text{if} \quad |\alpha| \le M$$

and $N_{y_m}(\alpha) = 0$, if $|\alpha| > M$. Thus, by the Banach Indicatrix Theorem it follows that for $m \ge L_0$,

$$\begin{aligned} V_c^d(y_m) &= \int_{-M}^{M} N_{y_m}(\alpha)\, d\alpha \\ &\le \int_{-M}^{M} (n-1)\left(\left[\frac{d-c}{\delta}\right] + 1\right) d\alpha \\ &= 2M(n-1)\left(\left[\frac{d-c}{\delta}\right] + 1\right). \end{aligned}$$

But, this contradicts $V_c^d(y_m) \to \infty$, as $m \to \infty$. Hence, $\{V_c^d(y_m)\}$ is a bounded sequence.

Theorem 5.3. (Helly's Selection (or Choice) Theorem, [22, p. 398]) *If $\{y_m(x)\}$ is a sequence of functions on $[c, d]$ such that for some M, $|y_m(x)| \le M$ on $[c, d]$ for all $m \ge 1$, and such that $|V_c^d(y_m)| \le H$, for all $m \ge 1$, and some $H > 0$, then there exists a subsequence $\{y_{m(j)}(x)\}$ which converges point-wise on $[c, d]$. Moreover, the limit function is of bounded variation on $[c, d]$.*

Corollary 5.4. *Assume that the differential equation (1.1) satisfies the conditions (A), (C) and (D_n). Then, if $[c, d]$ is a compact subinterval of (a, b) and if $\{y_m(x)\}$ is a sequence of solutions of (1.1) which is uniformly bounded on*

$[c,d]$, there is a subsequence $\{y_{m(j)}(x)\}$ which converges point-wise on $[c,d]$ and $z(x) = \lim_{j\to\infty} y_{m(j)}(x)$ is of bounded variation on $[c,d]$ (as a consequence $z(x)$ has finite derivative almost everywhere, and also $\int_c^d z(x)dx$ exists).

Proof. The result follows from Theorems 5.2 and 5.3.

Remark 5.1. In conjunction with Corollary 5.4 we remark that Schrader [19,20] has proven that, if $\{y_m(x)\}$ is a uniformly bounded sequence of functions on a compact interval $[c,d]$, and if the functions $y_m(x)$ satisfy only the uniqueness condition (D_n) on $[c,d]$, then there is a subsequence of $\{y_m(x)\}$ which converges point-wise on $[c,d]$.

6 Generalized Solutions

To prove the Proposition now we shall follow another possible approach. For this, if the differential equation (1.1) satisfies (A) – (D_n), then it is straightforward to show that the compactness condition (E) is equivalent to the following:

(E^*) If $\{y_m(x)\}$ is a sequence of solutions of (1.1) which is monotone and bounded on some compact subinterval $[c,d] \subset (a,b)$, then $\lim_{m\to\infty} y_m(x)$ is a solution of (1.1) on $[c,d]$.

Thus, to prove the Proposition it suffices to show that the conditions (A) – (D_n) imply that the limit of a bounded monotone sequence of solutions of (1.1) is also a solution.

Definition 6.1. A function $\phi(x)$ defined on an interval $J \subset (a,b)$ is said to be a *generalized solution* of (1.1) on J if for each set of points $a_1 < a_2 < \cdots < a_n$ contained in J and any solution $y(x)$ of (1.1), the inequalities $(-1)^{n+i}[y(a_i) - \phi(a_i)] < 0$, $1 \le i \le n$ imply that $y(x) < \phi(x)$ on $J \cap [a_n, b)$ and $(-1)^{n+1}[y(x) - \phi(x)] < 0$ on $J \cap (a, a_1]$, and the inequalities $(-1)^{n+i}[y(a_i) - \phi(a_i)] > 0$, $1 \le i \le n$ imply $y(x) > \phi(x)$ on $J \cap [a_n, b)$ and $(-1)^{n+1}[y(x) - \phi(x)] > 0$ on $J \cap (a, a_1]$.

Theorem 6.1 ([13,17]). *Assume that the differential equation (1.1) satisfies conditions (A) – (D_n), and that $\lim_{m\to\infty} y_m(x) = \phi(x)$ on $J \subset (a,b)$, where $\{y_m(x)\}$ is a sequence of solutions of (1.1). Then, $\phi(x)$ is a generalized solution of (1.1) on J.*

Proof. Assume that for $a_1 < a_2 < \cdots < a_n$ contained in J there is a solution $y(x)$ of (1.1) such that $(-1)^{n+i}[y(a_i) - \phi(a_i)] < 0$ for $1 \le i \le n$, but that also $y(a_0) > \phi(a_0)$ for some $a_0 > a_n$ in J. Then, since $\lim_{m\to\infty} y_m(x) = \phi(x)$, there is a solution $y_m(x)$ of (1.1) such that $(-1)^{n+i}[y(a_i) - y_m(a_i)] < 0$ for $1 \le i \le n$ and $y(a_0) > y_m(a_0)$. This contradicts the condition (D_n). The remaining inequalities can be proved in a similar way.

Thus, the limit of a bounded monotone sequence of solutions $\{y_m(x)\}$ of (1.1) satisfying (A) – (D_n) is a generalized solution.

Lemma 6.2. *Assume that the differential equation (1.1) satisfies condition (A) and that $\phi(x) \in C^{(n-1)}[c,d]$, where $[c,d]$ is a compact subinterval of (a,b). Assume that $M > 0$ is such that $|\phi^{(j)}(x)| \le M$ on $[c,d]$ for $0 \le j \le n-1$. Then, there exists a $\delta > 0$ such that, for any $c \le a_1 < a_2 < \cdots < a_n \le d$ with $a_n - a_1 \le \delta$, (1.1) has a solution $y(x)$ with $y(a_i) = \phi(a_i)$, $1 \le i \le n$ and $|y^{(j)}(x)| \le 2M$ on $[a_1, a_n]$ for $0 \le j \le n-1$. Furthermore, δ can be chosen in such a way that, for each fixed set $a_1 < a_2 < \cdots < a_n$ satisfying the above conditions, there is an $\varepsilon > 0$ such that for any y_i, $1 \le i \le n$ with $|y_i - \phi(a_i)| < \varepsilon$, $1 \le i \le n$, (1.1) has a solution $y(x)$ satisfying $y(a_i) = y_i$, $1 \le i \le n$, and $|y^{(j)}(x)| \le 3M$ on $[a_1, a_n]$ for $0 \le j \le n-1$.*

Proof. The proof follows from Corollary 2.7.

Theorem 6.3 ([13,17]). *Assume that the differential equation (1.1) satisfies conditions (A) and (D_n), and that $\lim_{m\to\infty} y_m(x) = \phi(x)$ on $[c,d] \subset (a,b)$, where $\{y_m(x)\}$ is a sequence of solutions of (1.1). Then, if $\phi(x) \in C^{(n-1)}[c,d]$, $\phi(x)$ is a solution of (1.1) on $[c,d]$ and $\lim_{m\to\infty} y_m^{(j)}(x) = \phi^{(j)}(x)$ uniformly on $[c,d]$ for each $0 \le j \le n-1$.*

Proof. Let $M > 0$ be such that $|\phi^{(j)}(x)| \le M$ on $[c,d]$ for $0 \le j \le n-1$. By Lemma 6.2 there is a $\delta > 0$ such that, if $c \le a_1 < a_2 < \cdots < a_n \le d$ is a fixed set of points with $a_n - a_1 \le \delta$, there is an $\varepsilon > 0$ with the property that $|y_i - \phi(a_i)| < \varepsilon$, $1 \le i \le n$ implies that (1.1) has a solution $y(x)$ satisfying $y(a_i) = y_i$, $1 \le i \le n$ and $|y^{(j)}(x)| \le 3M$ on $[a_1, a_n]$ for $0 \le j \le n-1$. It follows that there is an $N > 0$ such that $m \ge N$ implies $|y_m(a_i) - \phi(a_i)| < \varepsilon$, $1 \le i \le n$. Hence, by condition (D_n) and the choice of ε, $|y_m^{(j)}(x)| \le 3M$ on $[a_1, a_n]$ for $0 \le j \le n-1$ and all $m \ge N$. From this the conclusion follows.

Now let $\phi(x)$ be a real valued function defined on (c,d). At a point $x_0 \in (c,d)$ where $\phi(x)$ has a finite right limit $\phi(x_0+0)$, we define

$$D^1\phi(x_0+0) = \lim_{x\to x_0^+} \frac{\phi(x) - \phi(x_0+0)}{x - x_0}$$

provided the limit exists. The left derivative $D^1\phi(x_0-0)$ is similarly defined. Likewise, if $\phi(x_0+0)$ and $D^1\phi(x_0+0)$ exist and are finite, we define

$$D^2\phi(x_0+0) = \lim_{x\to x_0^+} \left\{ \frac{2}{(x-x_0)^2} \left[\phi(x) - \phi(x_0+0) - D^1\phi(x_0+0)(x-x_0) \right] \right\}$$

provided the limit exists. In general, if the limits defining $\phi(x_0+0)$ and $D^j\phi(x_0+0)$, $1 \le j \le k-1$ exist and are finite, we define

$$D^k\phi(x_0+0) = \lim_{x\to x_0^+} \left\{ \frac{k!}{(x-x_0)^k} \left[\phi(x) - \phi(x_0+0) - \sum_{j=1}^{k-1} \frac{D^j\phi(x_0+0)(x-x_0)^j}{j!} \right] \right\}$$

provided the limit exists. The left derivatives $D^j\phi(x_0-0)$ are defined correspondingly.

Theorem 6.4 ([13,17]). *Assume that the differential equation (1.1) satisfies conditions (A) and (D_n), and that $\phi(x)$ is a bounded generalized solution of (1.1) on $(c,d)\subset(a,b)$. Then, $\phi(x)$ has right and left limits at each point of (c,d) and $D^1\phi(x_0-0)$ and $D^1\phi(x_0+0)$ exist in the extended reals for all $x_0\in(c,d)$. Furthermore, if at a point $x_0\in(c,d)$, $D^j\phi(x_0+0)$ exists and is finite for each $1\le j\le k-1\le n-2$, then the limit defining $D^k\phi(x_0+0)$ exists in the extended reals. The same assertion applies to the left derivative $D^k\phi(x_0-0)$.*

Proof. Assume that for some $x_0\in(c,d)$, $\liminf_{x\to x_0^+}\phi(x)<\limsup_{x\to x_0^+}\phi(x)$ and choose a real number r such that $\liminf_{x\to x_0^+}\phi(x)<r<\limsup_{x\to x_0^+}\phi(x)$. Then, there exist sequences $\{t_m\}$ and $\{x_m\}$ in (c,d) such that $\lim t_m=\lim x_m=x_0$, $x_0<t_{m+1}<x_m<t_m$ for each $m\ge1$, $\lim\phi(t_m)=\limsup_{x\to x_0^+}\phi(x)$, and $\lim\phi(x_m)=\liminf_{x\to x_0^+}\phi(x)$. let $y(x)$ be a solution of (1.1) satisfying the initial conditions $y(x_0)=r$ and $y^{(j)}(x_0)=0$, $1\le j\le n-1$. This solution exists on $[x_0,x_0+\delta]$ for some $\delta>0$, and since $\lim_{x\to x_0}y(x)=r$, there is an $N>0$ such that $m\ge N$ implies that $x_0<t_m<x_0+\delta$ and $\phi(t_m)>y(t_m)$, $\phi(x_m)<y(x_m)$. This contradicts the fact that $\phi(x)$ is a generalized solution on (c,d). The existence of $\phi(x_0-0)$ can be proved similarly.

Now assume that for some $x_0\in(c,d)$ the limit defining $D^1\phi(x_0+0)$ does not exist in the extended reals. Then, choose the real number r such that

$$\liminf_{x\to x_0^+}\frac{\phi(x)-\phi(x_0+0)}{x-x_0}<r<\limsup_{x\to x_0^+}\frac{\phi(x)-\phi(x_0+0)}{x-x_0}.$$

If $y(x)$ is a solution of (1.1) satisfying the initial conditions $y(x_0)=\phi(x_0+0)$, $y'(x_0)=r$, and $y^{(j)}(x_0)=0$, $2\le j\le n-1$, again sequences $\{t_m\}$ and $\{x_m\}$ can be chosen so that $\lim t_m=\lim x_m=x_0$, $x_0<t_{m+1}<x_m<t_m$ for each $m\ge1$, and $\phi(t_m)>y(t_m)$, $\phi(x_m)<y(x_m)$ for all sufficiently large m. This again contradicts $\phi(x)$ being a generalized solution. Thus, $D^1\phi(x_0+0)$ and $D^1\phi(x_0-0)$ exist in the extended reals for all $x_0\in(c,d)$.

Finally, if we assume that for some $x_0\in(c,d)$, $D^j\phi(x_0+0)$ exists and is finite for each $1\le j\le k-1\le n-2$, then by considering a solution of (1.1) satisfying the initial conditions $y(x_0)=\phi(x_0+0)$, $y^{(j)}(x_0)=D^j\phi(x_0+0)$ for $1\le j\le k-1$, $y^{(k)}(x_0)=r$, and $y^{(j)}(x_0)=0$ for $k+1\le j\le n-1$, we can as above prove that the limit defining $D^k\phi(x_0+0)$ exists in the extended reals.

Corollary 6.5. *Assume that the differential equation (1.1) satisfies conditions (A) and (D_n), and that $\phi(x)$ is a bounded generalized solution of (1.1) on $(c,d)\subset(a,b)$. Then, $\phi(x)$ has a finite derivative $\phi'(x)$ almost everywhere on (c,d).*

Theorem 6.6 ([13,17]). *Assume that the differential equation (1.1) satisfies conditions (A) – (D_n). Let $\{y_m(x)\}$ be a sequence of solutions of (1.1) on $(c,d)\subset(a,b)$ such that $\{y_m(x)\}$ is uniformly bounded on (c,d) and $\lim y_m(x)=\phi(x)$ on*

(c,d). *Then, if for some* $x_0 \in (c,d)$ *the derivatives* $D^j\phi(x_0+0),\ 1 \le j \le n-1$ *all exist and are finite, or the derivatives* $D^j\phi(x_0-0),\ 1 \le j \le n-1$ *all exist and are finite, it follows that there is a subsequence* $\{y_{m(j)}(x)\}$ *such that* $\{y^{(i)}_{m(j)}(x)\}$ *converges uniformly on each compact subinterval of* (a,b) *for each* $0 \le i \le n-1$.

Proof. Assume that for some $x_0 \in (c,d)$ the derivatives $D^j\phi(x_0+0),\ 1 \le j \le n-1$ exist and are finite. Let $p(x)$ be the polynomial

$$p(x) = \phi(x_0+0) + \sum_{j=1}^{n-1} \frac{D^j\phi(x_0+0)(x-x_0)^j}{j!}$$

then, it follows from the definition of $D^{n-1}\phi(x_0+0)$ that given any $\varepsilon > 0$ there is a $\delta > 0$ such that $x_0+\delta < d$, and

$$|p(x) - \phi(x)| < \frac{\varepsilon(x-x_0)^{n-1}}{(n-1)!}$$

for $x_0 < x \le x_0+\delta$. Let d_0 be a fixed number satisfying $x_0 < d_0 < d$. By Lemma 6.2 there is a $\delta_0 > 0$ such that for $x_0 < x_1 < x_2 < \cdots < x_n \le d_0$ with $x_i - x_{i-1} = \eta \le \delta_0$ for each $1 \le i \le n$, (1.1) has a solution $y(x)$ with $y(x_i) = p(x_i)$ for $1 \le i \le n$ and $|y^{(j)}(x)| \le 2M$ on $[x_1, x_n]$ for $0 \le j \le n-1$ where $|p^{(j)}(x)| \le M$ on $[x_0, d_0]$ for $0 \le j \le n-1$. Furthermore, there is an $\varepsilon_0 > 0$ such that, if $|y_i - p(x_i)| < \varepsilon_0$ for $1 \le i \le n$, then (1.1) has a solution $y(x)$ with $y(x_i) = y_i$ for $1 \le i \le n$ and $|y^{(j)}(x)| \le 3M$ on $[x_1, x_n]$ for $0 \le j \le n-1$. It is not difficult to show that with equal spacing η between the $x_i's$ a suitable ε_0 has the form $\varepsilon_0 = Mh_n\eta^{n-1}$, where h_n is a fixed constant depending on n. Now as noted above, if we choose $\varepsilon = Mh_n/(2n^{n-1})$, there is a $\eta,\ 0 < \eta \le \delta_0$ such that $x_0 < x < x_0 + n\eta$ implies

$$|p(x) - \phi(x)| < \frac{\varepsilon(x-x_0)^{n-1}}{(n-1)!} \le \frac{\varepsilon_0}{2(n-1)!} \le \frac{\varepsilon_0}{2}.$$

For such a choice of $\eta > 0$, we have $|p(x_i) - \phi(x_i)| \le \varepsilon_0/2$ for $1 \le i \le n$ where $x_i - x_{i-1} = \eta$ for $1 \le i \le n$. Consequently, if $N > 0$ is such that $m \ge N$ implies $|y_m(x_i) - \phi(x_i)| < \varepsilon_0/2$ for $1 \le i \le n$, then $|p(x_i) - y_m(x_i)| < \varepsilon_0$ for $m \ge N$ and $1 \le i \le n$. It follows from our construction and condition (D_n) that $|y^{(j)}_m(x)| \le 3M$ on $[x_1, x_n]$ for $0 \le j \le n-1$ and all $m \ge N$. The conclusion of the theorem now follows.

Thus, we see that, in order to prove that conditions (A) – (D_n) imply the compactness condition (E), it is sufficient to prove that, if $\phi(x)$ is the pointwise limit of a bounded sequence of solutions of (1.1) on $(c,d) \subset (a,b)$, then there is at least one $x_0 \in (c,d)$ at which either $D^j\phi(x_0+0),\ 1 \le j \le n-1$ or $D^j\phi(x_0-0),\ 1 \le j \le n-1$ are finite.

7 The Case $q = 0$

As we have remarked in Section 6 for the differential equation (1.1) the compactness condition (E), under the assumptions (A) – (D_n), is equivalent to (E^*). This observation is used in the following result to establish the Proposition for the case $q = 0$.

Theorem 7.1 ([17]). *If the differential equation (1.1) with $q = 0$ satisfies conditions (A) – (D_n), then (1.1) with $q = 0$ also satisfies condition (E^*).*

Proof. Let $\{y_m(x)\}$ be a monotone, bounded sequence of solutions of (1.1) with $q = 0$ which converges point-wise to a function $\phi(x)$ on $[c, d] \subset (a, b)$. Let $c = a_1 < a_2 < \cdots < a_n = d$ and $p_m(x)$ be the unique polynomial of degree $n-1$ such that $p_m(a_i) = y_m(a_i)$, $1 \le i \le n$ and $m = 1, 2, \ldots$. Then, $p_m(x)$ converges uniformly to $p(x)$, where $p(x)$ is the unique polynomial of degree $n-1$ such that $p(a_i) = \lim_{m\to\infty} y_m(a_i)$, $1 \le i \le n$. Now, since $y_m(x)$ are uniformly bounded on $[c, d]$, it is clear that $M = \sup\{|f(x, y_m(x))| : c \le x \le d,\ m \ge 1\}$ exists. Further, from the properties of the Green's function $g(x, t)$, it follows that $\partial g/\partial x$ exists and is continuous on $[c, d] \times [c, d]$, and hence $|\partial g/\partial x| \le K$ for all $x, t \in [c, d]$. Thus, if $x \ne t$ from the integral representation

$$y_m(x) = p_m(x) + \int_c^d g(x, t) f(t, y_m(t)) dt, \tag{7.1}$$

which is the same

$$\omega_m(x) \equiv y_m(x) - p_m(x) = \int_c^d g(x, t) f(t, y_m(t)) dt$$

we find

$$|\omega_m(x) - \omega_m(s)| \le \int_c^d |g(x, t) - g(s, t)||g(t, y_m(t))| dt$$

$$\le MK|x - s|(d - c).$$

Hence, $\{\omega_m(x)\}$ is uniformly bounded and equicontinuous on $[c, d]$. Thus, a subsequence and by monotonicity the whole sequence $\{y_m(x)\}$ converges uniformly to $\phi(x)$ on $[c, d]$. Finally, taking limits through (7.1) yields that $\phi(x)$ is a solution of (1.1) with $q = 0$ on $[c, d]$, and hence the condition (E^*) is satisfied.

8 The Uniform Convergence

In [8] Henderson and Jackson in there closing remarks have mentioned the validity of the Proposition for fourth order differential equations. To prove the Proposition for arbitrary order differential equations we let P_n denote the set of all real-valued polynomials of degree at most n.

Definition 8.1 ([3]). Given $S \subset [c,d]$, x_0 is a *bilateral accumulation point* of S, in case x_0 is an accumulation point of both $S \cap [c,x_0]$ and $S \cap [x_0,d]$.

Definition 8.2. A function $g(x) : I \to \mathbb{R}$, I an interval, is said to be *n-convex (n-concave),* on I in case for any distinct points $x_0, x_1, \ldots, x_n$ in I,

$$\sum_{i=0}^{n} \frac{g(x_i)}{\omega'(x_i)} \geq 0, \ (\leq 0),$$

where

$$\omega(x) = \prod_{i=0}^{n}(x - x_i), \qquad \text{so that} \qquad \omega'(x_j) = \prod_{i=0, i\neq j}^{n} (x_j - x_i).$$

The following results for n-convex functions are well known.

Lemma 8.1. *Suppose $g(x) \in C^{(n)}(I)$. Then, $g(x)$ is n-convex on I, if and only if, $g^{(n)}(x) \geq 0$ on I.*

Lemma 8.2. *The function $g(x)$ is n-convex, if and only if, $g(x) \in C^{(n-2)}(I)$ and $g^{(n-2)}(x)$ is convex.*

Remark 8.1. In Lemmas 8.1 and 8.2, 'convex' can be replaced by 'concave'.

Lemma 8.3 ([3]). *Let $g(x) \in C[c,d]$ and assume that, for each $p(x) \in P_n$, the set $\{x : p(x) = g(x)\}$ does not have a bilateral accumulation point in (c,d). Then, there exists a subinterval $I \subseteq [c,d]$ on which $g(x)$ is either $(n+1)$-convex or $(n+1)$-concave.*

Theorem 8.4 ([21]). *Assume that the differential equation (1.1) satisfies the conditions (A) – (D_n). Then, (1.1) also satisfies condition (E).*

Proof. Let $\{y_m(x)\}$ be a sequence of solutions of (1.1) which is uniformly bounded on some subinterval $[c,d] \subset (a,b)$. Then, by Corollary 5.4 there exists a subsequence $\{y_{m(j)}(x)\}$ and a function $z(x) \in BV[c,d]$ such that $\lim_{j\to\infty} y_{m(j)}(x) = z(x)$ point-wise on $[c,d]$. Thus, $z'(x)$ exists a.e. on $[c,d]$ and $\int_c^d z(x)dx$ exists. We set

$$Z(x) = \int_c^x z(t)dt.$$

Then, $Z \in C[c,d]$, and by Lemma 8.3, either one of the following holds.

(i) $Z(x)$ is $(n+2)$-convex or $(n+2)$-concave on some $[c_1,d_1] \subseteq [c,d]$, or

(ii) there exists a $p(x) \in P_{n+1}$ such that $\{x : p(x) = Z(x)\}$ has a bilateral accumulation point in (c,d).

Case (i). Let us relabel the sequence $\{y_{m(j)}(x)\}$ as $\{y_m(x)\}$. By Lemma 8.2, $Z(x) \in C^{(n)}[c_1,d_1]$ and $Z^{(n)}(x)$ is convex, (or concave). Thus, $Z'(x) \in$

$C^{(n-1)}[c_1, d_1]$ and $Z'(x) = z(x)$ a.e. on $[c_1, d_1]$. Thus, as a consequence of Corollaries 2.6 and 2.7 there exists a $\delta = \delta(Z', d_1 - c_1) > 0$ such that, for fixed $c_1 \le a_1 < a_2 < \cdots < a_n \le d_1$, with $a_n - a_1 \le \delta$, there exists an $\varepsilon_0 > 0$ such that the boundary value problem for (1.1) satisfying $y(a_j) = Z'(a_j) + \varepsilon_j,\ 1 \le j \le n$, where $|\varepsilon_j| \le \varepsilon_0,\ 1 \le j \le n$, has a solution $y(x)$. Furthermore, the first $n-1$ derivatives of this solution are bounded, with bounds depending on Z' and $d_1 - c_1$. We call these bounds as $N_0 + 1, \ldots, N_{n-1} + 1$.

Let us now choose points $c_1 \le a_1 < a_2 < \cdots < a_n \le d_1$, with $a_n - a_1 \le \delta$ and such that $Z'(a_j) = z(a_j),\ 1 \le j \le n$. Then, there exists an M such that $|y_m(a_j) - z(a_j)| \le \varepsilon_0,\ 1 \le j \le n$ for all $m \ge M$. Now for $m \ge M$ and $1 \le j \le n$, let $\varepsilon_{m(j)} = y_m(a_j) - z(a_j)$. Then, for $m \ge M$,

$$y_m(a_j) = z(a_j) + \varepsilon_{m(j)} = Z'(a_j) + \varepsilon_{m(j)}, \quad 1 \le j \le n$$

and it follows from condition (D_n) that $y_m(x)$ is the solution referred to above resulting from Corollaries 2.6 and 2.7. As a consequence, we have for $m \ge M$, $|y_m^{(i)}(x)| \le N_i + 1$ on $[a_1, a_n]$ for each $0 \le i \le n-1$. Now we can apply Theorem 2.1 to obtain a further subsequence $\{y_{m(\ell)}(x)\}$ such that $\{y_{m(\ell)}^{(i)}(x)\}$ converges uniformly on each compact subinterval of (a, b) for each $0 \le i \le n-1$.

Case (ii). Assume that $\{x_m\} \downarrow x_0$ is such that $p(x_m) = Z(x_m)$, for all $m \ge 1$. Thus, if on some subinterval $[x_{j+1}, x_j]$, $Z'(x) = z(x) \ge p'(x)$ a.e., then we have

$$\begin{aligned} Z(x_j) - Z(x_{j+1}) &= \int_{x_{j+1}}^{x_j} z(t)dt \\ &\ge \int_{x_{j+1}}^{x_j} p'(t)dt \\ &= p(x_j) - p(x_{j+1}) \ = \ Z(x_j) - Z(x_{j+1}), \end{aligned}$$

so that the inequality is in fact an equality. However, from $z(x) - p'(x) \ge 0$, a.e. on $[x_{j+1}, x_j]$, we have $\int_{x_{j+1}}^{x_j} (z(t) - p'(t))dt = 0$, we conclude that $z(x) - p'(x) = 0$ a.e. on $[x_{j+1}, x_j]$. In particular,

$$Z'(x) = z(x) = p'(x) \quad \text{a.e.} \quad \text{on} \ \ [x_{j+1}, x_j].$$

Similarly, if we assume that on some subinterval $[x_{j+1}, x_j]$, $Z'(x) = z(x) \le p'(x)$ a.e., then we would arrive at $Z'(x) = z(x) = p'(x)$ a.e. on $[x_{j+1}, x_j]$.

Now that $p'(x) \in C^{(n-1)}[c, d]$, by Corollaries 2.6 and 2.7 there exists a $\delta = \delta(p', d - c) > 0$ such that for fixed $c \le a_1 < a_2 < \cdots < a_n \le d,\ a_n - a_1 \le \delta$, there exists an $\varepsilon_0 > 0$ such that the boundary value problem for (1.1) satisfying $y(a_j) = p'(a_j) + \varepsilon_j,\ 1 \le j \le n$, where $|\varepsilon_j| \le \varepsilon_0,\ 1 \le j \le n$, has a solution $y(x)$. As in Case (i) this solution has bounds on its first $n-1$ derivatives depending only on p' and $d - c$; again we call these bounds as $N_0 + 1, \ldots, N_{n-1} + 1$.

Let us choose points $x_{\ell+n} < x_{\ell+n-1} < \cdots < x_\ell$ from $\{x_m\}$ such that $x_\ell - x_{\ell+n} \le \delta$. We need to consider two subcases:

Case (a). For some $1 \le r \le n$, on $[x_{\ell+r}, x_{\ell+r-1}]$ we have $z(x) \ge p'(x)$ a.e., or $z(x) \le p'(x)$ a.e. From the above arguments we, however, have $z(x) = p'(x)$ a.e. on $[x_{\ell+r}, x_{\ell+r-1}]$. We repeat the arguments of Case (i). Choose points $x_{\ell+r} \le a_1 < \cdots < a_n \le x_{\ell+r-1}$ such that $z(a_j) = p'(a_j),\ 1 \le j \le n$. Then, there exists an M such that $|y_m(a_j) - z(a_j)| \le \varepsilon_0,\ 1 \le j \le n,\ m \ge M$. For $m \ge M$ and $1 \le j \le n$ let $\varepsilon_{m(j)} = y_m(a_j) - z(a_j)$. Then, for $m \ge M$,

$$y_m(a_j) = z(a_j) + \varepsilon_{m(j)} = p'(a_j) + \varepsilon_{m(j)}, \qquad 1 \le j \le n$$

and it follows that, from condition (D_n), $y_m(x)$ is the solution referred to the above problem arising from the Corollaries 2.6 and 2.7. Thus, for all $m \ge M$,

$$|y_m^{(i)}(x)| \le N_i + 1 \quad \text{on} \quad [a_1, a_n]$$

for each $0 \le i \le n-1$. Then, by Theorem 2.1 there exists a further subsequence $\{y_{m(s)}(x)\}$ such that $\{y_{m(s)}^{(i)}(x)\}$ converges uniformly on each compact subinterval of (a, b), for each $0 \le i \le n-1$.

Case (b). For each $1 \le r \le n$, there exist sets $A_r,\ B_r \subset [x_{\ell+r}, x_{\ell+r-1}]$, each having positive Lebesgue measure, and $z(x) > p'(x)$ on A_r, and $z(x) < p'(x)$ on B_r. However, since $\lim_{m\to\infty} y_m(x) = z(x)$, and so there exists a M such that for $m \ge M$, $y_m(x) > p'(x)$, for some A_r, and $y_m(x) < p'(x)$, for some $x \in B_r$. By continuity, for all $1 \le r \le n$, there exists $a_r \in (x_{\ell+r}, x_{\ell+r-1})$ such that $y_m(a_r) = p'(a_r)$.

In particular, there are points $x_{\ell+n} \le \tilde{a}_1 < \cdots < \tilde{a}_n \le x_\ell\ (\tilde{a}_n - \tilde{a}_1 \le \delta)$, so that, for some $M' \ge M$,

$$y_m(\tilde{a}_j) = p'(\tilde{a}_j) + \varepsilon_{m(j)}, \qquad 1 \le j \le n$$

where $|\varepsilon_{m(j)}| \le \varepsilon_0,\ 1 \le j \le n$ and all $m \ge M'$.

It follows from condition (D_n) that $y_m(x)$ is the solution referred to before Case (a) arising from Corollaries 2.6 and 2.7. Thus, for all $k \ge M'$,

$$|y_m^{(i)}(x)| \le N_i + 1 \quad \text{on} \quad [\tilde{a}_1, \tilde{a}_n],$$

for each $0 \le i \le n-1$. Now an application of Theorem 2.1 leads to a subsequence $\{y_{m(j)}(x)\}$ such that $\{y_{m(j)}^{(i)}(x)\}$ converges uniformly on each compact subinterval of (a, b), for each $0 \le i \le n-1$.

9 Problems and Comments

The establishment of Theorem 8.4 implies that in the known results on conjugate boundary value problems the condition (E) is, in fact, superfluous. As an example we state an important result, which was independently proved by Hartman [6] and Klassen [17] with the additional condition (E).

Theorem 9.1. *Assume that for the differential equation* (1.1) *conditions* (A) – (D_n) *are satisfied. Then, each* r *point boundary value problem, i.e., for any* $a < a_1 < a_2 \cdots < a_r < b$ *and any* $A_{j+1,i}$, $0 \le j \le k_i$, $1 \le i \le r$ *the problem* (1.1), (2.3) *has a unique solution.*

Problem 1. For the third order differential equations in Theorem 4.5 we have proved that conditions (A), (C) and (D_3) imply condition (E). It will be interesting to extend this result to equations of arbitrary order, i.e., whether it is possible to prove Theorem 8.4 without the assumption (B).

In [16] Jackson has indicated that for n-point boundary value problems, Klassen has used a result he established in [18] to prove the existence of solutions under the assumptions (A), (C), (D_n) and (E). Thus, if the answer to Problem 1 is affirmative, then for $r = n$, Theorem 9.1 holds without the assumption (B).

Let $2 \le r \le n$ and let m_i, $1 \le i \le r$, be positive integers such that $\sum_{i=1}^{r} m_i = n$. Let $s_0 = 0$ and for $1 \le k \le r$, let $s_k = \sum_{i=1}^{k} m_i$. A boundary value problem for (1.1) with the boundary conditions

$$y^{(i)}(a_k) = y_{i,k}, \qquad s_{k-1} \le i \le s_k - 1, \quad 1 \le k \le r \tag{9.1}$$

where $a < a_1 < a_2 < \cdots < a_r < b$ is called a *right* $(m_1, \ldots, m_r)$*-focal point boundary value problem for* (1.1) *on* (a, b).

With respect to the boundary conditions (9.1) we replace the condition (D_n) by the following:

(D_n^{rf}) For any $a < a_1 < a_2 < \cdots < a_n < b$ and any solutions $y(x)$ and $z(x)$ of (1.1), it follows that $y^{(i-1)}(a_i) = z^{(i-1)}(a_i)$, $1 \le i \le n$ implies $y(x) \equiv z(x)$, i.e., the differential equation (1.1) is right $(1, 1, \ldots, 1)$ disfocal on (a, b).

As an application of Rolle's theorem it follows that condition (D_n^{rf}) implies the condition (D_n). Thus, in Theorem 8.4 condition (D_n) can be replaced by (D_n^{rf}). We state this observation in the following result.

Theorem 9.2. *Assume that the differential equation* (1.1) *satisfies conditions* (A) – (C) *and* (D_n^{rf}). *Then,* (1.1) *also satisfies condition* (E).

Of course, in Theorem 8.4, we can always replace condition (D_n) by a stronger condition. The point is now whether it is possible to replace condition (D_n) by some other condition which does not imply (D_n). The first 'round about' result in this direction is the following:

Theorem 9.3. *If the differential equation* (1.1) *is of third order and satisfies conditions* (A), (C) *and* (D_2), *then* (1.1) *also satisfies condition* (E).

Proof. The proof is similar to that of Theorems 4.5.

Problem 2. A result similar to that of Theorem 9.3 for arbitrary order differential equations (1.1) remains undecided, i.e., does conditions (A), (C) and (D_r) imply condition (E).

For arbitrary order differential equations (1.1) Jackson [14] has established that conditions (A), (B) and (D_n) imply (D_r). A converse of this result for third order differential equations (1.1) is that the conditions (A), (C) and (D_2) imply (D_3). Jackson's proof [15] of this converse result uses Theorem 9.3, i.e., under the assumptions, condition (E) is implied, and then this fact is used to prove (D_3). Thus, if we accept Jackson's converse result without looking at its proof, then we can argue that conditions (A), (C) and (D_2) imply (D_3), and therefore Theorem 4.5 gives condition (E).

Problem 3. The question for arbitrary order differential equations (1.1) which remains open is whether conditions (A), (C) and (D_r) imply condition (D_n).

Finally, we state one more result which is similar in nature to that of Theorem 9.3.

Theorem 9.4 ([7]). *If the differential equation* (1.1) *is of third order and satisfies conditions* (A), (C) *and*

(D_2^{rf}) each right $(2,1)$-focal point boundary value problem for (1.1) on (a,b) has at most one solution,

then (1.1) also satisfies condition (E).

Problem 4. A result similar to that of Theorem 9.4 for arbitrary order differential equations (1.1) is not known.

The author is grateful to Professor Johnny Henderson for his help in the preparation of this lecture.

References

1. R. P. Agarwal, *Boundary Value Problems for Higher Order Differential Equations*, World Scientific, Singapore, 1986
2. R. P. Agarwal and P. J. Y. Wong, *Error Inequalities in Polynomial Interpolation and their Applications*, Kluwer, The Netherlands, 1993
3. S. Agronsky, A. M. Bruckner, M. Laczkovich and D. Preiss, *Convexity conditions and intersections with smooth functions*, Trans. Amer. Math. Soc. **289** (1985), 659–677
4. P. Hartman, *Unrestricted n-parameter families*, Rend. Circ. Mat. Palermo **7** (1958), 123–142
5. P. Hartman, *Ordinary Differential Equations*, Wiley, New York, 1964
6. P. Hartman, *On N-parameter families and interpolation problems for nonlinear ordinary differential equations*, Trans. Amer. Math. Soc. **154** (1971), 201–226
7. J. Henderson, *Existence and uniqueness of solutions of right focal point boundary value problems for third and fourth order equations*, Rocky Mountain J. Math. **14** (1984), 487–497

8. J. Henderson and L. Jackson, *Existence and uniqueness of solutions of k-point boundary value problems for ordinary differential equations*, J. Differential Equations **48** (1983), 373–385
9. E. Hewitt and K. Stromberg, *Real and Abstract Analysis. A Modern Treatment of the Theory of Functions of a Real Variables*, Springer-Verlag, New York, 1965
10. L. Jackson and G. Klaasen, *Uniqueness of solutions of boundary value problems for ordinary differential equations*, SIAM J. Appl. Math. **19** (1970), 542–546
11. L. Jackson and K. Schrader, *Subfunctions and third order differential inequalities*, J. Differential Equations **8** (1970), 180–194
12. L. Jackson and K. Schrader, *Existence and uniqueness of solutions of boundary value problems for third order differential equations*, J. Differential Equations **9** (1971), 46–54
13. L. Jackson, *Uniqueness and existence of solutions of boundary value problems for ordinary differential equations*, Proc. NRL-MRC Conference on Ordinary Differential Equations, Washington, D. C., Academic Press, New York, 1972, 137–149
14. L. Jackson, *Uniqueness of solutions of boundary value problems for ordinary differential equations*, SIAM J. Appl. Math. **24** (1973), 535–538
15. L. Jackson, *Existence and uniqueness of solutions of boundary value problems for third order differential equations*, J. Differential Equations **13** (1973), 432–437
16. L. Jackson, *A compactness condition for solutions of ordinary differential equations*, Proc. Amer. Math. Soc. **57** (1976), 89–92
17. G. Klaasen, *Existence theorems for boundary value problems of nth order ordinary differential equations*, Rocky Mountain J. Math. **3** (1973), 457–472
18. G. Klaasen, *Continuous dependence for N-point boundary value problems*, SIAM J. Appl. Math. **29** (1975), 99–102
19. K. Schrader, *A pointwise convergence theorem for sequences of continuous functions*, Trans. Amer. Math. Soc. **159** (1971), 155–163
20. K. Schrader, *A generalization of the Helly selection theorem*, Bull. Amer. Math. Soc. **78** (1972), 415–419
21. K. Schrader, *Uniqueness implies existence for solutions of nonlinear boundary value problems*, Abstracts Amer. Math. Soc. **6** (1985), 235
22. A. E. Taylor, *General Theory of Functions and Integration*, Blaisdell Pub. Comp., Waltham, 1965

EQUADIFF 9 CD ROM, Brno 1997 PROCEEDINGS
Masaryk University pp. 25–48

Parabolic Equations with Multiple Singularities

Emmanuele DiBenedetto*

Department of Mathematics
Northwestern Univ. Evanston ILL. 60208–2730, USA
Email: dibenede@giotto.math.nwu.edu

Abstract. Models of flow of multiple immiscible fluids in a porous matrix and/or phenomena of multiple transitions of phase, result into quasilinear parabolic equations, with measurable coefficients and exhibiting multiple singularities and/or degeneracies (in the sense made precise in Section 1.1 below). We discuss the problem of the continuity of the transition parameters, for example saturation in the flow of immiscible fluids, or temperature in isothermal phase transitions. We review and summarize the main points of the theory and will present some recent results in this direction, pointing to the new mathematical tools generated by these investigations. We will also indicate the main open questions of physical and mathematical interest and discuss their relevance.

AMS Subject Classification. 35K20, 35K40, 35K55, 35K65, 35K99, 35Q35

Keywords. Degenerate Parabolic Equations, Singular parabolic equations, immiscible fluids, phase transition, Harnack estimates, Hadamard growth, modulus of continuity, DeGiorgi estimates, intrinsic geometry, measure theory

1 Introduction

We will present some recent results concerning the local behavior of weak solutions of singular parabolic equations with measurable coefficients. We will indicate the main points of the theory and will trace back their motivation to physical phenomena, such as transition of phase and/or flow of immiscible fluids in a porous matrix. In this connection, we will also indicate some novel analytical ideas of measure theory which we feel are of independent interest. Along the presentation we will point out the main open problems, which we feel have both a theoretical and physical interest.

* Partially supported by NSF Grant DMS–9706388

1.1 Singular Parabolic Equations

Let $\beta(\cdot)$ be a maximal monotone graph in $\mathbb{R}\times\mathbb{R}$ and consider parabolic inclusions of the type

$$\frac{\partial}{\partial t}\beta(u) - \operatorname{div}\mathbf{A}(x,t,u,\nabla u) + B(x,t,u,\nabla u) \ni 0 \quad \text{in } \Omega_T, \tag{1.1}$$

where Ω is a domain in $\mathbb{R}^N$ and ∇ denotes the gradient with respect to the space variables only. Also, for $T>0$ we have set $\Omega_T \equiv \Omega\times(0,T]$. We assume that the graph $\beta(\cdot)$ is coercive, i.e., there exists a positive constant γ_o, such that for all pairs of real numbers $(s_1,\, s_2)$ and all selections $w_1\in\beta(s_1)$ and $w_2\in\beta(s_2)$,

$$w_1 - w_2 \geq \gamma_o(s_1 - s_2). \tag{1.2}$$

We also assume that $\beta(\cdot)$ is bounded for bounded values of its argument, i.e.,

$$\text{for every } M>0, \qquad \sup_{|s|<M}\ \sup_{w\in\beta(s)} |w| < \infty. \tag{1.3}$$

No further condition is formulated on the behavior of $\beta(\cdot)$. In particular in any finite interval $(-M, M)$, the graph $\beta(\cdot)$ might exhibit countably many jumps or might become vertical countably many times, exponentially fast or faster. If a graph $\beta(\cdot)$ exhibits this behavior we call it a *singular* graph and refer to (1.1) as singular parabolic equations. Examples of such a $\beta(\cdot)$ are

$$\beta(s)\equiv\begin{cases} s & \text{if } s<0,\\ [0,1] & \text{if } s=0,\\ 1+s & \text{if } s>0;\end{cases} \qquad \beta(s)\equiv\begin{cases} 2+s & \text{if } s>1,\\ [2,3] & \text{if } s=1,\\ 1+s & \text{if } 0<s<1,\\ [0,1] & \text{if } s=0,\\ s & \text{if } s<0;\end{cases} \tag{i}$$

$$\begin{aligned} &\beta(s)\equiv |s|^{\frac{1}{m}}\operatorname{sign} s, \qquad m>1;\\ &\beta(s)\equiv 1+s^{\alpha_1}-(1-s)^{\alpha_2}, \quad \begin{cases} s\in[0,1],\\ \alpha_i\in(0,1), i=1,2.\end{cases}\end{aligned} \tag{ii}$$

The first of (**i**) is the enthalpy function in the weak formulation of a Stefan-like problem modeling a water-ice transition of phase.[1] The second might serve as a prototype of the enthalpy in a double transition of phase. The first of (**ii**) is the graph arising from the classical porous media equation, modeling the flows

[1] There exists a vast literature on each of the several aspects of the classical Stefan problem. For a summary of the main results we refer to the monograph of Meirmanov [42], the review article of Danilyuk [13] as well as the Proceedings [8,26,30] and the references therein. Here we review only those aspects connected with the local continuity of weak solutions of (1.1) with $\beta(\cdot)$ exhibiting multiple singularities.

of a *single* fluid in a porous matrix.[2] The second, is a first approximation for a model of two immiscible fluids moving within a porous matrix.[3] The simplest example of (1.1) is,

$$\frac{\partial}{\partial t}\beta(u) - \Delta u \ni 0 \quad \text{in} \quad \Omega_T. \tag{1.1$'$}$$

The diffusion field $\mathbf{A}$ and the forcing term B in (1.1), are real valued and measurable over $\Omega_T \times \mathbb{R} \times \mathbb{R}^N$, and satisfy the structure conditions,

$$\begin{aligned} \mathbf{A}(x,t,\eta,\xi)\cdot\xi &\geq \mu_o\,|\xi|^2 - \varphi_o(x,t);\\ |\mathbf{A}(x,t,\eta,\xi)| &\leq \mu_1\,|\xi| - \varphi_1(x,t);\\ |B(x,t,\eta,\xi)| &\leq \mu_2\,|\xi|^2 - \varphi_2(x,t), \end{aligned} \tag{1.4}$$

for a.e. $(x,t,\eta,\xi) \in \Omega_T \times \mathbb{R} \times \mathbb{R}^N$. Here μ_i, $i = 0,1,2$ are prescribed positive numbers and φ_i, $i = 0,2$ are prescribed nonnegative functions defined a.e. in Ω_T, satisfying

$$\varphi_o + \varphi_1^2 + \varphi_2 \in L^{\overline{q},\overline{r}}_{\rm loc}(\Omega_T). \tag{1.5}$$

The numbers $\overline{q}$ and $\overline{r}$ are positive, are linked by

$$\frac{1}{\overline{r}} + \frac{N}{\overline{q}} = 1 - \overline{\kappa}, \qquad \overline{\kappa} \in (0,1), \tag{1.6}$$

and can be taken out of their admissible range

$$\begin{aligned} &\overline{q} \in \left[\tfrac{N}{2(1-\overline{\kappa})}\,,\,\infty\right], \quad \overline{r} \in \left[\tfrac{1}{1-\overline{\kappa}}\,,\,\infty\right], \quad 0 < \overline{\kappa} < 1, \quad \text{for } N \geq 2;\\ &\overline{q} \in (1\,,\,\infty), \quad \overline{r} \in \left[\tfrac{1}{1-\overline{\kappa}}\,,\,\tfrac{1}{1-2\overline{\kappa}}\right], \quad 0 < \overline{\kappa} < \tfrac{1}{2}, \quad \text{for } N = 1. \end{aligned} \tag{1.7}$$

The inclusion in (1.1) is meant weakly and in the sense of graphs. Precisely, a function

$$u \in L^2_{\rm loc}\left\{0,T;W^{1,2}_{\rm loc}(\Omega)\right\}, \tag{1.8}$$

is a local weak solution to (1.1) if there exists a measurable selection $w \subset \beta(u)$, such that

$$t \to w(\cdot,t) \text{ is weakly continuous in } L^2_{\rm loc}(\Omega), \tag{1.9}$$

[2] Also the porous medium equation has been widely investigated in the literature and we refer to the same Proceedings [8,26,30] and their references, for an overview. An overview of the main results regarding the local regularity of the solutions, is in the Bibliographical Notes of the monograph [22].

[3] A 1-dimensional model in hydrology is investigated by Van Duijn and Zhang in [15], and numerically by Hoff [29]. Most of the models of multiphase flows in porous medium are multidimensional. For such models we refer to the monographs [6,7,11,12,49].

and in addition,

$$\int_\Omega w(x,t)\varphi(x,\tau)\,dx\Big|_{\tau=t_1}^{\tau=t_2} + \int_{t_1}^{t_2}\int_\Omega \left\{ -w(x,\tau)\varphi_t + \mathbf{A}(x,\tau,u,\nabla u)\cdot D\varphi \right\} dxd\tau + \int_{t_1}^{t_2}\int_\Omega B(x,\tau,u,\nabla u)\varphi\,dxd\tau = 0, \quad (1.10)$$

for all testing functions

$$\varphi \in W^{1,2}_{\rm loc}\left\{0,T;L^2_{\rm loc}(\Omega)\right\} \cap L^2_{\rm loc}\left\{0,T;W^{1,2}_o(\Omega)\right\}, \quad (1.11)$$

and for all intervals $(t_1,t_2)\subset(0,T]$.

2 The Problem of Continuity of Weak Solutions

It is natural to ask whether *locally bounded* weak solutions to (1.1) are continuous in Ω_T and whether one can estimate *quantitatively* their modulus of continuity. To simplify the setting of the problem, we assume that u is a solution of (1.1) bounded in the whole Ω_T and set,

$$\|u\|_{\infty,\Omega_T} \equiv M. \quad (2.1)$$

This is not restrictive, by regarding Ω_T as a subset of the domain of definition of u. By the same token we also assume that the integrability requirement in (1.5) holds in Ω_T and set,

$$\left\|\varphi_o+\varphi_1^2+\varphi_2\right\|_{\overline{q},\overline{r};\Omega_T} \equiv \Phi. \quad (2.2)$$

We refer to the numbers,

$$N,\quad \gamma_o,\quad M,\quad \Phi,\quad \mu_i, i=0,1,2,$$

as the *data*. For a constant C or γ, or a *continuous* function $\omega(\cdot)$ we say

$$C\equiv C(\text{data}),\quad \gamma\equiv\gamma(\text{data}),\quad \omega(\cdot)=\omega_{\rm data}(\cdot),$$

if they can be determined a priori only in terms of the indicated parameters. Having fixed an arbitrary subset $\mathcal{K}\subset\Omega_T$, one can ask whether u is continuous in $\mathcal{K}$ with a modulus of continuity $\omega_{\rm data}(\cdot)$ depending only upon the data and the distance from $\mathcal{K}$ to the parabolic boundary of Ω_T.

Remark 2.1. If $\beta(\cdot)\equiv\mathbb{I}$, then locally bounded solutions of (1.1) are locally Hölder continuous in Ω_T, and the assumptions (1.5)–(1.7) are optimal for this to occur.[4] Thus the issue at hand is to investigate to what extend the singularity of $\beta(\cdot)$ might affect the continuity of u.

[4] For a general account of the theory of local regularity of solutions of non-singular parabolic equations with measurable coefficients, we refer to the monograph [39], and in particular Chap. I, §3,4; Chap. II, §6,7; Chap. V, §1,2.

Remark 2.2. The assumption that u be locally bounded is essential. Indeed even if $\beta(\cdot) \equiv \mathbb{I}$, weak solutions of (1.1) need not be bounded. This is due to the critical growth of the forcing term $B(x,t,u,\nabla u)$ with respect to $|\nabla u|$ as indicated in the last of (1.4). We refer Stampacchia [51] for counterexamples even in the elliptic case.[5] If the last of (1.4) were replaced with

$$|B(x,t,\eta,\xi)| \leq \mu_2 |\xi| + \varphi_2(x,t), \tag{1.4'}$$

then weak solutions of (1.1), for any coercive $\beta(\cdot)$ as in (1.2)–(1.3) would be locally bounded. This would follow from a simple adaptation of the methods of [39].[6]

In what follows we will assume in addition that the local solution u can be constructed as the limit in the topology of (1.8), of a sequence of smooth local solutions of (1.1) for *smooth* $\beta(\cdot)$. This assumption is formulated only to justify some of the calculations.[7] We stress that the modulus of continuity of u must be independent of any approximating procedure and must depend only upon the data.

3 Some Degenerate Parabolic Equations

The full generality indicated in (1.4)–(1.6) seems to be natural in physical models, such as the simultaneous flow of two immiscible fluids in a porous matrix.[8] These models typically lead to degenerate parabolic equations of the type,[9]

$$v_t - \operatorname{div} \mathbf{a}(x,t,v,\nabla v) + b(x,t,v,\nabla v) = 0 \quad \text{in } \Omega_T. \tag{3.1}$$

[5] Thus (1.1) even with $\beta(\cdot) = \mathbb{I}$ might have unbounded solutions. However if one had some a priori qualitative knowledge of the boundedness of the solution, such a qualitative bound could be turned into a quantitative one. See for example Vespri [53] and references therein.

[6] Indeed a slightly faster growth is allowed; for example $|\nabla u|^q$ where $0 \leq q < \frac{N+4}{N+2}$. See [39] Chap. V, §1.

[7] If the forcing term $B(x,t,u,\nabla u)$ has at most a linear growth with respect to $|\nabla u|$, then questions of existence and uniqueness are well understood. We refer for example to the monographs [27,39,41] and the Proceedings [8,26,30] and references therein. Here we only remark that a modulus of continuity uniform with respect to the approximating procedure, would supply the necessary compactness to establish existence of solutions.

[8] For these models we refer to the monographs of J. Bear [6] (Chap. 9) and [7] (Chap. 6), R. E. Collins [12] (Chap. 6), and A.E. Scheidegger [49] (Chap. 10), and the article of Leverett [40]. These models consist of a system of two parabolic equations, written in terms of the saturations and pressures of each of the two fluids.

[9] The transformation of Kruzkov-Sukorjanski [37], transforms the physical models of [6,7,12,40,49] into a system of one parabolic equation like (3.1) in terms of the saturation v of only one of the two fluids, and another degenerate-elliptic equation in terms of a mean pressure. In such a formulation, the term $b(x,t,v,\nabla v)$ in (3.1) would depend on such a mean pressure. The local continuity for the saturations was first raised in [1] and [21]. The analysis of [1,21] permits to reduce the question of the continuity of the saturations to the continuity of solutions to (3.1).

The leading vector field $\mathbf{a}$ and the forcing term b, are measurable and satisfy,

$$\begin{aligned}\mathbf{a}(x,t,v,\nabla v)\cdot\nabla v &\geq C_o\varphi(v)|\nabla v|^2-\varphi_o(x,t);\\ |\mathbf{a}(x,t,v,\nabla v)| &\leq C_1\varphi(v)|\nabla v|-\varphi_1(x,t);\\ |b(x,t,v,\nabla v)| &\leq C_2\varphi(v)|\nabla v|^2-\varphi_2(x,t),\end{aligned}\tag{3.2}$$

for a.e. $(x,t)\in\Omega_T$ and all smooth functions $(x,t)\to v(x,t)$ defined in Ω_T. Because of the physical origin of the p.d.e., it is natural to assume that the solutions are bounded, say for example $v\in[0,1]$.[10] The equation is degenerate in the sense that $\varphi(\cdot)$ is permitted to vanish. Precisely we assume that $v\to\varphi(v)$ is continuous, non-negative and vanishes at the extreme values of its argument, i.e.,

$$\varphi(v)>0 \text{ for } v\in(0,1) \text{ and } \varphi(0)=\varphi(1)=0.\tag{3.3}$$

The functions φ_i, $i=0,1,2$ satisfy the assumptions (1.5)–(1.7). A notion of solution to (3.1) is introduced along the lines of (1.8)–(1.11), by requiring that $t\to v(\cdot,t)$ satisfies (1.9) and that

$$\nabla\varphi(v)\in L^2_{\rm loc}(\Omega_T).$$

The main difficulty in establishing the local continuity of v resides in the double degeneracy of $\varphi(\cdot)$ and, more importantly, in the lack of precise quantitative and/or qualitative information on its modulus of continuity. Such a limited information on the nature of the degeneracy is typical of the physical models of flows of a mixture of fluids in a porous medium.[11] Thus in particular $\varphi(\cdot)$ might degenerate at $v=0$ and $v=1$ at different rates, and perhaps exponentially fast or faster.[12] The problem of continuity of weak solutions to (3.1) consists in showing that v is continuous *whatever* the nature of the degeneracy of $\varphi(\cdot)$, provided (3.3) is satisfied.

Let $u\in[0,1]$ be a solution of (1.1) with $\beta(\cdot)\in C(0,1)$ and singular at the extreme values $u=0$ and $u=1$ of its argument, i.e. for example

$$\lim_{u\searrow 0}\beta'(u)=\lim_{u\nearrow 1}\beta'(u)=+\infty.$$

[10] The function $(x,t)\to v(x,t)$ is the local relative saturation of one of the two fluids. Thus $v\in[0,1]$. See for example [1,6,7,12,37,49].

[11] The function $\varphi(\cdot)$ is related to the permeability of both fluids. The permeability of one of the fluids vanishes as the fluid is displaced by the other (i.e., either $v=0$ or $v=1$). This is the physical origin of the degeneracy of $\varphi(\cdot)$. The behaviour of the permeabilities as functions of the saturations are derived from hydrostatic (rather than dynamic) experiments, [6,7,12,49], dimensional analysis [40], and heuristic arguments. For this reason the information on their rate of vanishing is rather limited.

[12] In fact, because of the phenomenon of the connate water it might be even completely flat in a small right neighborhood of zero or a left interval of 1, or both. See Bear [6] Chap. 9, §2.3 and 2.4; Collins [12] Chap. 2, §24, and Chap.6, §10; Scheidegger [49] Chap. 3, §4 and Chap.10, §6.

Then by setting $v \equiv \beta(u)$ and $\varphi(\cdot) \equiv \beta^{-1\prime}(\cdot)$ the singular p.d.e. in (1.1), in terms of u, can be recast as the degenerate p.d.e. in (3.1), in terms of v, and one checks that the conditions (1.4) yield (3.2). For this reason, the methods introduced in the context of (1.1) and those connected to (3.1) bear a considerable similarity and/or overlap. Starting now from (3.1) one might set

$$u \equiv \int^v \varphi(s)\, ds, \qquad \beta(u) = v,$$

and attempt to recast the degenerate p.d.e. (3.1) as the singular equation (1.1). One verifies that the resulting leading coefficients $\mathbf{A}$ would satisfy the first two of (1.4). The resulting free term B however might not satisfy the last (1.4), due to its faster than linear growth with respect to $|\nabla v|$.[13] In what follows we will outline the analogies and point to the main differences.

4 The Classical Approach to Continuity

For positive ρ, let K_ρ and Q_ρ denote respectively the cube of wedge 2ρ centered at the origin of $\mathbb{R}^N$, and the parabolic cylinder with "vertex" at the origin of $\mathbb{R}^{N+1}$, with cross sections K_ρ, i.e.,

$$K_\rho \equiv \{x \in \mathbb{R}^N | \max_{1\le i\le N} |x_i| < \rho\}, \qquad Q_\rho \equiv K_\rho \times (-\rho^2, 0). \tag{4.1}$$

A cube centered at some $x_o \in \mathbb{R}^N \backslash \{0\}$ and congruent to K_ρ will be denoted by $\{x_o + K_\rho\}$ and a parabolic cylinder with "vertex" at some $(x_o, t_o) \in \mathbb{R}^{N+1}$ and congruent to Q_ρ, will be denoted by $\{(x_o, t_o) + Q_\rho\}$. In what follows we will fix a point $(x_o, t_o) \in \Omega_T$ and let ρ_o be the largest radius so that $\{(x_o, t_o) + Q_{\rho_o}\}$ is contained in Ω_T. Also for a constant $\delta \in (0, 1)$ we consider the sequence of decreasing radii,

$$\rho_n \equiv \delta^n \rho_o, \qquad n = 0, 1, 2 \ldots, \tag{4.2}$$

and the family of nested shrinking cylinders, with the same vertex at (x_o, t_o),

$$\{(x_o, t_o) + Q_{\rho_n}\}, \qquad n = 0, 1, 2, \ldots .$$

4.1 Non Singular Parabolic Equations

Suppose for the moment that in (1.1), the graph $\beta(\cdot)$ is the *identity*, i.e., that (1.1) is a quasilinear, *non-singular*, parabolic equation with measurable coefficients. If u is a weak solution to such an equation, we set

$$\mu_n^+ \equiv \operatorname*{ess\,sup}_{\{(x_o,t_o)+Q_{\rho_n}\}} u, \quad \mu_n^- \equiv \operatorname*{ess\,inf}_{\{(x_o,t_o)+Q_{\rho_n}\}} u, \quad \omega_n \equiv \operatorname*{ess\,osc}_{\{(x_o,t_o)+Q_\rho\}} u.$$

[13] One verifies this for the equation $v_t - \Delta v^2 = v|\nabla v|^2$. The equivalence of the two formulations would hold if B had a linear growth with respect to $|\nabla u|$. Equations such as (3.1) arising from the flow of immiscible fluids in a porous medium bear lower order terms with a behavior technically similar to a super-linear growth with respect to $|\nabla v|$. See [1], §3–6.

Proposition 4.1. *Let u be a weak solution of (1.1) with $\beta(\cdot) \equiv \mathbb{I}$. Then there exists constants $C > 1$ and $\delta, \eta \in \left(0, \frac{1}{2}\right)$ that can be determined a priori only in terms of the data, such that for every $(x_o, t_o) \in \Omega_T$,*

$$\omega_{n+1} \le (1-\eta)\omega_n + C\rho_n^\lambda, \quad n = 0, 1, 2, \dots . \tag{4.3}$$

Here $\lambda \in (0,1)$ is a number determined only in terms of the integrability conditions (1.5)–(1.7) and is independent of δ and η. As a consequence u is locally Hölder continuous in Ω_T.

Proof of Hölder continuity assuming (4.3). Having fixed $(x_o, t_o) \in \Omega_T$, from (4.3) by iteration we derive,

$$\omega_n \le (1-\eta)^n \omega_o + \frac{C}{\delta^\lambda} \sum\nolimits_{i=1}^{n} \left(\frac{1-\eta}{\delta^\lambda}\right)^n, \quad \forall n \in \mathbb{N}. \tag{4.4}$$

The two numbers $(1-\eta)$ and δ can be related by

$$(1-\eta) = \delta^\alpha, \qquad \text{where} \qquad \alpha = \frac{\ln(1-\eta)}{\ln \delta} \in (0,1).$$

Moreover without loss of generality we may assume that $\rho_o \in (0,1)$. Then, having determined δ and η, the iterative inequalities (4.3) continue to hold if λ is replaced by a smaller number. We will choose it so that $(1-\eta)\delta^{-\lambda} < 1$. This way the sum on the right hand side of (4.4) can be majorized with a convergent series. Therefore from (4.4) and the definition (4.2) of the sequence ρ_n, it follows that,

$$\omega_n \le \omega_o \left(\frac{\rho_n}{\rho_o}\right)^\alpha + \gamma(\text{data}; \eta, \delta)\rho_n^\lambda, \quad \forall n \in \mathbb{N}. \tag{4.5}$$

Since $(x_o, t_o) \in \Omega_T$ is arbitrary, this implies that u is locally Hölder continuous in Ω_T, with Hölder exponent $\min\{\alpha; \lambda\}$. □

Remark 4.1. Having fixed $(x_o, t_o) \in \Omega_T$, the starting cylinder $\{(x_o, t_o) + Q_{\rho_o}\}$ must be contained in Ω_T. Thus from the form of (4.5) it follows that the Hölder continuity can be claimed only within compact subsets $\mathcal{K}$ of Ω_T and that the Hölder constant $\omega_o \rho_o^{-\alpha}$ deteriorates as (x_o, t_o) approaches the parabolic boundary of Ω_T.

Remark 4.2. The constant C appearing on the right hand side of (4.3) is due only to the functions φ_i in the structure conditions (1.4) and it would be zero for the prototype equation (1.1′), with $\beta(\cdot) \equiv \mathbb{I}$.

This is the parabolic version of the classical DeGiorgi's approach to continuity introduced in the context of elliptic equations with measurable coefficients [14]. The adaptation to parabolic equations is far from simple and it appears in the book [39]. The same point of view of reducing the oscillation of u in a family

of *shrinking cylinders* has influenced, one way or another, most of the literature on the subject, including Moser [43,44], Trudinger [52], Kruzkov [34,35,36], Aronson-Serrin [5] and Krylov-Safonov [38]. The reduction of the oscillation (4.3) is realized by the following Proposition that can be regarded as some sort of a weak maximum principle.[14]

Proposition 4.2. *Let u be a weak solution of (1.1) with $\beta(\cdot) \equiv \mathbb{I}$. Then there exists constants $C > 1$ and $\delta, \eta \in \left(0, \frac{1}{2}\right)$, that can be determined a priori only in terms of the data, such that for every $(x_o, t_o) \in \Omega_T$ and every $n \in \mathbb{N}$, either $\omega_n < C\rho_n^\lambda$, or at least one of the following two inequalities holds,*

$$\begin{aligned} u(x,t) &\le \mu_n^+ - \eta\omega_n, \\ u(x,t) &\ge \mu_n^- + \eta\omega_n, \end{aligned} \qquad \text{for a.e. } (x,t) \in \left\{(x_o,t_o) + Q_{\rho_{n+1}}\right\}. \tag{4.6}$$

Proof of (4.3) assuming (4.6). Fix $(x_o, t_o) \in \Omega_T$ and $n \in \mathbb{N}$. If the first of (4.6) holds true, then

$$\operatorname*{ess\,sup}_{\{(x_o,t_o)+Q_{\rho_{n+1}}\}} u = \mu_{n+1}^+ \le \mu_n^+ - \eta\omega_n.$$

Subtracting μ_{n+1}^- from the left hand side and μ_n^- from the right hand side, gives

$$\omega_{n+1} = \mu_{n+1}^+ - \mu_{n+1}^- \le \mu_n^+ - \mu_n^- - \eta\omega_n = (1-\eta)\omega_n.$$

A similar argument proves the claim if the second of (4.6) holds. □

The proof of this Proposition is in [39] and is a parabolic version of a similar *elliptic* Proposition proved by DeGiorgi [14].

5 Parabolic Equations with One-Point Singularity

We consider now (1.1) where $\beta(\cdot)$ is singular at *only one point*. This would include the Stefan graph indicated in the first of (**i**) and the porous medium graph indicated in the first of (**ii**).

The first regularity results for weak solutions of these equations with such a $\beta(\cdot)$, appear in [9,18,19,47,48,55]. In all these contributions, the basic approach to continuity is analogous to that of Propositions 4.1–4.2. The proofs differ essentially from the technical ways of establishing an alternative similar to that in Proposition 4.2. The singularity of $\beta(\cdot)$ affects Proposition 4.2 in two ways, i.e., the reduction factor δ that determines the sequence of radii ρ_n in (4.2), and the number η that determines the reduction of the oscillation in (4.3), are both functions of the oscillation itself. Given two continuous, monotone increasing functions

$$(0, 2M] \ni s \to \delta(s), \eta(s) \in (0,1), \quad \text{such that} \quad \delta(0) = \eta(0) = 0, \tag{5.1}$$

[14] Suppose the first of (4.4) holds and assume without loss of generality that (x_o, t_o) coincides with the origin of $\mathbb{R}^{N+1}$. Then the sup u over the smaller cylinder $Q_{\rho_{n+1}}$ is strictly less than the sup u over the larger and coaxial cylinder Q_{ρ_n}. Thus the sup u over the larger cylinder can only be achieved in the parabolic shell $Q_{\rho_n} \backslash Q_{\rho_{n+1}}$. This can be regarded as some sort of parabolic boundary for the larger box.

we construct inductively, the decreasing sequences of numbers

$$\begin{aligned} \omega_o &= \max\left\{2M\,;\,C\rho_o^\lambda\right\},\\ \rho_{n+1} &= \delta(\omega_n)\,\rho_n, \qquad\qquad \forall n\in\mathbb{N},\\ \omega_{n+1} &= \max\left\{\big(1-\eta(\omega_n)\big)\omega_n; C\rho_n^\lambda\right\} \end{aligned} \tag{5.2}$$

and the corresponding family of shrinking nested cylinders $\{(x_o,t_o)+Q_{\rho_n}\}$, with the same "vertex" at (x_o,t_o). Here $C>1$ and $\lambda\in(0,1)$ are two given constants.

Lemma 5.1. *$\{\omega_n\}\to 0$ as $n\to\infty$.*

Proof. From the definition it follows that the sequences $\{\rho_n\}$ and $\{\omega_n\}$ are non-increasing, so that their limits as $n\nearrow\infty$ exist. Since $\delta(\cdot)\in(0,1)$, it is apparent that $\{\rho_n\}\searrow 0$. If

$$\lim_{n\to\infty}\omega_n=\omega_\infty>0,$$

then, using the monotonicity of $\eta(\cdot)$, we derive from (5.2),

$$\omega_{n+1}\le\max\left\{\big(1-\eta(\omega_\infty)\big)\omega_n; C\delta^{\lambda n}(\omega_\infty)\rho^\lambda\right\},\quad \forall n\in\mathbb{N}.$$

Thus $\{\omega_n\}\searrow 0$ against the contradiction assumption. □

Proposition 5.1. *Let u be a weak solution of (1.1) with $\beta(\cdot)$ either of Stefan-type (i.e., the first of (**i**)) or of the type of porous media (i.e., the first of (**ii**)). Then there exist constants $C>1$ and $\lambda\in(0,1)$, and two continuous increasing functions $\delta(\cdot)$ and $\eta(\cdot)$ as in (5.2), that can be determined a priori only in terms of the data, such that for every $(x_o,t_o)\in\Omega_T$,*

$$\operatorname*{ess\,osc}_{\{(x_o,t_o)+Q_{\rho_n}\}} u\le\omega_n\qquad n=0,1,2,\dots\,. \tag{5.3}$$

Here $\lambda\in(0,1)$ is a number determined only in terms of the integrability conditions (1.5)–(1.7) and is independent of δ and η. As a consequence u is locally continuous in Ω_T.

Remark 5.1. The constants C and λ in (5.2), depend only upon the functions φ_i, $i=0,1,2$ in the structure conditions (1.4) and can be taken to be zero for the prototype equation (1.1′).

Remark 5.2. While Proposition 4.1 implies a precise Hölder modulus of continuity, this is not longer the case for Proposition 5.1. The sequences (5.2) and the recursive bound (5.3) supply a *quantitative* but *not explicit* modulus of continuity for u.[15]

[15] In the case of graphs of Stefan-type the functions $s\to\delta(s)$, $\eta(s)$ have the explicit form $K^{-h/s}$ where K and h are large constants (see [18]). It would be of interest to generate an explicit modulus of continuity for u, in terms of K and h.

The contributions in [9,18,19,47,48,55] all establish recursive inequalities similar to those of Proposition 5.1, even though with technically different points of view. In [18,19] the Proposition was established by means of DeGiorgi-type iterations, in the parabolic setting of [39]. The proof of [55] follows the Harnack-type techniques of Moser, as appearing in [43,44,5,52]. The results of [9] make use of local representations in terms of heat potentials, and for this reason are limited to the prototype equations (1.1′). The results of [47,48] are also limited to (1.1′), being based on the non-divergence structure *shrinking technique* of Krylov and Safonov [38].

Whatever the approach however, it is essential that $\beta(\cdot)$ be singular at *only one point*, say for example at $u = 0$.

All these proofs have a common pattern, i.e., having fixed a cylinder $\{(x_o, t_o) + Q_\rho\}$, either the singularity occupies a small portion of such a box or a large one. The first case is a *favorable*, in the sense that the singularity plays a negligible role. If the second case occurs, then since off the singular set the partial differential equation (1.1) is uniformly parabolic, the solution cannot grow too fast and remains "close" to a fixed value, for example μ^+, and it does not oscillate too much. This supplies a control on the oscillation which in turn can be rephrased as in Proposition 5.1.

Technically, the solution u remains "close" to μ^+ within $\{(x_o, t_o) + Q_\rho\}$, if the functions

$$(u-k)_+ \equiv \max\{u-k; 0\}, \qquad 0 < k < \mu^+,$$

are subsolutions of a *uniformly parabolic* equation. This in turn is possible if, for $u \geq k > 0$, the graph $\beta(\cdot)$ does not suffer any other singularity. By working with the infimum μ^- a similar argument indicates that $\beta(\cdot)$ cannot have a singularity for $u < 0$. Thus the only singularity permitted is at a single point. This is the main limitation of these proofs.

6 Power-Like One Point Singularity

Consider now (1.1) with $\beta(\cdot)$ given by the first of (**ii**). In such a case we rewrite the p.d.e. as

$$|u|^{\frac{1-m}{m}} u_t - \operatorname{div} \mathbf{A}(x, t, u, \nabla u) + B(x, t, u, \nabla u) = 0 \quad \text{in } \Omega_T. \tag{6.1}$$

If the coefficient of u_t were constant, one might perform a change of the time variable to transform (6.1) into a *non-singular*, uniformly parabolic equation. Following this remark, one might introduce an *intrinsic time scale* in the parabolic cylinders $\{(x_o, t_o) + Q_\rho\}$, with respect to which (6.1) would exhibits properties typical of uniformly parabolic equations. This idea has been introduced and implemented in [23]. The new *intrinsic geometry* is constructed as follows. For $\omega > 0$ let $Q_\rho(\omega)$ denote the cylindrical domain with "vertex" at the origin of $\mathbb{R}^{N+1}$,

$$Q_\rho(\omega) \equiv K_\rho \times \{-\rho^2 \omega^{\frac{1-m}{m}}, 0\}. \tag{6.2}$$

For $(x_o, t_o) \in \Omega_T$, we let $\{(x_o, t_o) + Q_\rho(\omega)\}$ denote the cylinder congruent to $Q_\rho(\omega)$ and with "vertex" at (x_o, t_o). Having fixed $(x_o, t_o) \in \Omega_T$ and $\omega > 0$, we will choose $\rho > 0$ so that $\{(x_o, t_o) + Q_\rho(\omega)\} \subset \Omega_T$. Let us fix two *constants* δ, $\eta \in \left(0, \frac{1}{2}\right)$ satisfying

$$\delta < (1-\eta)^{\frac{m-1}{m}}, \tag{6.3}$$

and construct, inductively, sequences $\{\omega_n\}$, $\{\rho_n\}$ and a family of nested and shrinking cylinders as follows.

$$\begin{aligned} &\omega_o = 2M, \rho_o \text{ such that } \{(x_o, t_o) + Q_{\rho_o}(\omega_o)\} \subset \Omega_T; \\ &\omega_{n+1} = (1-\eta)\omega_n + C\rho_n^\lambda, \quad \rho_{n+1} = \delta^n \rho_o, \quad \forall n \in \mathbb{N}; \\ &\{(x_o, t_o) + Q_{\rho_n}(\omega_n)\}. \end{aligned} \tag{6.4}$$

Here $C > 1$ and $\lambda \in (0,1)$ are fixed constants. These cylinders all have the same "vertex". Therefore, to verify that they are nested it suffices to verify that

$$\rho_{n+1}^2 \omega_{n+1}^{\frac{1-m}{m}} < \rho_n^2 \omega_n^{\frac{1-m}{m}}.$$

By making use of the definitions of ρ_n and ω_n, this is verified if (6.3) holds.

Proposition 6.1. *Let u be a weak solution of (1.1) with $\beta(\cdot)$ of the type of porous media (i.e., the first of (**ii**)). Then there exist constants δ, $\eta \in (0,1)$ that can be determined a priori only in terms of the data, such that for every $(x_o, t_o) \in \Omega_T$,*

$$\operatorname*{ess\,osc}_{\{(x_o,t_o)+Q_{\rho_n}(\omega_n)\}} u \le \omega_n \qquad n = 0, 1, 2, \dots. \tag{6.5}$$

Here $C > 1$ and $\lambda \in (0,1)$ are numbers determined only in terms of the integrability conditions (1.5)–(1.7) and are independent of δ and η. As a consequence u is locally Hölder continuous in Ω_T.

Remark 6.1. The Hölder modulus of continuity can be derived as in the proof of Proposition 4.1, since the "shrinking" numbers δ and η are independent of the solution.[16]

Remark 6.2. It is natural to ask whether the same idea of working with intrinsically rescaled cylinders could be used for graphs of the Stefan-type. In such a case $\beta'(\cdot)$ is the Dirac mass at the origin. As a consequence the time should be intrinsically rescaled into another which, loosely speaking, would remain constant on the transition set $[u = 0]$.[17] We do not know of a general technical way of operating such a rescaling. However in [20] we have devised a variant of it, in the context of the boundary regularity of weak solutions of (1.1).

[16] The same idea of introducing an intrinsic geometry, can be applied to doubly non linear parabolic equations, as long as the singularities and/or degeneracies are power-like. We refer to Ivanov [31,32], Porzio-Vespri [46] and Vespri [54] for the main points of the theory.

[17] Presumably, a technical implementation of this idea, if at all possible, would require some preliminary information on the relative size of the singular set $[u = 0]$.

6.1 Remarks on Boundary Regularity

Suppose that *continuous* Dirichlet data are prescribed on the lateral part

$$S_T \equiv \bigcup_{t\in(0,T]} \partial\Omega \times \{t\},$$

of the parabolic boundary of Ω_T. The boundary data on S_T are taken in the sense of the traces of the functions $u(\cdot,t) \in W^{1,1}(\Omega)$. In such a case we establish in [20] that weak solutions of (1.1) are continuous up to S_T, both for Stefan-type graphs and for graphs of the type of porous media. We introduce a time scale which becomes progressively small as as the essential oscillation of u decreases to zero. The method however could be implemented only because of the information contained in the boundary data.

7 Parabolic Equations with Multiple Singularities

Equations with $\beta(\cdot)$ exhibiting *multiple* singularities arise naturally from the flows of two immiscible fluids in a porous medium. The model example is (3.1) which, as indicated in §3, presents difficulties of similar nature as (1.1). The first attempt to establish the local continuity of solutions of (3.1) is in [1] under some assumption on the nature of the degeneracy of the function $\varphi(\cdot)$ introduced in (3.3), near *at least one* of the degeneracy points $v = 0$ and $v = 1$. For example $\varphi(\cdot)$ could degenerate at *any unrestricted rate* near $v = 1$, provided near $v = 0$ it degenerates no faster than logarithmically. The result was improved in [21] by allowing the second degeneracy to be power-like, with no restriction on the power. This last work employs a "one sided" intrinsic geometry, of the type discussed in §6, by introducing, roughly speaking, two parabolic scales. When working near the unrestricted degeneracy, say for example $v = 1$, we employ the standard parabolic cylinders $\{(x_o,t_o)+Q_\rho\}$, as in (4.1). This is because, due to the lack of information on the nature of the degeneracy, no natural rescaling is available. When working near $v = 0$, if the degeneracy is power-like, we work with cylinders coaxial with $\{(x_o,t_o)+Q_\rho\}$, with the same "vertex" at (x_o,t_o) and whose time scale is of the form (6.2).

Because on the restriction placed on the degeneracy of $\varphi(\cdot)$, both contributions [1,21] leave open the main issue of local continuity of solutions, as outlined in §3. The restriction imposed in [1,21], are used to exploit the parabolic nature of (1.1) on some side of a point of singularity of $\beta(\cdot)$.

However for *general* graphs $\beta(\cdot)$, the equation in (1.1) is not uniformly parabolic on either side of a singular point. For this reason, any continuity result for weak solutions of (1.1) with *general* $\beta(\cdot)$, would require a "non-parabolic" approach.

The first approach in this direction appears in [25], where the role played by (1.1) is reduced essentially to some energy estimates and a major role is

played instead, by some novel measure-theoretical facts.[18] The results of [25] are optimal in space-dimension $N = 2$. For $N \geq 3$ they are still not complete. We will state these results and point out the main open problems regarding $N \geq 3$.

Theorem 7.1 ($N = 2$). *Let u be a locally bounded weak solution to (1.1), in the sense of (1.8)–(1.11), where $\beta(\cdot)$ is any maximal monotone graph satisfying the coercivity and boundedness conditions (1.2)–(1.3). Assume moreover that $N = 2$ and that the structure conditions (1.4)–(1.7) are satisfied for $N = 2$. Then u is locally continuous in Ω_T. Moreover, for every compact subset $\mathcal{K} \subset \Omega_T$, there exists a continuous, non-negative, increasing function*

$$s \longrightarrow \omega_{\text{data}}(s), \qquad \omega_{\text{data}}(0) = 0, \tag{7.1}$$

that can be determined a priori only in terms of the data and the distance from $\mathcal{K}$ to the parabolic boundary of Ω_T, such that

$$|u(x_1, t_1) - u(x_2, t_2)| \leq \omega_{\text{data}} \left(|x_1 - x_2| + |t_1 - t_2|^{\frac{1}{2}} \right), \tag{7.2}$$

for every pair of points $(x_i, t_i) \in \mathcal{K}$, $i = 1, 2$.

Remark 7.1. The result is optimal in that no restrictions are placed on the singularities of $\beta(\cdot)$, and the parabolic equations is permitted to bear the full quasilinear structure (1.1)–(1.7). For $N \geq 3$ on the other hand, while no restrictions are placed on $\beta(\cdot)$, the p.d.e. in (1.1) is required to have a limited structure.

Theorem 7.1 ($N \geq 3$). *Let u be a locally bounded weak solution to (1.1′), where $\beta(\cdot)$ is any maximal monotone graph satisfying the coercivity and boundedness conditions (1.2)–(1.3). Then u is locally continuous in Ω_T, with a modulus of continuity that can be determined quantitatively, a priori only in terms of the data as in (7.1)–(7.2).*

8 Main Ideas of the Proof

In outlining the main points of the proof, we let u be a weak solution of (1.1) with the full quasilinear structure (1.1)–(1.7) in any number of dimensions, and will point out later the differences between $N = 2$ and $N \geq 3$. To establish Proposition 5.1, we fix $(x_o, t_o) \in \Omega_T$ and assume, after a translation, that it coincides with the origin. We will work with the cubes K_ρ and the cylinders Q_ρ introduced in (4.1). The numbers $\mu^\pm$ and ω are defined as in Section 4.1.

Proposition 8.1. *Let $\delta \in \left(0, \frac{1}{4}\right)$ be a parameter to be chosen and assume that there exists a time level $\tilde{t} \in (-\rho^2, -\delta^2\rho^2)$, such that*

$$u\left(x, \tilde{t}\right) \leq \mu^+ - \tfrac{1}{4}\omega, \qquad \forall x \in K_{2\delta\rho}. \tag{8.1$^+$}$$

[18] The main one these is stated in Section 11 and is independent of partial differential equations. For this reason we feel that it might be applicable to other branches of Analysis.

Then there exist numbers $\eta \in (0,1)$ and $C > 1$, $\lambda \in (0,1)$, depending upon the data and δ, but independent of ω and ρ, such that either $\omega \le C\rho^{\lambda}$ or,

$$u(x,t) \le \mu^{+} - \eta\omega \qquad \forall (x,t) \in Q_{\delta\rho} \equiv K_{\delta\rho} \times \left(-\delta^2\rho^2, 0\right). \tag{8.2^{+}}$$

Likewise if for some $\tilde{t} \in (-\rho^2, -\delta^2\rho^2)$, there holds,

$$u\left(x,\tilde{t}\right) \ge \mu^{-} + \tfrac{1}{4}\omega, \qquad \forall x \in K_{2\delta\rho}, \tag{8.1^{-}}$$

then either $\omega \le C\rho^{\lambda}$, or

$$u(x,t) \ge \mu^{-} + \eta\omega, \qquad \forall (x,t) \in Q_{\delta\rho} \equiv K_{\delta\rho} \times \left(-\delta^2\rho^2, 0\right), \tag{8.2^{-}}$$

for the same constants η, C, λ.

The constants $C > 1$ and $\lambda \in (0,1)$ depend only on the various parameters appearing in the structure conditions (1.5)–(1.7) and are independent of ω and the singularities of $\beta(\cdot)$. From now on we will consider them fixed.

As indicated in the proof of Proposition 4.2, either one of (8.2^{+}), (8.2^{-}) implies that going down from Q_ρ to the smaller cylinder $Q_{\delta\rho}$, the oscillation of u decreases of a factor $(1-\eta)$.

The proof of Proposition 8.1 hinges upon recursive inequalities based on the logarithmic estimates introduced in [18]. Due to the "initial conditions" (8.1^{+}), (8.1^{-}) these logarithmic estimates are analogous to those one would derive for solutions of non-singular equations.[19] Another feature of Proposition 8.1 is that the number η depends upon δ but not upon the oscillation ω. This is precisely the parameter dependence of Proposition 4.1.

Thus, the starting point of the proof is that if one had some information, such as (8.1^{+}), (8.1^{-}) on the status of the system at some "initial" time $t = \tilde{t}$, then the p.d.e. in (1.1) would behave like a quasilinear non-singular parabolic equation.

To achieve an information of the type (8.1^{+}), (8.1^{-}) we consider cylinders, coaxial with Q_ρ, with "vertex" at $(0,\tilde{t})$ and congruent to $Q_{4\delta\rho}$, i.e.,

$$\left\{(0,\tilde{t}) + Q_{4\delta\rho}\right\} \equiv K_{4\delta\rho} \times \left\{\tilde{t} - (4\delta\rho)^2\ ,\ \tilde{t}\right\}.$$

As the time level $\tilde{t}$ ranges over

$$\left\{-\left(1 - 16\delta^2\right)\rho^2\ ,\ -16\delta^2\,\rho^2\right\}, \tag{8.3}$$

the cylinders $\{(0,\tilde{t}) + Q_{4\delta\rho}\}$, move inside Q_ρ remaining coaxial with it. By moving them in the indicated range, we seek to locate some position of $\tilde{t}$ where one could derive some "initial" information of the type of (8.1^{+}), (8.1^{-}). Precisely, we will look for those positions of $\tilde{t}$, where the subset of $\{(0,\tilde{t}) + Q_{4\delta\rho}\}$ where u is

[19] The proof of Proposition 8.1 results from combining the logarithmic estimates of §4 of [25] with Propositions 3.2$^{\pm}$. We refer to [25] for full proofs.

close either to μ^+ or μ^- is small, i.e., for which either one of the following two inequalities holds,

$$\begin{aligned}&\text{meas}\left\{(x,t)\in\{(0,\tilde t)+Q_{4\delta\rho}\}\;\Big|\;u(x,t)\ge\mu^+-\tfrac12\omega\right\}\le\nu\,|Q_{4\delta\rho}|\,;\\&\text{meas}\left\{(x,t)\in\{(0,\tilde t)+Q_{4\delta\rho}\}\;\Big|\;u(x,t)\ge\mu^-+\tfrac12\omega\right\}\le\nu\,|Q_{4\delta\rho}|\,,\end{aligned}\tag{8.4$^\pm$}$$

for some $\nu\in(0,1)$ to be determined in terms of the data.

Proposition 8.2. *There exists a number $\nu\in(0,1)$, that can be determined a priori only in terms of the data and ω, such that if (8.4$^+$) holds for some $\tilde t$ in the range (8.3), then either $\omega\le C\rho^\lambda$ or,*

$$u(x,t)\le\mu^+-\tfrac14\omega\qquad\forall(x,t)\in\{(0,\tilde t)+Q_{2\delta\rho}\}.\tag{8.5$^+$}$$

Analogously, if (8.4$^-$) holds for some $\tilde t$, then either $\omega\le C\rho^\lambda$ or,

$$u(x,t)\ge\mu^-+\tfrac14\omega\qquad\forall(x,t)\in\{(0,\tilde t)+Q_{2\delta\rho}\}.\tag{8.5$^-$}$$

The proof is based on iterative inequalities starting from energy estimates, similar to those one would obtain for quasilinear, *non-singular* equations. The singularity of $\beta(\cdot)$ contributes to these energy estimates with a large constant depending upon the data and ω. For this reason the number ν, in (8.4$^\pm$) has to be chosen to depend upon ω.[20]

A consequence of Proposition 8.2 is that if either one of (8.4$^\pm$) is verified for *some* time level $\tilde t$ in the indicated range, then at least one of (8.2$^+$), (8.2$^-$) would hold true, and the proof could be concluded as indicated in Proposition 4.2.

Therefore the unfavorable case is when both (8.4$^\pm$) are violated for *every* time level $\tilde t$ in the range (8.3). The parameter δ introduced in Proposition 8.1 is still to be chosen. We will choose it in such a way that if the unfavorable case occurs for all $\tilde t$ in the range (8.3) and for *arbitrarily small* valued of δ, then this would imply a contradiction.

Consider any one of the cylinders $\{(0,\tilde t)+Q_{4\delta\rho}\}$. If (8.4$^\pm$) are both violated for arbitrarily small δ, then near the axis of Q_ρ, at the time level $\tilde t$, there is a relatively large set where the solution u is close to μ and another relatively large set where u is close to μ^-. Since δ is arbitrarily small and $\tilde t$ is arbitrary in the range (8.3), these two sets are arbitrarily close to each other. Therefore the space gradient ∇u must be large on a relatively large set. Since however $\nabla u\in L^2(Q_\rho)$, this would create a contradiction.

9 Identifying Regions of Concentration of the Energy

The technical implementation of this idea requires that we locate those regions within $\{(0,\tilde t)+Q_{4\delta\rho}\}$ where the energy is sufficiently large. For this we identify

[20] Proposition 8.2 corresponds to Propositions 3.1$^\pm$ of [25] to which we refer for a full proof and further details.

two sub-cylinders

$$\{(y_i, \tilde{t}) + Q_{\delta^2\rho}\} \subset \{(0, \tilde{t}) + Q_{4\delta\rho}\}, \quad i = 1, 2, \tag{9.1}$$

where

$$\begin{aligned} u(x,t) &\geq \tfrac{1}{4}\mu^+ \qquad \forall (x,t) \in \{(x_1, \tilde{t}) + Q_{\delta^2\rho}\}; \\ u(x,t) &\leq \tfrac{1}{4}\mu^- \qquad \forall (x,t) \in \{(x_2, \tilde{t}) + Q_{\delta^2\rho}\}. \end{aligned} \tag{9.2}$$

At first these two cylinders are found within $\{(0, \tilde{t}) + Q_{4\delta\rho}\}$. Then by using the arbitrariness of $\tilde{t}$ we identify them has having their "vertices" at $(y_i, \tilde{t})$, i.e., at the same time level $\tilde{t}$. Also, using the arbitrariness of δ we may insure that their cross sections are mutually separated by a distance of at least $\delta^2\rho$.[21]

It is in this process that the new Lemma on measure theory plays a role. Assume for the moment that (8.4$^+$) is violated so that the set where u is close to μ^+ is relatively large. The Lemma asserts that u must be close to μ^+ in some sufficiently small box within $\{(0, \tilde{t}) + Q_{4\delta\rho}\}$, i.e., the set where u is close to μ^+, even though it might be scattered in $\{(0, \tilde{t}) + Q_{4\delta\rho}\}$, it must have, loosely speaking, some *concentration regions* within it. We will state and discuss this Lemma in Section 11. Here we observe that (9.2) implies

$$\tfrac{1}{4}\omega \leq u(x_1, t) - u(x_2, t) \quad \forall x_i \in \left\{y_i + K_{\delta^2\rho}\right\}, \quad i = 1, 2, \tag{9.3}$$

for all the time levels

$$t \in \left(\tilde{t} - \delta^4\rho^2, \tilde{t}\right). \tag{9.3'}$$

For t fixed in the indicated range, we first integrate (9.3) over a path, piecewise parallel to the coordinate axes and joining

$$x_1 \in \left\{y_1 + K_{\delta^2\rho}\right\} \quad \text{and} \quad x_2 \in \left\{y_2 + K_{\delta^2\rho}\right\}.$$

Then using the arbitrariness of these points within their ranges, we integrate the resulting segment-integrals, over the remaining $(N-1)$ variables, and then over the time in the range (9.3′). This yields

$$\gamma(\omega)\ (\delta\rho)^N \leq \int_{\tilde{t}-\delta^2\rho^2}^{\tilde{t}} \int_{K_{\delta\rho} \setminus K_{\delta^2\rho}} |\nabla u|^2 dx d\tau, \tag{9.4}$$

where $\gamma(\omega)$ is a constant depending upon the data and ω. This inequality has been derived for all $\tilde{t}$ in the range (8.3) for which (8.4$^\pm$) are *both* violated. We observe that, for $\tilde{t}$ in such a range, the number of disjoint cylinders of the type $\{(0, \tilde{t}) + Q_{4\delta\rho}\}$ is of the order of δ^{-2}. Thus adding (9.4) over the corresponding boxes, gives

$$\gamma(\omega)\, \delta^{N-2}\rho^N \leq \int_{-\rho^2}^{0} \int_{K_{\delta\rho} \setminus K_{\delta^2\rho}} |\nabla u|^2 dx d\tau, \tag{9.4'}$$

[21] The proof of these assertions is in §5–8 of [25].

The argument can now be repeated with δ replaced by δ^2, since δ can be chosen to be arbitrarily small. Therefore we conclude that for all $n \in \mathbb{N}$,

$$\gamma(\omega)\delta^{n(N-2)}\rho^N \le \int_{-\rho^2}^{0}\int_{K_{\delta^n\rho}\setminus K_{\delta^{n+1}\rho}} |\nabla u|^2 dxd\tau, \tag{9.4$_n$}$$

On the other hand, a standard energy estimate, gives

$$\iint_{Q_\rho} |\nabla u|^2 dxd\tau \le (\text{const})\rho^N. \tag{9.5}$$

We seek to derive a contradiction by iterating and adding (9.4$_n$) and comparing the resulting integral with (9.5).[22]

9.1 The case $N = 2$

Adding (9.4$_n$) for $n = 1, 2, \dots, n_o$, and taking into account (9.5) implies that,

$$\gamma(\omega)\, n_o \le (\text{const}).$$

This is a contradiction if n_o is sufficiently large depending on the data and ω. It follows that at least one of (8.4$^{\pm}$) must hold for $\tilde{t}$ in the range (8.3) and for some radius $\rho_o \in [\rho, \delta^{n_o}\rho]$. In view of Propositions 8.2 and 8.1 this would imply the result.

Remark 9.1. The same argument could be applied whenever one has information that essentially reduce the space dimension N to 1 or 2. This for example would occur for radial solutions of (1.1).

10 The Case $N \ge 3$

The key observation here is that, even though the previous argument fails if $N > 2$, an information of the type of (9.4) continues to hold within any sub-cylinder of Q_ρ not necessarily coaxial with it. With the number δ to be chosen, we assume, without loss of generality that $(4\delta)^{-1}$ is an integer, say for example m, and partition the original cube K_ρ, up to a set of measure zero, into m^N pairwise disjoint sub-cubes of wedge $(8\delta\rho)$ and centered at points $x_\ell \in K_\rho$, i.e.,

$$\begin{aligned}
&\{x_\ell + K_{4\delta\rho}\} \subset K_\rho, \qquad\qquad \ell = 1, 2, \dots, m^N;\\
&\{x_\ell + K_{4\delta\rho}\} \cap \{x_j + K_{4\delta\rho}\} = \emptyset \quad \text{if } \ell \ne j;\\
&K_\rho = \bigcup_{\ell=1}^{m^N} \{x_\ell + K_{4\delta\rho}\}.
\end{aligned}$$

[22] These inequalities are proved in Sections 9–12 of [25].

Then we partition the original cylinder Q_ρ, up to a set of measure zero, into m^{N+2} pairwise disjoint sub-cylinders with vertices at (x_ℓ, t_h) and all congruent to $Q_{4\delta\rho}$, i.e.,

$$\begin{aligned}
&\{(x_\ell, t_h) + Q_{4\delta\rho}\} \qquad\qquad \ell = 1, 2, \dots, m^N; h = 1, 2, \dots, m^2;\\
&\{(x_\ell, t_h) + Q_{4\delta\rho}\} \cap \{(x_j, t_k) + Q_{4\delta\rho}\} = \emptyset \text{ if } \ell \neq j \text{ or } h \neq k;\\
&Q_\rho = \bigcup_{\ell=1}^{m^N} \bigcup_{h=1}^{m^2} \{(x_\ell, t_h) + Q_{4\delta\rho}\}.
\end{aligned} \tag{10.1}$$

Returning to ($8.4^{\pm}$), we claim that at least one of them must be satisfied, for at least one of the cylinders $\{(x_\ell, t_h) + Q_{4\delta\rho}\}$ making up the partition of Q_ρ. Indeed if *both* of ($8.4^{\pm}$) are violated for *all* these cylinders, then inequality (9.4) must hold for all of them. We rewrite such inequalities in a sightly different form, i.e.,

$$\gamma(\omega)\ (\delta\rho)^N \leq \int_{t_h - \delta^2\rho^2}^{t_h} \int_{\{x_\ell + K_{\delta\rho}\}} |\nabla u|^2 dx d\tau, \qquad \begin{matrix} \ell = 1, 2, \dots, m^N; \\ h = 1, 2, \dots, m^2. \end{matrix}$$

Adding these inequalities over the indicated indices and taking into account (9.5) gives,

$$\gamma(\omega)\, m^2 (m\delta)^N \leq (\text{const}) \quad \Longrightarrow \quad \delta^{-2} \leq \gamma_{\text{data}}(\omega).$$

This is a contradiction for δ sufficiently small, depending on ω. It follows that at least one of ($8.4^{\pm}$) must hold for at least one of the cylinders (10.1) making up the partition of Q_ρ. Suppose for example that (8.4^{-}) holds true for the cylinder $\{(x_\ell, t_h) + Q_{4\delta\rho}\}$. Then, by Proposition 8.2,

$$u(x,t) \geq \mu^- + \tfrac{1}{4}\omega \qquad \forall (x,t) \in \{(x_\ell, t_h) + Q_{2\delta\rho}\}. \tag{$8.5^-_{(\ell,h)}$}$$

If $x_\ell \equiv 0$, then the cylinder $\{(x_\ell, t_h) + Q_{4\delta\rho}\}$ would be coaxial with Q_ρ, and the proof could be concluded as indicated in Proposition 8.1. Thus the main point of the proof for $N \geq 3$ is to establish that a version of ($8.5^-_{(\ell,h)}$) actually holds for a cylinder coaxial with Q_ρ. Alternatively we seek to establish that some bound below for u, within a region, would yield a bound below in a larger region. Estimates of this kind are typical of solutions of quasilinear parabolic equations and are contained for example in [44,38,52]. The difficulty here is the presence of the singularity of $\beta(\cdot)$.

In our proof such *space propagation of a bound below*, is technically realized by means of a suitable comparison function. To construct such a comparison function as well as to make full use of the comparison principle, the p.d.e. in (1.1) is required to have the restricted structure (1.1′).

10.1 Open Problems

We omit here the presentation of such a construction as we feel that the *space extension of positivity* should hold for equations with the full quasilinear structure (1.4) and it should be independent of the comparison principle. What seems

to be missing is some sort of weak form of the Harnack inequality ([44,52]), for solutions of *singular* parabolic equations.

We feel that an understanding of this point would permit one also to establish a, still missing, regularity Theorem up to the parabolic boundary of Ω_T.

11 A Lemma of Measure Theory

For $r > 0$ let K_r be a cube of wedge $2r$ and centered at the origin, as in (4.1). Let $v \in W^{1,p}(K_r)$, $p > 1$ satisfy

$$\int_{K_r} |\nabla v|^p dx \le \gamma\, r^{N-p}. \tag{11.1}$$

Inequalities of this type are satisfied by harmonic functions in a domain Ω containing K_r, or more generally by solutions of quasilinear elliptic equations in divergence form.

Lemma 11.1. *Suppose that for some $\alpha \in (0,1)$ there holds*

$$\text{meas}\left\{x \in K_r \;\middle|\; v(x) < 1\right\} \ge \alpha |K_r|. \tag{11.2}$$

Then for every $\varepsilon \in (0,1)$ and $\theta > 1$, there exists some $x^* \in K_r$ and a number $\delta \in (0,1)$ that can be determined a priori only in terms of N, α, ε, θ, such that

$$\text{meas}\left\{x \in \{x^* + K_{\delta r}\} \;\middle|\; v(x) < \theta\right\} \ge (1-\varepsilon)|K_{\delta r}|. \tag{11.3}$$

If v were continuous in K_r, then by the Theorem of the permanence of positivity, the Lemma would be trivial. However a function $v \in W^{1,p}(K_r)$, has some regularity. Thus the Lemma can be regarded as some sort of permanence of positivity for functions in $W^{1,p}(K_r)$. It asserts that if the set where $(v-1)$ is negative, is quantitatively non negligible, then the set where $(v-\theta)$ is negative, mighty be partly scattered within K_r, provided some of it is concentrated within a full cube $\{x^* + K_{\delta r}\}$.

11.1 An Open Question

The proof is independent of (1.1) or any partial differential equations and makes only use of measure-theoretical arguments, starting from (11.1). The number δ deteriorates as either $\varepsilon \searrow 0$ or $\theta \searrow 1$.

The proof also uses in an essential way that $v \in W^{1,p}(K_r)$ for $p > 1$. It would be of interest to investigate it when $p = 1$.

11.2 Use of the Lemma in the Context of (1.1)

The Lemma is applied to a solution u of (1.1) in the following manner. Suppose for example that (8.4$^-$) is violated for some $\tilde{t}$ in the range (8.3). Then for some time level

$$t \in (\tilde{t} - 16r^2, \tilde{t}), \qquad \text{where} \quad r = \delta\rho, \tag{11.4}$$

there holds,

$$\text{meas}\left\{x \in K_r \;\middle|\; u(x,t) > \mu^- + \tfrac{1}{2}\omega\right\} > \nu|K_r|. \tag{11.5}$$

By setting

$$v(x,t) = \frac{2\{u(x,t) - \mu^-\}}{\omega},$$

we rewrite (11.5) as

$$\text{meas}\left\{x \in K_r \;\middle|\; v(x,t) < 1\right\} \geq \nu|K_r|. \tag{11.2$_t$}$$

Then, by possibly modifying the positive number ν into a new quantifiable positive number $\alpha \in (0,\nu)$, we establish the existence of a time τ in the range (11.4) such that the following two inequalities both hold,

$$\begin{aligned}
&\text{meas}\left\{x \in K_r \;\middle|\; v(x,t) < 1\right\} \geq \alpha|K_r|,\\
&\int_{K_r} |\nabla v(x,\tau)|^2 dx \leq \gamma_{\text{data}}(\omega)\, r^{N-2},
\end{aligned}$$

for a constant $\gamma_{\text{data}}(\omega)$ depending only upon the data and ω, and independent of τ. Therefore by Lemma 11.1, having fixed $\varepsilon \in (0,1)$ and $\theta = \frac{3}{2}$ there exists a number $\delta \in (0,1)$ and a cube $\{x^* + K_{\delta r}\} \subset K_r$, such that

$$\text{meas}\left\{x \in \{x^* + K_{\delta r}\} \;\middle|\; u(x,\tau) < \mu^- + \frac{1+\sigma}{2}\,\omega\right\} \geq (1-\varepsilon)|K_{\delta r}|.$$

It follows that one has

$$\begin{aligned}
u(x,\tau) &\leq \mu^- + \tfrac{3}{4}\omega\\
&= \mu^+ - \mu^+ + \tfrac{3}{4}(\mu^+ - \mu^-)\\
&= \mu^+ - \tfrac{1}{4}\omega,
\end{aligned} \tag{11.6}$$

everywhere in $\{x^* + K_{\delta r}\}$ except at most a set of measure less than $\varepsilon|K_{\delta r}|$. The information in (11.6) is similar to (8.1$^+$) where some information is available at some "initial" time $\tilde{t}$. Here the time is τ and the information is not as complete since out of $\{x^* + K_{\delta r}\}$ one has to remove a set of measure less than $\varepsilon|K_{\delta r}|$. However since $\varepsilon \in (0,1)$ is arbitrary, we establish that ε can be chosen so small that (11.6) is sufficient to apply a version of Proposition 8.1.

References

1. H. W. Alt and E. DiBenedetto, Nonsteady Flow of Water and Oil Through Inhomogeneous Porous Media, *Ann. Scuola Norm. Sup. Pisa*, Classe di Sc. Ser. IV, **XII** (3) (1985), 335–392.
2. H. W. Alt and S. Luckhaus, Quasilinear Elliptic-Parabolic Differential Equations, *Math. Z.*, **183** (1983), 311–341.

3. S. N. Antoncev, On the Solvability of Boundary Value Problems for Degenerate Two-Phase Porous Flow Equations, *Dynamika Splosnoi Sredy Vyp.*, **10** (1972), 28–53 (Russian).
4. S. N. Antoncev, A. V. Kazhikov and V. N. Monakov, *Boundary-Value Problems in the Mechanics of Nonuniform Fluids*, Studies in Mathematics and its Applications, Amsterdam, 1990.
5. D. G. Aronson and J. Serrin, Local Behavior of Solutions of Quasilinear Parabolic Equations, *Arch. Rat. Mech. Anal.* **25** (1967), 81–123.
6. J. Bear, *Dynamics of Fluids in Porous Media*, Amer. Elsevier, New York, 1972.
7. J. Bear, *Hydraulics of Groundwater*, Mc-Graw Hill, New York, 1979.
8. A. Bossavit, A. Damlamian and M. Fremond Eds., *Free Boundary Problems: Applications and Theory*, Research Notes in Mathematics, **118, 119, 120**, London 1985.
9. L. Caffarelli and L. E. Evans, Continuity of the Temperature in the two Phase Stefan Problem, *Arch. Rat. Mech. Anal.* **81** (1983), 199–220.
10. L. Caffarelli and A. Friedman, Continuity of the Density of a Gas Flow in a Porous Medium, *Trans. Amer. Math. Soc.*
11. G. Chavent and J. Jaffré, *Mathematical Models and Finite Elements Methods for Reservoir Simulation*, North-Holland, Amsterdam, 1986.
12. R. E. Collins, *Flow of Fluids Through Porous Materials*, Reinhold, New York, 1961.
13. I. I. Danilyuk, The Stefan Problem, *Russian Math. Surveys*, **40** (5), (1985), 157–223.
14. E. DeGiorgi, Sulla Differenziabilità e l'Analiticità delle Estremali degli Integrali Multipli Regolari, *Mem. Accad. Sc. Torino,* Cl. Sc. Fis. Mat. Nat., **3** (3), (1957), 25–43.
15. J. van Duijn and Zhang Hongfei, Regularity Properties of a Doubly Degenerate Equation in Hydrology, *Comm. Part. Diff. Equ.* **13**, (1988), 261–319.
16. J. Douglas Jr., Finite Differences Methods for Two-Phase Incompressible Flow in Porous Media, *SIAM J. Numer. Anal.* **20** (1983), 681–696.
17. J. Douglas Jr., B. L. Darlow, M. F. Wheeler and R. P. Kendall, Self-Adaptive Finite Element and Finite Difference Methods for Two-Phase Immiscible Flow.
18. E. DiBenedetto, Continuity of Weak Solutions to Certain Singular Parabolic Equations, *Ann. Mat. Pura ed Appl.* (IV) **CXXX** (1982), 131–177.
19. E. DiBenedetto, Continuity of Weak Solutions to a General Porous Medium Equation, *Indiana Univ. Math. Jour.,* **32** (1) (1983), 83–118.
20. E. DiBenedetto, A Boundary Modulus of Continuity for a Class of Singular Parabolic Equations, *J. of Diff. Equ.,* **63** (3), (1986), 418–447.
21. E. DiBenedetto, The Flow of Two Immiscible Fluids Through a Porous Medium: Regularity of the Saturation, *Theory and Appl. of Liquid Crystals*, Eds. J. L. Ericksen and D. Kinderlehrer, Springer-Verlag, New York 1987, 123–141.
22. E. DiBenedetto, *Degenerate Parabolic Equations*, Springer Verlag, New York 1993.
23. E. DiBenedetto and A. Friedman, Hölder Estimates for Nonlinear Degenerate Parabolic Systems, *Jour. für die Reine und Angew. Math.* **357**, (1985), 1–22.
24. E. DiBenedetto and R. Gariepy, On the Local Behavior of Solutions for an Elliptic-Parabolic Equation, *Arch. Rat. Mech. Anal.* **97** (1) (1987), 1–17.
25. E. DiBenedetto and V. Vespri, On the Singular Equation $\beta(u)_t = \Delta u$, *Arch. Rat. Mech. Anal.* **132** (1995), 247–309.
26. A. Fasano and M. Primicerio Eds., *Free Boundary Problems: Theory and Applications*, Research Notes in Mathematics, **78, 79**, Pitman, London 1983.
27. A. Friedman, *Variational Principles and Free Boundary Problems*, Wiley-Interscience, New York, 1982.

28. J. Glimm, D. Marchesin and O. McBryan, Unstable Fingers in Two Phase Flow, *Comm. Pure Appl. Math.* **34** (1981), 53–76.
29. D. Hoff, A Scheme for Computing Solutions and Interface Curves for a Doubly Degenerate Parabolic Equation, *SIAM J. Numer. Anal.* **22** (4), (1985), 687–712.
30. K. H. Hoffmann and J. Sprekels Eds., Free Boundary Problems: Theory and Applications, Pitman Research Notes in Mathematics, **185, 186**, Longman Scientific & Technical, New York, 1990.
31. A. V. Ivanov, Uniform Hölder Estimates for Generalized Solutions of Quasilinear Parabolic Equations Admitting a Double Degeneracy, *Algebra Anal.* **3** (2), (1991), 139–179.
32. A. V. Ivanov, The Classes $\mathcal{B}_{m\ell}$ and Hölder Estimates for Quasilinear Doubly Degenerate Parabolic Equations, *Zap. Nauchn. Sem. St. Petersburg Otdel. Math. Inst. Steklov (LOMI)* **197** (1991), 3–28.
33. D. Kröner and S. Luckhaus, Flow of Oil and Water in a Porous Medium *J. Diff. Equ.,* **55** (1984), 276–288.
34. S. N. Kruzkov, On the a Priori Estimation of Solutions of Linear Parabolic Equations and of Solutions of Boundary Value Problems for a Certain Class of Quasi-Linear Parabolic Equations, *Dokl. Akad. NAUK SSSR* **138** (1961), 1005–1008; (Engl. Transl. in Soviet Math. Doklady **2** (1961), 764–767).
35. S. N. Kruzkov, A Priori Estimates and Certain Properties of the Solutions of Elliptic and Parabolic Equations of Second Order, *Mat. Sbornik* **65** (107), (1968), 522–570; (Engl. Transl. Amer. Math. Soc. Transl. **2** (68), (1968), 169–220).
36. S. N. Kruzkov, Results Concerning the Nature of the Continuity of Solutions of Parabolic Equations and Some of Their Applications, *Math. Zametki* **6**, (1969), 97–108 (Russian).
37. S. N. Kruzkov and S. M. Sukorjanski, Boundary Value Problems for Systems of Equations of two Phase Porous Flow Type: Statement of the Problems, Questions of Solvability, Justification of Approximate Methods, *Mat. Sbornik,* **44** (1977), 62–80.
38. N. V. Krylov and M. V. Safonov, A Certain Property of Solutions of Parabolic Equations with Measurable Coefficients, *Math. USSR Izvestija* **16** (1) (1981), 151–164.
39. O. A. Ladyzenskaja, V. S. Solonnikov and N. N. Ural'ceva, *Linear and Quasilinear Equations of Parabolic Type*, Amer. Math. Soc., Providence RI, 1968.
40. M. C. Leverett, Capillary Behavior in Porous Solids, *Trans. Amer. Inst. Mining and Metallurgical Engrs.,* **142** (1941), 151–169.
41. J. L. Lions, *Quelques Méthodes de Résolution des Problèmes aux Limites non Linéaires*, Dunod, Gauthiers-Villars, Paris, 1969.
42. A. Meirmanov, *The Stefan Problem*, Walter de Gruyter, Berlin, 1992.
43. J. Moser, A New Proof of DeGiorgi's Theorem Concerning the Regularity Problem for Elliptic Differential Equations, *Comm. Pure Appl. Math.* **13**, (1960), 457–468.
44. J. Moser, A Harnack Inequality for Parabolic Differential Equations, *Comm. Pure Appl. Math.* **17**, (1964), 101–134.
45. W. M. Ni, L. Peletier and J. L. Vazquez, Eds., *Degenerate Diffusion*, The IMA Volumes in Mathematics and its Appl. **47**, Springer-Verlag, New York, 1993.
46. M. M. Porzio and V. Vespri, Hölder Estimates for Local Solutions of Some Doubly Nonlinear Degenerate Parabolic Equations, *J. Diff. Equ.* **103**, (1993), 146-178.
47. P. Sachs, Continuity of Solutions of Singular Parabolic Equations, *Nonlinear Anal. TMA,* **7** (1983), 387–409.
48. P. Sachs, The Initial and Boundary Value Problem for a Class of Degenerate Parabolic Equations, *Comm. Part. Diff. Equ.* **8** (1983), 693–734.

49. A. E. Scheidegger, *The Physics of Flow Through Porous Media*, 3rd Ed., Univ. of Toronto Press, Toronto, Ontario, 1974.
50. A. Spivak, H. S. Price and A. Settari, Solution of the Equations for Multidimensional Two-Phase, Immiscible Flow by Variational Methods, *Soc. Petr. Eng. J.,* Feb. 1977, 27–41.
51. G. Stampacchia, *Equations Elliptiques du Second Ordre à Coefficients Discontinues*, Sem. Mth. Sup. **16**, Les Presses de l'Univ. de Montréal, Montréal, (1966).
52. N. S. Trudinger, Pointwise Estimates and Quasilinear Parabolic Equations, *Comm. Pure Appl. Math.* **21** (1968), 205–226.
53. V. Vespri, L^∞ Estimates for Non-Linear Parabolic Equations with natural Growth Conditions, *Rend. Sem. Mat. Univ. Padova,* **90** (1993), 1–8.
54. V. Vespri, On the Local Behavior of a Certain Class of Doubly non Linear Parabolic Equations, *Manuscripta Math.* **75**, (1992), 65–80.
55. W. P. Ziemer, Interior and Boundary Continuity of Weak Solutions of Degenerate Parabolic Equations, *Trans. Amer. Math. Soc.* **271** (1982), 773–748.

EQUADIFF 9 CD ROM, Brno 1997 PROCEEDINGS
Masaryk University pp. 49–61

Transformations and Oscillatory Properties of Linear Hamiltonian Systems — Continuous versus Discret

Ondřej Došlý

Department of Mathematics, Masaryk University, Janáčkovo nám. 2a,
662 95 Brno, Czech Republic
Email: dosly@math.muni.cz

Abstract. Transformations and oscillatory properties of discrete and continuous linear Hamiltonian systems are investigated. A particular attention is devoted to the so-called reciprocity principle and trigonometric transformation for these systems.

AMS Subject Classification. 34C10, 39A10

Keywords. Linear Hamiltonian system, trigonometric tranformation, reciprocity principle, symplectic system, principal solution, time scale

1 Introduction

The aim of this contribution is to present a survey of the recent results on transformations and oscillatory behaviour of solutions of linear Hamiltonian systems — both differential and difference – and to suggest some directions for the further investigation.

We consider the differential Hamiltonian system

$$x' = A(t)x + B(t)u, \quad u' = C(t)x - A^T(t)u, \tag{1.1}$$

and its difference (= discrete) counterpart

$$\Delta x_k = A_k x_{k+1} + B_k u_k, \quad \Delta u_k = C_k x_{k+1} - A_k^T u_k. \tag{1.2}$$

We suppose that $t \in I \subseteq \mathbb{R}$, $k \in [0, N] \cap \mathbb{N}$, $N \in \mathbb{N}$, both in continuous and discrete case A, B, C are $n \times n$ matrices, B, C are symmetric, i. e. $B = B^T$, $C = C^T$. Moreover, in the continuous case we suppose that the matrix B is nonnegative definite and in the discrete case that the matrix $(I - A_k)$ is nonsingular, its inverse we denote by $\tilde{A}_k$.

Linear Hamiltonian systems cover a large variety of linear equations. For example, the even order, self-adjoint, differential equation

$$\sum_{\nu=0}^{n} (-1)^\nu \left(r_\nu(t) y^{(\nu)} \right)^{(\nu)} = 0 \tag{1.3}$$

can be written as LHS (1.1) using the substitution

$$x = \begin{pmatrix} y \\ y' \\ \vdots \\ y^{(n-1)} \end{pmatrix}, \quad u = \begin{pmatrix} (-1)^{n-1}(r_n y^{(n)})^{(n-1)} + \cdots + r_1 y' \\ \vdots \\ -(r_n y^{(n)})' + r_{n-1} y^{(n-1)} \\ r_n y^{(n)} \end{pmatrix}.$$

Then (x, u) solves (1.1) with A, B, C given by

$$B(t) = \operatorname{diag}\{0, \dots, 0, r_n^{-1}(t)\}, \quad C(t) = \operatorname{diag}\{r_0(t), \dots, r_{n-1}(t)\},$$
$$A = A_{ij} = \begin{cases} 1, & \text{if } j = i+1,\ i = 1, \dots, n-1, \\ 0, & \text{elsewhere.} \end{cases}$$

Another example is the second order system

$$\left[R(t)x' + Q(t)x\right]' - \left[Q^T(t)x' + P(t)x\right] = 0 \tag{1.4}$$

with $n \times n$ matrices P, Q, R, whereby P, R are symmetric and R nonsingular. Putting $u = R(t)x' + Q(t)x$, the pair of n-vectors (x, u) solves (1.1) with $B = R^{-1}$, $A = -R^{-1}Q$, $C = P - Q^T R^{-1} Q$. Discrete analogies of (1.3) and (1.4) can be written in the form (1.2) using essentially the same substitutions as in the continuous case.

Investigation of oscillatory properties of continuous system (1.1) has a relatively long history and was initiated by the paper of Morse [17] from 1930. Since that time oscillation theory of (1.1) attracted a considerable attention and the results of this investigation up to seventies of this century can be found in the monograph of Reid [18]. In 1995 Kratz [16] published another comprehensive monograph which in addition to the classical results contains also the results achieved in the period 1980–95.

In contrast to the continuous case, oscillation theory of discrete systems (1.2) is much less developed and the fundamental result of this theory, a discrete version of the so-called Roundabout Theorem, was established only very recently by Bohner [8]. This paper accomplished the effort of several mathematicians in the last decade to prove the discrete Roundabout Theorem in its full generality, see [4].

Here we concentrate our attention to the investigation of oscillatory properties and transformations of Hamiltonian systems (1.1) and (1.2). The paper is organized as follows. In the next section we present basic facts of oscillation and transformation theory of continuous systems (1.1), in particular, we formulate trigonometric transformation and reciprocity principle for these systems. Section 3 is devoted to some aspects of transformation theory of Hamiltonian difference systems (1.2) and we give here essentially discrete versions of statements of Section 2. In the last section we discuss some aspects of unified approach to continuous and discrete systems via theory of differential equations on the so-called time scales.

2 Continuous Hamiltonian Systems

We start with basic concepts of oscillation theory of (1.1).

Definition 1. Two points t_1, t_2 are said to be *conjugate* relative to (1.1) if there exists a solution (x, u) such that $x(t_1) = 0 = x(t_2)$ and $x(t) \not\equiv 0$ in $[t_1, t_2]$. System (1.1) is said to be *conjugate* in an interval $[a, b]$ if there exist $t_1, t_2 \in [a, b]$ which are conjugate relative to (1.1), in the opposite case (1.1) is said to be *disconjugate*. System (1.1) is said to be *oscillatory* if for every $c \in \mathbb{R}$ this system is conjugate in $[c, \infty)$, in the opposite case (1.1) is said to be *nonoscillatory*.

As mentioned in the previous section, principal statement concerning oscillatory properties of (1.1) is the so-called Reid Roundabout Theorem [18]. Before formulating it, we recall some very elementary properties of solutions of Hamiltonian systems (1.1).

Simultaneously with (1.1) we consider its matrix analogy

$$X' = A(t)X + B(t)U, \quad U' = C(t)X - A^T(t)U, \tag{2.1}$$

where X, U are $n \times n$ matrices. If $(X, U), (\tilde{X}, \tilde{U})$ are two solutions of (2.1) then the "Wronskian-type" identity $X^T\tilde{U} - U^T\tilde{X} \equiv K$ holds, where K is a constant $n \times n$ matrix. A solution (X, U) of (2.1) is said to be *conjoined* if X^TU is symmetric and it is said to be *conjoined basis* if, moreover, rank $(X^T, U^T) = n$. Recall also that (1.1) is said to be *controllable* in an interval I whenever the trivial solution $(x, u) \equiv (0, 0)$ is the only solution of (1.1) for which $x(t) \equiv 0$ on some nondegenerate subinterval $I_0 \subseteq I$.

Proposition 1 (Reid [18]). *Suppose that the matrix $B(t)$ is nonnegative in the interval $[a, b]$ and that (1.1) is controllable in this interval. Then the following statements are equivalent:*

(i) *System (1.1) is disconjugate in the interval $[a, b]$.*

(ii) *The quadratic functional*

$$\mathcal{F}(x, u) = \int_a^b [u^T(t)B(t)u(t) + x^T(t)C(t)x(t)]dt$$

is positive for every nontrivial (x, u) satisfying $x' = A(t)x + B(t)u$ and $x(a) = 0 = x(b)$.

(iii) *The solution (X, U) of (2.1) given by the initial condition $X(a) = 0$, $U(a) = I$ satisfies $\det X(t) \neq 0$, $t \in (a, b]$.*

(iv) *There exists a conjoined basis (X, U) of (2.1) such that $X(t)$ is nonsingular for $t \in [a, b]$.*

(v) *There exists a symmetric matrix Q which for $t \in [a, b]$ solves the Riccati matrix differential equation*

$$Q' - C(t) + A^T(t)Q + QA(t) + QB(t)Q = 0 \tag{2.2}$$

related to (2.1) by the substitution $Q = UX^{-1}$.

For a better understanding of this statement we suggest the reader to see (1.1) as a rewritten second order equation and to compare this statement with the well known results of oscillation theory of second order equations, see e.g. Swanson [20].

Now we state some results concerning transformations of LHS. A $2n \times 2n$ matrix $\mathcal{R}$ is said to be *symplectic* if $\mathcal{R}^T \mathcal{J} \mathcal{R} = \mathcal{J}$, where $\mathcal{J} = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$, I being the $n \times n$ identity matrix. If $\mathcal{R} = \begin{pmatrix} H & K \\ M & N \end{pmatrix}$, where H, K, M, N are $n \times n$ matrices then $\mathcal{R}$ is symplectic if and only if

$$H^T K = K^T H, \ M^T N = N^T M, \ H^T N - K^T M = I. \tag{2.3}$$

Consider the transformation

$$\begin{pmatrix} x \\ u \end{pmatrix} = \mathcal{R}(t) \begin{pmatrix} y \\ z \end{pmatrix}, \ \text{where } \mathcal{R}(t) = \begin{pmatrix} H(t) & M(t) \\ K(t) & N(t) \end{pmatrix} \tag{2.4}$$

is symplectic and continuously differentiable. Then the new variables y, z satisfy the LHS

$$y' = \bar{A}(t)y + \bar{B}(t)z, \quad z' = \bar{C}(t)y - \bar{A}^T(t)z, \tag{2.5}$$

where

$$\begin{aligned} \bar{A} &= N^T[-H' + AH + BK] - M^T[-K' + CH - A^T K], \\ \bar{B} &= N^T[-M' + AM + BN] - M^T[-N' + CM - A^T N], \\ \bar{C} &= -K^T[-H' + AH + BK] + H^T[-N' + CM - A^T N]. \end{aligned}$$

Observe that in the case $M(t) \equiv 0$ transformation (2.4) preserves oscillatory properties of transformed systems since then $H(t)$ is nonsingular (compare (2.3)), hence t_1, t_2 are conjugate relative to (1.1) if and only if they are conjugate relative to (2.5). Consequently, transformation (2.4) with $M(t) \equiv 0$ is the powerful tool for the investigation of oscillatory properties of (1.1). This system is transformed into an "easier" system and from oscillatory properties of this "easy" system we deduce oscillatory properties of (1.1). One of such "easy systems" is the so-called trigonometric system.

Theorem 1 (Došlý [10]). *There exist continuously differentiable $n \times n$ matrices H, K such that H is nonsingular, $H^T K \equiv K^T H$ and the transformation*

$$\begin{pmatrix} x \\ u \end{pmatrix} = \begin{pmatrix} H(t) & 0 \\ K(t) & (H^T(t))^{-1} \end{pmatrix} \begin{pmatrix} y \\ z \end{pmatrix} \tag{2.6}$$

transforms (1.1) into the trigonometric system

$$y' = Q(t)z, \quad z' = -Q(t)y, \tag{2.7}$$

where Q is a nonnegative definite symmetric $n \times n$ matrix.

The trigonometric system was introduced in [7] in connection with the Prüfer-type transformation for (1.1) and the terminology *trigonometric system* is justified by the fact that in the scalar case, i.e. when y, z, q are scalar quantities, then

$$(y_1, z_1) = (\sin \int^t q(s)\,ds, \cos \int^t q(s)\,ds),$$
$$(y_2, z_2) = (\cos \int^t q(s)\,ds, -\sin \int^t q(s)\,ds)$$

form the basis of the solution space of (2.7). In the n-dimensional case system (2.7) cannot be in general solved explicitly, but it may be proved that its solutions have many properties which in the scalar case reduce to the well-known trigonometric identities and these properties we may use to study properties of (1.1). For example, (2.7) is oscillatory if and only if

$$\int^\infty \operatorname{Tr} Q(t)\,dt = \infty,$$

where Tr stands for the trace, i.e. the sum of the diagonal entries of the matrix indicated.

Now turn our attention to the reciprocity principle for LHS. We start with the following elementary example. Consider the second order equation

$$(r(t)y')' + p(t)y = 0 \tag{2.8}$$

with *positive* coefficients r, p. If we denote $z = r(t)y'$ then this function verifies the so-called *reciprocal equation*

$$\left(\frac{1}{p(t)}z'\right)' + \frac{1}{r(t)}z = 0. \tag{2.9}$$

Using an elementary argument it is easy to see that a solution y of (2.8) oscillates if and only if its derivatives y' oscillates, i.e. (2.8) is oscillatory if and only if (2.9) is oscillatory. These equations may be written in the form of Hamiltonian system (1.1)

$$y' = \frac{1}{r(t)}z, \quad z' = -p(t)y \tag{2.10}$$

and

$$\tilde{y}' = p(t)\tilde{z}, \quad \tilde{z}' = -\frac{1}{r(t)}\tilde{y} \tag{2.11}$$

and these systems are related by the transformation

$$\begin{pmatrix} y \\ z \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} \tilde{y} \\ \tilde{z} \end{pmatrix}. \tag{2.12}$$

The statement concerning relation between oscillatory behaviour of these systems may be now formulated as follows: If the functions r, p are positive then transformation (2.12) preserves oscillation properties of transformed 2×2 systems, i.e. (2.10) is oscillatory if and only (2.11) is oscillatory.

Reciprocity principle concerns extension of this statement to Hamiltonian system (1.1).

Theorem 2 (Ahlbrandt [2]). *Suppose that $B(t) \geq 0$, $C(t) \leq 0$ (this means that B is nonnegative definite and C nonpositive definite) for large t and both system (1.1) and its reciprocal system*

$$y' = -A^T(t)y - C(t)z, \quad z' = -B(t)y + A(t)z \tag{2.13}$$

are eventually controllable (i.e., the trivial solution $(x, u) = (0, 0)$ is the only solution of (1.1) for which one of the components x, u is eventually vanishing). Then (1.1) is oscillatory if and only if (2.13) is oscillatory.

Obviously, this statement is a generalization of the relationship between (2.10) and (2.11) and claims, roughly speaking, that (1.1) is oscillatory with respect to the first component x if and only if it is oscillatory with respect to the second component u (compare Definition 1). Indeed, (2.13) results from (1.1) upon the transformation

$$\begin{pmatrix} x \\ u \end{pmatrix} = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \begin{pmatrix} y \\ z \end{pmatrix} = \mathcal{J} \begin{pmatrix} y \\ z \end{pmatrix} \tag{2.14}$$

which essentially only reverses the order of equations in (1.1). In another words, under definiteness assumption on the matrices B, C, transformation (2.14) preserves oscillatory properties of transformed systems.

The above mentioned reciprocity principle may be easily shown to be a particular case of the following general statement concerning transformations of (1.1) preserving oscillatory behaviour of transformed systems.

Theorem 3 (Došlý [11]). *Consider Hamiltonian systems (1.1) and (2.5) related by transformation (2.4) and suppose that matrices $B(t), \bar{B}(t)$ in these systems are nonnegative definite for large t. Then (1.1) is oscillatory if and only if (2.5) is oscillatory.*

This statement is proved using the trigonometric transformation given in Theorem 1. Systems (1.1) and (2.5) are transformed into trigonometric systems (using transformation of the form (2.6) which preserves oscillatory properties) with matrices Q and $\bar{Q}$ and then it is shown that $\int^\infty \operatorname{Tr} Q(t)\,dt = \infty$ if and only if $\int^\infty \operatorname{Tr} \bar{Q}(t)\,dt = \infty$. This means that these trigonometric systems and hence also systems (1.1), (2.5) are simultaneously oscillatory or nonoscillatory.

3 Discrete Hamiltonian Systems

Similar to the continuous case, we start with the definition of basic concepts.

Definition 2. We say that an interval $(k, k+1]$, $k \in \mathbb{N}$, contains a *generalized zero* of a solution (x, u) of (1.2) if $x_k \neq 0$ and there exists $c \in \mathbb{R}^n$ such that

$$x_{k+1} = \tilde{A}_k B_k c, \quad \text{and} \quad x_k^T B_k^\dagger (I - A_k) x_{k+1} \leq 0.$$

System (1.2) is said to be *disconjugate* in an interval $[n, m]$ if any solution of (1.2) has at most one generalized zero in $[n, m+1]$ and, moreover, any solution satisfying $x_n = 0$ has no generalized zero in $(n, m+1]$, in the opposite case (1.2) is said to be *conjugate* in $[n, m]$. System (1.2) is said to be *nonoscillatory* if there exists $n \in \mathbb{N}$ such that (1.2) is disconjugate on $[n, m]$ for every $m > n$, in the opposite case (1.2) is said to be *oscillatory*.

In the above definition † denotes the Moore-Penrose generalized inverse matrix, for an $n \times n$ matrix V its generalized inverse $V^\dagger$ is the (unique) $n \times n$ matrix such that $VV^\dagger$, $V^\dagger V$ are symmetric and $V^\dagger V V^\dagger = V^\dagger$, $VV^\dagger V = V$.

Basic oscillatory properties of discrete Hamiltonian systems are summarized in the discrete version of Roundabout Theorem.

Proposition 2 (Bohner [8]). *The following statements are equivalent:*

(i) *System (1.2) is disconjugate in the interval* $[0, N]$, $N \in \mathbb{N}$.

(ii) *The discrete quadratic functional*

$$\mathcal{F}(x, u) = \sum_{k=0}^{N} \left\{ u_k^T B_k u_k + x_{k+1}^T C_k x_{k+1} \right\}$$

is positive for every (x, u) *satisfying* $\Delta x_k = A_k x_{k+1} + B_k u_k$ *with* $x_0 = 0 = x_{N+1}$ *and* $x \not\equiv 0$.

(iii) *The matrix solution* (X, U) *of (1.2) given by the initial condition* $X_0 = 0$, $U_0 = I$ *satisfies*

$$\operatorname{Ker} X_{k+1} \subseteq \operatorname{Ker} X_k \qquad \text{and} \qquad X_k X_{k+1}^\dagger \tilde{A}_k B_k \geq 0, \ k = 1, \ldots, N.$$

(iv) *There exists a conjoined basis* (X, U) *of (1.2) (this is defined in the same way as for (1.1)) such that* X_k *are nonsingular and* $X_k X_{k+1}^{-1} \tilde{A}_k B_k \geq 0$, $k = 0, \ldots, N$.

(v) *There exist symmetric matrices* Q_k *such that* $(I + B_k Q_k)$ *are nonsingular,* $(I + B_k Q_k)^{-1} B_k \geq 0$, *and verify the discrete Riccati matrix difference equation*

$$Q_{k+1} = C_k + (I - A_k^T) Q_k (I + B_k Q_k)^{-1} (I - A_k), \tag{3.1}$$

$k = 0, \ldots, N$.

Concerning transformations of discrete LHS (1.2), the situation is not so easy as in the continuous case. In the discrete case it is supposed that the matrices $(I - A_k)$ are nonsingular and this assumption must satisfy also the system resulting after a transformation. To ensure this, we need an extra assumption as shows the next theorem.

Theorem 4 (Došlý [12]). *Let $\mathcal{R}_k$ be a $2n \times 2n$ symplectic matrix consisting of $n \times n$ matrices $\mathcal{R}_k = \begin{pmatrix} H_k & M_k \\ K_k & N_k \end{pmatrix}$ such that the matrix*

$$\begin{pmatrix} H_k + B_k K_k & (I - A_k)M_{k+1} \\ (I - A_k^T)K_k & N_{k+1} - C_k M_{k+1} \end{pmatrix} \tag{3.2}$$

is nonsingular and denote $\begin{pmatrix} D_k & F_k \\ E_k & G_k \end{pmatrix}$ its inverse. The transformation

$$\begin{pmatrix} x \\ u \end{pmatrix} = \mathcal{R}_k \begin{pmatrix} y \\ z \end{pmatrix} \tag{3.3}$$

transforms (1.2) into the system

$$\Delta y_k = \bar{A}_k y_{k+1} + \bar{B}_k z_k, \quad \Delta z_k = \bar{C}_k y_{k+1} - \bar{A}_k^T z_k, \tag{3.4}$$

where

$$\begin{aligned}
\bar{A}_k &= D_k(-\Delta H_k + A_k H_{k+1} + B_k K_k) + F_k(-\Delta K_k + C_k H_{k+1} - A_k^T K_k),\\
\bar{B}_k &= D_k(-\Delta M_k + A_k M_{k+1} + B_k N_k) + F_k(-\Delta N_k + C_k M_{k+1} - A_k^T N_k),\\
\bar{C}_k &= E_k(-\Delta H_k + A_k H_{k+1} + B_k K_k) + G_k(-\Delta K_k + C_k H_{k+1} - A_k^T K_k),
\end{aligned}$$

in particular, the matrices $\bar{B}_k, \bar{C}_k$ are symmetric and $(I - \bar{A}_k)$ are nonsingular, i.e. (3.4) is again a difference LHS.

Having now in disposal the above given statements, we may try to extend the reciprocity principle and trigonometric transformation to discrete systems. Let us start with the reciprocity principle. If we apply transformation (3.3) with $\mathcal{R} = \mathcal{J}$ to (1.2) (this transformation relates (1.1) and (2.13) in the continuous case), it is easy to see that the assumption of Theorem 4 concerning nonsingularity of the matrix in (3.2) is not generally satisfied, i.e. the resulting *(reciprocal)* system

$$\Delta y_k = -A_k^T y_k - C_k z_{k+1}, \quad \Delta z_k = -B_k y_k + A_k z_{k+1} \tag{3.5}$$

is the system of a different kind than (1.2). In fact, the variable x which defines oscillatory properties of (1.2) appears in the right-hand-sides of this system with indices $k + 1$, whereas the variable y which should define oscillations of (3.5) appears there with indices k. For this reason, Definition 2 and Proposition 2 do not apply to (3.5). However, as suggests the equivalence between oscillatory properties of the pair of second order equations $\Delta(r_k \Delta x_k) + p_k x_{k+1} = 0$ and $\Delta(p_k^{-1} \Delta z_k) + r_{k+1}^{-1} z_{k+1} = 0$ with positive r_k, p_k, which follows using the same

argument as in the continuous case, one can expect some kind of similarity between oscillatory properties of (1.2) and (3.5).

In studying the relationship between (1.2) and (3.5), the principal role play the so-called *symplectic systems*, i.e. systems of the form

$$\begin{pmatrix} x_{k+1} \\ u_{k+1} \end{pmatrix} = S_k \begin{pmatrix} x_k \\ u_k \end{pmatrix}, \quad S_k = \begin{pmatrix} \mathcal{A}_k & \mathcal{B}_k \\ \mathcal{C}_k & \mathcal{D}_k \end{pmatrix}, \tag{3.6}$$

where S_k are symplectic $2n \times 2n$ matrices. Expanding forward differences in (1.2) and (3.5), it is not difficult to see that these systems are symplectic systems. Oscillation theory of symplectic systems was established in [9] and fundamental definition is the following:

Definition 3. We say that the interval $(k, k+1]$ contains the *generalized zero* of a solution (x, u) of (3.6) if $x_k \neq 0$, there exists $c \in \mathbb{R}^n$ such that

$$x_{k+1} = \mathcal{B}_k c \quad \text{and} \quad x_{k+1}^T \mathcal{B}_k^\dagger x_k \leq 0.$$

Oscillation and nonoscillation of symplectic systems are defined via generalized zeros in the same way as for Hamiltonian systems. Applying these definitions to (1.2) and (3.5) we get the following discrete version of the reciprocity principle.

Theorem 5 (Došlý-Bohner [9]). *Suppose that both systems* (1.2) *and* (3.5) *are eventually controllable. If* $C_k \leq 0$ *for large* k *and* (1.2) *is nonoscillatory, then reciprocal system* (3.5) *is also nonoscillatory. Conversely, if* $B_k \geq 0$ *for large* k *and* (3.5) *is nonoscillatory then* (1.2) *is also nonoscillatory.*

Essentially the same difficulty as in the the case of the reciprocity principle we meet when trying to extend the trigonometric transformation to difference Hamiltonian systems (1.2). Trigonometric system (2.7) may be characterized as a Hamiltonian system which complies with its reciprocal system. Since the reciprocity transformation does not preserve the Hamiltonian structure of transformed difference systems, also in this case we have to pass to symplectic systems. By a direct computation one may verify that the transformation $\binom{x}{u} = \mathcal{J}\binom{\tilde{x}}{\tilde{u}}$ transforms (3.6) into itself if and only if $\mathcal{D} = \mathcal{A}$, $\mathcal{C} = -\mathcal{B}$. A symplectic system (3.6) having this property we will call *self-reciprocal* and such system may be regarded as a discrete analogue of the trigonometric differential system (2.7). However, it is an open problem whether any symplectic system may be transformed (by a transformation preserving oscillatory properties, i.e. by (3.3) with $M \equiv 0$) into a self-reciprocal system. Moreover, in contrast to trigonometric systems, till now no necessary and sufficient condition for oscillation of self-reciprocal symplectic systems is known.

We finish this section with discrete version of Theorem 5. To introduce this statement, consider the transformation of symplectic system (3.6)

$$\begin{pmatrix} x \\ u \end{pmatrix} = \mathcal{R}_k \begin{pmatrix} \tilde{x} \\ \tilde{u} \end{pmatrix}, \quad R_k = \begin{pmatrix} H_k & M_k \\ K_k & N_k \end{pmatrix} \tag{3.7}$$

with a symplectic $2n \times 2n$ matrix $\mathcal{R}$. Directly one can verify that this transformation transforms (2.7) into another symplectic system

$$\begin{pmatrix} \tilde{x}_{k+1} \\ \tilde{u}_{k+1} \end{pmatrix} = \tilde{S}_k \begin{pmatrix} \tilde{x}_k \\ \tilde{u}_k \end{pmatrix}, \quad \tilde{S}_k = \begin{pmatrix} \tilde{\mathcal{A}}_k & \tilde{\mathcal{B}}_k \\ \tilde{\mathcal{C}}_k & \tilde{\mathcal{D}}_k \end{pmatrix}. \tag{3.8}$$

The $n \times n$ matrices $\tilde{\mathcal{A}}, \tilde{\mathcal{B}}, \tilde{\mathcal{C}}, \tilde{\mathcal{D}}$ can be expressed via matrices H, K, M, N in a similar way as in Theorem 4 but we will not need these formulas.

Theorem 6 (Došlý - Hilscher [13]). *Suppose that systems (3.6) and (3.8) are related by transformation (3.7) with a symplectic matrix $\mathcal{R}$ and consider the following hypotheses:*

(i) *Both systems (3.6) and (3.8) are eventually controllable;*

(ii) *The matrices M and $\mathcal{A}M + \mathcal{B}N$ are eventually nonsingular;*

(iii) *Eventually, $R(NM^{-1}) \geq 0$, where $R(\cdot)$ is given by*

$$R(Q)_k \equiv Q_{k+1} - (\mathcal{C}_k + \mathcal{D}_k Q_k)(\mathcal{A}_k + \mathcal{B}_k Q_k)^{-1}.$$

(iv) *Eventually, $\tilde{R}(-M^{-1}H) \geq 0$, where*

$$\tilde{R}(\tilde{Q}) := -\tilde{Q}_k + (-\tilde{Q}_{k+1}\tilde{\mathcal{B}}_k + \tilde{\mathcal{D}}_k)^{-1}(\tilde{Q}_{k+1}\tilde{\mathcal{A}}_k - \tilde{\mathcal{C}}_k).$$

If the assumptions (i), (ii), (iii) *hold and (3.6) is eventually disconjugate then (3.8) is also eventually disconjugate. Conversely, if* (i), (ii), (iv) *hold and (3.8) is eventually disconjugate then (3.6) is eventually disconjugate.*

Obviously, if $\mathcal{R} = \mathcal{J}$, i.e. $H = 0 = N$, $-K = I = M$ and (3.6) corresponds to (1.2), i.e.

$$S = \begin{pmatrix} \tilde{A} & \tilde{A}B \\ C\tilde{A} & C\tilde{A}B + I + A^T \end{pmatrix},$$

then this statement reduces to reciprocity principle given in Theorem 6.

4 Hamiltonian Systems on Time Scales

In this section we discuss briefly possibilities of a unified approach to the investigation of discrete and continuous Hamiltonian systems. One of such possibilities consists in transforming both (1.1) and (1.2) into an integral equation with Riemann-Stiltjes integrals. This approach has been offered by Reid in [19], where the Roundabout Theorem for these generalized systems is presented. However, as pointed out in [5], this method when applied to difference systems (1.2) requires the matrix B to be nonnegative definite and as shows the Roundabout Theorem for difference systems (Proposition 2) this assumption is not needed there.

Another unified approach to continuous and discrete systems is based on the theory of equations on the so-called time scales. A *time scale* $\mathbb{T}$ is defined to be

any closed subset of real numbers $\mathbb{R}$ (an alternative terminology for time scale is *measure chain* [14]). On this set there are defined operators $\sigma, \rho : \mathbb{T} \to \mathbb{T}$

$$\sigma(t) := \inf\{s \in \mathbb{T},\ s > t\}, \quad \rho(t) := \sup\{s \in \mathbb{T},\ s < t\}.$$

A point $t \in \mathbb{T}$ is said to be *left-dense* (l-d) if $\rho(t) = t$, *right-dense* (r-d) if $\sigma(t) = t$, *left-scattered* (l-s) if $\rho(t) < t$, *right-scattered* (r-s) if $\sigma(t) > t$ and it is said to be *dense* if it is r-d or l-d. The *graininess* μ of a time scale $\mathbb{T}$ is defined by $\mu(t) := \sigma(t) - t$. For a function $f : \mathbb{T} \to \mathbb{R}$ (the range $\mathbb{R}$ of f may be actually replaced by any Banach space) it is defined the *generalized derivative*

$$f^{\Delta}(t) = \lim_{s \to t} \frac{f(\sigma(t)) - f(s)}{\sigma(t) - s}, \quad \text{where } s \in \mathbb{T} \backslash \{\sigma(t)\}.$$

As a basic reference concerning the differential and integral calculus on time scales we suggest the monograph [6] and the paper [14]. In particular cases $\mathbb{T} = \mathbb{R}$ and $\mathbb{T} = \mathbb{Z}$ the generalized derivative $f^{\Delta}(t)$ reduces to the usual derivative $f'(t)$ and to the usual forward difference $\Delta f(t) = f(t+1) - f(t)$, respectively.

Linear Hamiltonian system on a time scale $\mathbb{T}$ is the system

$$x^{\Delta}(t) = A(t)x(\sigma(t)) + B(t)u(t), \quad u^{\Delta}(t) = C(t)x(\sigma(t)) - A^T(t)u(t),$$

where it is supposed that $A, B, C : \mathbb{T} \to \mathbb{R}^{n \times n}$, B, C are symmetric and $\tilde{A} = (I - \mu A)^{-1}$ exists. The corresponding quadratic functional and the Riccati matrix equation are of the form

$$\mathcal{F}(x, u) = \int_a^b \left\{u^T(t)B(t)u(t) + x^T(\sigma(t))C(t)x(\sigma(t))\right\} \Delta t \tag{4.1}$$

and

$$Q^{\Delta}(t) - C(t) + A^T(t)Q + (Q(\sigma(t)) - \mu(t)C(t))\tilde{A}(t)(A(t) + B(t)Q(t)) = 0. \tag{4.2}$$

respectively. Concerning the definition of the integral over a subset of a time scale appearing in (4.1), we will not specify explicitly this definition and we note only that this integral reduces to the usual Riemann integral in case $\mathbb{T} = \mathbb{R}$ and to the usual sum if $\mathbb{T} = \mathbb{Z}$, the exact definition of this integral is given e.g. in [6]. Substituting $\mu \equiv 0$ (continuous case) in (4.2) we get equation (2.2) and substituting $\mu \equiv 1$ (discrete case) we have (3.1). As a basic reference concerning qualitative theory of Hamiltonian systems on time scales may be regarded the recent papers of Agarwal and Bohner [1] and of Hilscher [15]. Here the main result is the "Partly Roundabout Theorem", relating positivity of the functional (4.1), existence of a symmetric solution of (4.2) and the existence of a self-conjoined basis of the matrix system

$$X^{\Delta}(t) = A(t)X(\sigma(t)) + B(t)U(t), \quad U^{\Delta}(t) = C(t)X(\sigma(t)) - A^T(t)U(t) \tag{4.3}$$

without focal points (the conjoined basis of (4.3) is defined in the same way as for (1.1) and (1.2)). The word "partly" in the name of this statement is motivated by the fact that the proofs of some implications with respect to the "classical" Roundabout Theorem are still missing and these proofs are subject of the present investigation.

The advantages of the time scale approach to Hamiltonian systems well illustrates the explanation why in the discrete oscillation theory no assumption concerning definiteness of the matrix B and controllability of (1.2) is needed, whereas in the continuous oscillation theory controllability of (1.1) it is necessary (at least for the formulation of the Roundabout Theorem in the form given here) and the assumption $B(t) \geq 0$ plays there a crucial role here — in the calculus of variations it is known as the Legendre necessary condition for positivity of the functional $\mathcal{F}$ given in Proposition 1.

Following [15], a conjoined basis (X, U) of (4.3) has no focal point in an interval $\mathcal{I} := (a, b] \cap \mathbb{T}$ provided $X(t)$ is invertible in all dense points of $\mathcal{I}$,

$$\operatorname{Ker} X(\sigma(t)) \subseteq \operatorname{Ker} X(t) \quad \text{and} \quad D(t) := X(t)(X(\sigma(t))^{\dagger}\tilde{A}(t)B(t) \geq 0 \tag{4.4}$$

in this interval. Consequently, $\mathcal{I}$ contains a focal point whenever one of the following conditions holds:

(i) There exists $s \in \mathcal{I}$ such that $\operatorname{Ker} X(\sigma(t)) \not\subseteq \operatorname{Ker} X(t)$, or

(ii) $\operatorname{Ker} X(\sigma(t)) \subseteq \operatorname{Ker} X(t)$ on $\mathcal{I}$ and X is singular at some dense point $s \in \mathcal{I}$, or

(iii) For every $t \in \mathcal{I}$ we have $\operatorname{Ker} X(\sigma(t)) \subseteq \operatorname{Ker} X(t)$, X is nonsingular in all dense points of $\mathcal{I}$, but $D(s) := X(s)X^{\dagger}(\sigma(s))\tilde{A}(s)B(s) \not\geq 0$ at some $s \in \mathcal{I}$.

Nonexistence of a focal point of the matrix solution (X, U) of (4.3) given by the initial condition $X(a) = 0, U(a) = I$ is sufficient for positivity of the functional (4.1) in the class of n-dimensional pairs (x, u) satisfying $x^{\Delta}(t) = A(t)x(\sigma(t)) + B(t)u(t)$ and $x(a) = 0 = x(b)$, see [15]. The proof of this statements is based on the generalized Picone identity where the quantity $D(t)$ defined in (4.4) plays a crucial role.

In the continuous case $\mathbb{T} = \mathbb{R}$, controllability of (1.1) implies that singularities of X are isolated, in particular, that X is nonsingular in some right neighbourhood of $t = a$. Since $\sigma(t) = t$, $\mu(t) \equiv 0$, we have $D(t) = B(t)$ and focal points of X are singularities of X or points where B fails to be nonnegative definite. However, the last possibility is eliminated by the *apriori* assumption $B \geq 0$ and focal points of X are just singularities of this matrix as it is usual in oscillation theory of differential systems. In the discrete case $\mathbb{T} = \mathbb{Z}$ all points are automatically isolated and this explains why controllability assumption is not needed in this case.

Finally, one may also easily see why the assumption of invertibility of the matrix $(I - A_k)$ (supposed in the discrete case) has no continuous analogue. This is a particular case of the general assumption of invertibility of $(I - \mu(t)A(t))$ which is in the continuous case $\mu(t) \equiv 0$ trivially satisfied.

Supported by the Grant No. 201/98/0677 of the Czech Grant Agency (Prague).

References

1. R. P. Agarwal, M. Bohner, *Quadratic functionals for second order matrix equations on time scales*, submitted
2. C. D. Ahlbrandt, *Equivalent boundary value problems for self-adjoint differential system*, J. Diff. Equations, **9** (1971), 420–435.
3. C. D. Ahlbrandt, *Recessive solutions of symmetric three term recurrence relations*, Canad. Math. Soc. Proceedings, **8** (1987), 3–42.
4. C. D. Ahlbrandt, A. C. Peterson, *Discrete Hamiltonian Systems: Difference Equations, Continued Fractions, and Riccati Equations*, Kluwer, Boston 1996.
5. C. D. Ahlbrandt, S. L. Clark, J. W. Hooker, W. T. Patula, *A Discrete interpretation of Reid's Roundabout Theorem for generalized differential systems*, Comput. Appl. Math. **28** (1994), 11–21.
6. B. Aulbach, S. Hilger, *Linear dynamic processes with inhomogeneous time scale*, In Nonlinear Dynamics and Quantum Dynamical Systems, Akademie Verlag, Berlin, 1990.
7. J. H. Barrett, *A Prüfer transformation for matrix differential system*, Proc. Amer. Math. Soc., **8** (1957), 510–518.
8. M. Bohner, *Linear Hamiltonian difference systems: disconjugacy and Jacobi-type conditions*, J. Math. Anal. Appl., **199** (1996), 804–826.
9. M. Bohner, O. Došlý, *Disconjugacy and transformations for symplectic systems*, Rocky Mountain J. Math., **27** (1997), 707–743.
10. O. Došlý, *On transformations of self-adjoint linear differential systems and their reciprocals*, Annal. Pol. Math., **50** (1990), 223–224.
11. O. Došlý, *Transformations of linear Hamiltonian systems preserving oscillatory behaviour*, Arch. Math. (Brno), **27b** (1991), 211–219.
12. O. Došlý, *Transformations of linear Hamiltonian difference systems and some of their applications*, J. Math. Anal. Appl., **191** (1995), 250–265.
13. O. Došlý, R. Hilscher, *Linear Hamiltonian difference systems: transformations, recessive solutions, generalized reciprocity,* submitted, 1997.
14. S. Hilger, *Analysis on measure chains — a unified approach to continuous and discrete calculus*, Res. Math. **18** (1990), 18–56.
15. R. Hilscher, *A unified approach to continuous and discrete linear Hamiltonian systems via the calculus on time scales*, submitted 1998.
16. W. Kratz, *Quadratic Functionals in Variational Analysis and Control Theory*, Akademie Verlag, Berlin, 1995.
17. M. Morse, *A generalization of Sturm separation and comparison theorems in n-space*, *Math. Ann.*, **103** (1930), 52–69.
18. W. T. Reid, *Sturmian Theory for Ordinary Differential Equations*, Springer Verlag, New York-Heidelberg-Berlin 1980.
19. W. T. Reid, *Generalized linear differential systems and related Riccati matrix integral equation*, Illinois J. Math. **10** (1966), 701–722.
20. C. A. Swanson, *Comparison and Oscillation Theorems for Linear Differential Equations*, Acad. Press, London 1968.

EQUADIFF 9 CD ROM, Brno 1997 PROCEEDINGS
Masaryk University pp. 63–80

On the Steady Translational Self-Propelled Motion of a Body in a Navier-Stokes Fluid

Giovanni P. Galdi

Department of Mathematics & Statistics,
301 Thackeray Hall, University of Pittsburgh,
Pittsburgh 15260, PA U.S.A.
Permanent address: Istituto di Ingegneria, Università di Ferrara,
Ferrara 44100 Italy
Email: galdi@math.pitt.edu

Abstract. There is a large interest, from both mathematical [4], [5], [14], [15] and physical [20], [3] point of view, in the motion of self-propelling bodies in a viscous fluid. This latter means that the body $\mathcal{B}$, say, moves without the action of an external force (like gravity, for instance), but just because of the interaction between its boundary Σ, say, and the fluid. Therefore, Σ serves as the "driver" of $\mathcal{B}$ and the distribution of velocity on Σ as its "thrust". In this paper we shall consider steady, translational self-propelled motion of a body in a Navier-Stokes fluid. In particular, we show the existence of a space $\mathcal{T}(\mathcal{B})$ of velocity distributions on $\mathcal{S}$ with the property that for any given translational velocity V of $\mathcal{B}$ there is one and only one element in $\mathcal{T}(\mathcal{B})$ which can move the body with velocity V. $\mathcal{T}(\mathcal{B})$ depends only on the geometric properties of $\mathcal{B}$ such as size or shape. In particular, it is independent of the orientation of $\mathcal{B}$ and on the fluid property.

AMS Subject Classification. 35Q, 76C

Keywords. Steady-state Navier-Stokes equations, self-propelled body, existence and uniqueness

1 Introduction

A body $\mathcal{B}$ moving in an infinite viscous fluid $\mathcal{F}$ undergoes a *self-propelled motion* if the net total force and torque, external to $\mathcal{B}$ and $\mathcal{F}$, acting on $\mathcal{B}$ are identically zero. Examples of self-propelled motions can be those performed by rockets, submarines, fishes, microorganisms, etc. This type of motion is possible because of the interaction between the boundary of the body Σ, say, and the fluid. Therefore, Σ serves as the "driver" of $\mathcal{B}$ and and the distribution of velocity on Σ as its "thrust".

Since the pioneering work of G. I. Taylor [20] on the propulsion of microscopic organisms, these problems have attracted the attention of many scientists, particularly with the objective of giving a fluid mechanical interpretation of the

self-motion of ciliated and flagellated organisms, see, *e.g.* [3], [19], [2] and the bibliography cited therein. It should also be noticed that most of these results are derived under the assumption of zero Reynolds number, that is, Stokes approximation.

In this paper we shall consider steady, translational self-propelled motion of a body $\mathcal{B}$ in a Navier-Stokes fluid $\mathcal{F}$. By this we mean that $\mathcal{B}$ moves in $\mathcal{F}$ by purely translational motion, with constant velocity $-\xi \neq 0$, and that the motion of $\mathcal{F}$, as seen by an observer attached to $\mathcal{B}$, is independent of time. The shape of $\mathcal{B}$ is, of course, independent of time as well. Our goal is to investigate the class of velocity distributions on Σ (the "thrust") which makes $\mathcal{B}$ move with the given velocity $-\xi$. It is simple to see that this problem admits an *infinite* number of solutions corresponding to the same ξ; see Section 2. The objective of this work is to characterize a class of boundary velocities for which the problem admits *one and only one* solution. We shall show, among other things, that there exists a six-dimensional subspace $\mathcal{T}(\mathcal{B})$ of the space $L^2(\Sigma)$ with the following properties. $\mathcal{T}(\mathcal{B})$ depends only on the geometric properties of $\mathcal{B}$ such as size or shape and for any given ξ, there exists one and only one element of $\mathcal{T}(\mathcal{B})$ moving $\mathcal{B}$ with prescribed velocity $-\xi$. We thus give a general answer to a question which was addressed and/or partially solved by several authors. In this regard, we recall the paper of Lugovtsov & Lugovtsov [12] where particular examples are given of flow past a self-propelled body and to the contributions of Sennitskii [16], [17] who, by the method of matched asymptotic expansion, has constructed, for sufficiently small values of the Reynolds number an approximate solution in the case when $\mathcal{B}$ is a cylinder or a sphere, under different prescriptions of boundary velocity.[1] A similar type of question (momentumless flow) for $\mathcal{B}$ of arbitrary shape has been investigated and solved by Pukhnachev [14], [15] within the Stokes approximation. Recently, I have given a general existence and uniqueness theory for the full nonlinear problem, in the particular case when $\mathcal{B}$ has rotational symmetry [6].[2]

The paper is organized as follows. In Section 2 we formulate the problem and introduce some notation. In Section 3, we study the linearized version of the self-motion of $\mathcal{B}$ within the Stokes approximation and furnish, in particular, necessary and sufficient conditions on the distribution of velocity on Σ in order that $\mathcal{B}$ performs a steady, translational flow. These results contain, as a particular case, those of [14], [15] and are of fundamental importance in the investigation of the nonlinear problem which is the object of Section 4. There, we shall show that for any translational motion of $\mathcal{B}$ with velocity -ξ, there exists a uniquely determined velocity distribution on Σ, which lies in a six-dimensional "control" space $\mathcal{T}(\mathcal{B})$, provided $|\xi|$ is not "too large". $\mathcal{T}(\mathcal{B})$ depends only on $\mathcal{B}$. Furthermore, the set

[1] See also [18] for a dynamical counterpart of these problems.

[2] We would like to mention also a series of works aimed at investigating the asymptotic behaviour of the velocity field of the fluid within the wake behind $\mathcal{B}$. We refer, in particular, to the work of Birkhoff and Zarantonello [4], Finn [5] and, more recently, Pukhnachev [13], Kozono and Sohr [10], and Kozono, Sohr and Yamazaki [11].

of all translational motions of $\mathcal{B}$ is set in a one-to-one correspondence with a subspace $\mathcal{T}'(\mathcal{B})$ of $\mathcal{T}(\mathcal{B})$.

An interesting question we leave open is that of the uniqueness of the space $\mathcal{T}'(\mathcal{B})$. In other words, assume there is another space $\tilde{\mathcal{T}}(\mathcal{B})$ with the property that any translational motion of $\mathcal{B}$ determines and is determined by a unique element of $\tilde{\mathcal{T}}(\mathcal{B})$. The question is if and how $\mathcal{T}'(\mathcal{B})$ and $\tilde{\mathcal{T}}(\mathcal{B})$ are related to each other.

Acknowledgment. This paper is part of a keynote lecture I gave at the Conference Equadiff 9, held in Brno in August 1997. I am particularly grateful to Professors F. Neuman and J. Vosmanský for their kind invitation and warm hospitality.

2 Formulation of the Problem

Assume a body $\mathcal{B}$ moves of translational motion, with constant velocity $-\xi$ in a Navier-Stokes fluid. We denote by $\mathcal{D}$ the region occupied by the fluid (the exterior of $\mathcal{B}$) and by Σ the boundary of $\mathcal{B}$. If the fluid performs a time-independent flow, the relevant equations, written in a frame $\mathcal{S}$ attached to $\mathcal{B}$ and in dimensionless form, become[3]

$$\left.\begin{aligned} \Delta v = \lambda v \cdot \nabla v + \nabla p \\ \operatorname{div} v = 0 \end{aligned}\right\} \quad \text{in } \mathcal{D}$$
$$v = v_* \quad \text{on } \Sigma \tag{2.1}$$
$$\lim_{|x|\to\infty} v(x) = \xi.$$

Here v, p are velocity and pressure field, respectively, associated to the particles of the fluid. Moreover, λ is the dimensionless Reynolds number which has the form LU/ν, where L is a characteristic length (the diameter of $\mathcal{B}$, for example), U is a characteristic speed (the speed of $\mathcal{B}$, in which case ξ is of modulus one) and ν is the kinematical viscosity coefficient of the fluid. We shall now append to these equations the conditions describing that $\mathcal{B}$ is self-propelling. Since $\mathcal{B}$ moves at steady pace, these latter are expressed by the requirement that, relative to an inertial frame, the total momentum flux and moment of momentum flux through Σ balance the total force and total moment of force exerted by the fluid on $\mathcal{B}$, respectively; see [20] pp. 448–449. Reformulating these conditions in the moving frame $\mathcal{S}$, we then obtain

$$\int_\Sigma [-T(v,p)\cdot n + \lambda(v_* - \xi)v_* \cdot n] = 0 \tag{2.2}$$

and

$$\int_\Sigma x \times [-T(v,p)\cdot n + \lambda(v_* - \xi)v_* \cdot n] = 0, \tag{2.3}$$

[3] We suppose that there is no body force acting on the fluid.

where n is the unit inward normal to $\mathcal{B}$ and $T = T(v,p)$ is the stress tensor whose components are given by

$$T_{ij}(v,p) = \frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} - p\,\delta_{ij}, \quad i,j = 1,2,3.$$

The main goal of this paper is to investigate the following Problem $\mathcal{P}$: *Given* $\xi \neq 0$, *find a solution to problem* (2.1)–(2.3).

Notice that, unlike the "classical" formulation of the exterior problem, the value of the velocity field at Σ is *not* prescribed.

As stated, it becomes clear that Problem $\mathcal{P}$ always admits the trivial solution $v = \xi, p = \text{const}$ and that, in general, it admits an infinite number of solutions corresponding to the same ξ. Actually, assume $\mathcal{B}$ is a body of revolution around the x_1-axis (say) and let Φ be *any* harmonic function in $\mathcal{D}$ approaching $\xi \cdot x$ at large distances and satisfying the following parity condition

$$\Phi(x_1,x_2,x_3) = \Phi(x_1,-x_2,x_3) = \Phi(x_1,x_2,-x_3).$$

Set $v = \nabla\Phi, p = \frac{1}{2}(\nabla\Phi)^2$. By a straightforward calculation, one shows that $v,\ p$ is a solution to (2.1), with $v_* = \nabla\Phi\,|_\Sigma$. Moreover, using the parity requirements it is obvious that also condition (2.3) is satisfied. Also, integrating $(2.1)_1$ on the subdomain of $\mathcal{D}$ delimited by the surface Σ and by the surface Σ_R of a ball of radius R centered in $\mathcal{B}$, we find

$$\int_\Sigma (T(v,p)\cdot n - \lambda(v_* - \xi)v_*\cdot n) = \int_{\Sigma_R} (T(v,p)\cdot n - \lambda(v-\xi)v\cdot n). \tag{2.4}$$

Thus, taking into account that $D^\sigma\Phi(x) = O(|x|^{-1-|\sigma|})$, $|\sigma| = 1,2$, we let $R \to \infty$ in (2.4), to deduce that also condition (2.2) holds.

This example shows that, in order to preserve uniqueness for Problem $\mathcal{P}$, we must impose some other restrictions on the class of solutions. We shall therefore require that the trace v_* of v at Σ belongs to a suitable "control" space $\mathcal{T}$. Our objective is to determine $\mathcal{T}$ in such a way that Problem $\mathcal{P}$ admits (one and) only one solution. We shall show that this is always the case, provided $\lambda|\xi|$ is not too large.

We shall briefly recall the main notation used in this paper.

$\mathbb{R}^3$ is the three-dimensional Euclidean space and (e_1,e_2,e_3) the canonical orthonormal basis.

Unless otherwise explicitly stated, we shall assume $\mathcal{D}$ sufficiently regular, for instance, of class C^2.

For $\beta = (\beta_1,\beta_2,\beta_3)$, $\beta_i \geq 0$, we set

$$D^\beta = \frac{\partial^{|\beta|}}{\partial x_1^{\beta_1}\partial x_2^{\beta_2}\partial x_3^{\beta_3}}, \quad |\beta| = \beta_1+\beta_2+\beta_3.$$

If $u = \{u_i\}$ is a vector function, by $D(u) = \{D_{ij}(u)\}$ we denote the symmetric part of $\nabla u = \left\{ \dfrac{\partial u_i}{\partial x_j} \right\}$, that is

$$D_{ij}(u) = \tfrac{1}{2} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right).$$

We shall use standard notations for function spaces, see [1]. So, for instance, $L^q(\mathcal{A})$, $W^{m,q}(\mathcal{A})$, etc., will denote the usual Lebesgue and Sobolev spaces on the domain $\mathcal{A}$, with norms $\| \cdot \|_{q,\mathcal{A}}$ and $\| \cdot \|_{m,q,\mathcal{A}}$, respectively. Whenever confusion will not arise, we shall omit the subscript $\mathcal{A}$. The trace space on $\partial\mathcal{A}$ for functions from $W^{m,q}(\mathcal{A})$ will be denoted by $W^{m-1/q,q}(\partial\mathcal{A})$ and its norm by $\| \cdot \|_{m-1/q,q,\partial\mathcal{A}}$. For other notation we follow [1].

3 The Stokes Approximation

We shall begin to consider the limiting situation of $\lambda \to 0$ of Problem $\mathcal{P}$. If we formally take $\lambda = 0$ into equations (2.1), (2.2), (2.3) we get the following problem

$$\left.\begin{aligned} \Delta v_0 &= \nabla p_0 \\ \operatorname{div} v_0 &= 0 \end{aligned}\right\} \quad \text{in } \mathcal{D}$$

$$\begin{aligned} v_0 &= v_{0*} \quad \text{on } \Sigma \\ \lim_{|x|\to\infty} v_0(x) &= \xi \\ \int_\Sigma T(v_0, p_0) \cdot n &= 0 \\ \int_\Sigma x \times T(v_0, p_0) \cdot n &= 0. \end{aligned} \tag{3.1}$$

We shall show that for any $\xi \neq 0$ there is a unique solution to (3.1) with v_{0*} in a suitable "control" space, see (3.17)

To prove this, we introduce some auxiliary fields, see [9] Chapter 5, [6]. Let $\{h_i, p_i\}$, $\{H_i, P_i\}$, $i = 1, 2, 3$, be the solutions to the following Stokes problems

$$\left.\begin{aligned} \Delta h^{(i)} &= \nabla p^{(i)} \\ \operatorname{div} h^{(i)} &= 0 \end{aligned}\right\} \quad \text{in } \mathcal{D}$$

$$\begin{aligned} h^{(i)} &= e_i \quad \text{on } \Sigma \\ \lim_{|x|\to\infty} h^{(i)}(x) &= 0, \end{aligned} \tag{3.2}$$

and

$$\left.\begin{aligned}\Delta H^{(i)} &= \nabla P^{(i)}\\ \operatorname{div} H^{(i)} &= 0\end{aligned}\right\} \quad \text{in } \mathcal{D}$$
$$H^{(i)} = e_i \times x \quad \text{on } \Sigma \tag{3.3}$$
$$\lim_{|x|\to\infty} H^{(i)}(x) = 0.$$

We also set

$$\begin{aligned} g^{(i)} &:= T(h^{(i)}, p^{(i)}) \cdot n\,|_\Sigma\,, \quad i = 1,2,3,\\ G^{(i)} &:= T(H^{(i)}, P^{(i)}) \cdot n\,|_\Sigma\,, \quad i = 1,2,3. \end{aligned} \tag{3.4}$$

The vector functions $g^{(i)} = g^{(i)}(x)$ and $G^{(i)} = G^{(i)}(x)$ depend only on the geometric properties of $\mathcal{B}$ such as size or shape. In particular, they do not depend on the orientation of $\mathcal{B}$ and on the fluid property. These functions $g^{(i)} = g^{(i)}(x)$ and $G^{(i)} = G^{(i)}(x)$ will play an important role and, in particular, we are interested in their linear independence properties. In this regard, we have.

Lemma 3.1. *The system of vector functions*

$$\mathsf{S} = \{g^{(i)}, G^{(i)}\}$$

is linearly independent.

Proof. Assume there are scalars γ_i, δ_i, $i = 1,2,3$ such that

$$\gamma_i g^{(i)}(x) + \delta_i G^{(i)}(x) = 0, \quad \text{for all } x \in \Sigma. \tag{3.5}$$

Setting

$$H = \gamma_i h^{(i)} + \delta_i H^{(i)}, \quad P = \gamma_i p^{(i)} + \delta_i P^{(i)}, \tag{3.6}$$

from (3.2), (3.3) we immediately deduce that H, P satisfy the following problem

$$\left.\begin{aligned}\Delta H &= \nabla P\\ \operatorname{div} H &= 0\end{aligned}\right\} \quad \text{in } \mathcal{D}$$
$$\lim_{|x|\to\infty} H(x) = 0 \tag{3.7}$$
$$H(x) = \gamma + \delta \times x, \quad x \in \Sigma.$$

Moreover, condition (3.5) gives

$$T(H, P) \cdot n = 0 \quad \text{at } \Sigma. \tag{3.8}$$

Multiplying $(3.7)_1$ by H, integrating by parts over $\mathcal{D}$, and and using well-known asymptotic properties of solutions to the Stokes problem, [7] Chapter V, we obtain

$$\int_{\mathcal{D}} |D(H)|^2 = \int_{\Sigma} (\gamma + \delta \times x) \cdot T(H, P) \cdot n.$$

In view of (3.8) we then deduce $H(x) = \gamma + \delta \times x$, for all $x \in \mathcal{D}$, and since $H(x)$ vanishes for $|x| \to \infty$ we conclude $\gamma = \delta = 0$, proving the assertion. □

We shall next furnish necessary and sufficient conditions in order that $\mathcal{B}$ performs a steady, translational self-propelled motion within the Stokes approximation with prescribed translational velocity; see (3.15). To this end, we multiply $(3.1)_1$ by $h^{(i)}$ and integrate by parts over $\mathcal{D}$ to find

$$e_i \cdot \int_{\Sigma} T(v_0, p_0) \cdot n = 2 \int_{\mathcal{D}} D(h^{(i)}) : D(v_0), \quad i = 1, 2, 3.$$

Likewise, multiplying $(3.2)_1$ by $v_0 + V_0$ and integrating by parts over $\mathcal{D}$, we obtain

$$\int_{\Sigma} (v_{0*} - \xi) \cdot g^{(i)} = 2 \int_{\mathcal{D}} D(h^{(i)}) : D(v_0), \quad i = 1, 2, 3. \tag{3.9}$$

These two displayed relations then imply

$$\int_{\Sigma} (v_{0*} - \xi) \cdot g^{(i)} = e_i \cdot \int_{\Sigma} T(v_0, p_0) \cdot n. \tag{3.10}$$

In a similar fashion, multiplying $(3.1)_1$ by $H^{(i)}$ and (3.3) by $v_0 - \xi$, respectively, and integrating by parts over $\mathcal{D}$ we find

$$e_i \cdot \int_{\Sigma} x \times T(v_0, p_0) \cdot n = 2 \int_{\mathcal{D}} D(H^{(i)}) : D(v_0), \quad i = 1, 2, 3.$$

and

$$\int_{\Sigma} (v_{0*} - \xi) \cdot G^{(i)} = 2 \int_{\mathcal{D}} D(H^{(i)}) : D(v_0), \quad i = 1, 2, 3 \tag{3.11}$$

which in turn give

$$\int_{\Sigma} (v_{0*} - \xi) \cdot G^{(i)} = e_i \cdot \int_{\Sigma} x \times T(v_0, p_0) \cdot n. \tag{3.12}$$

Let us consider the matrices $(i, j = 1, 2, 3)$, see [9] Chapter 5[4]

$$K_{ij} = -\int_{\Sigma} g_j^{(i)}, \quad C_{ij} = -\int_{\Sigma} \left(x \times g^{(i)} \right)_j, \tag{3.13}$$

[4] Notice that, in general, the matrix C depends on the origin of the axis.

and the vectors

$$\mathcal{V}_i = \int_\Sigma v_{0*} \cdot g^{(i)}, \quad \mathcal{W}_i = \int_\Sigma v_{0*} \cdot G^{(i)}, \quad i = 1, 2, 3, \tag{3.14}$$

from (3.10), (3.12) we find that v_{0*} generates a steady, translational self-propelled motion if and only if the following condition holds

$$\begin{aligned} \mathcal{V} &= K \cdot \xi_0 \\ \mathcal{W} &= C^\dagger \cdot \xi_0. \end{aligned} \tag{3.15}$$

In view of the linear independence of the system $\mathsf{S} = \{g^{(i)}, G^{(i)}\}$, see Lemma 3.1, we have that, setting

$$M_{ij} = \int_\Sigma g^{(i)} \cdot g^{(j)}, \quad N_{ij} = \int_\Sigma g^{(i)} \cdot G^{(j)}, \quad O_{ij} = \int_\Sigma G^{(i)} \cdot G^{(j)},$$

the 6×6 matrix

$$\begin{pmatrix} M & N \\ N^\dagger & O \end{pmatrix} \tag{3.16}$$

is invertible. Therefore, for any $\xi \in \mathbb{R}^3$ there exists a vector field $v_{0*} = \alpha_i g^{(i)} + \beta_i G^{(i)}$ with uniquely determined α, β satisfying (3.15), (3.14). However, it is also clear from (3.15) that in general it is *not* true that all vectors of the form (3.14) will generate a steady translational flow. This will happen *only* if certain compatibility conditions are satisfied. For example, if $\mathcal{B}$ has spherical symmetry, then one proves [9] Chapter 5, that $C^\dagger \equiv 0$. Consequently, from $(3.15)_2$ it follows that one must prescribe v_* in such a way that $\mathcal{W} = 0$. More generally, (3.15) will admit a solution ξ_0, if the data $\mathcal{V}$ and $\mathcal{W}$ satisfy the compatibility condition $\mathcal{W} = C^\dagger K^{-1} \mathcal{V}$.[5] An *arbitrary* prescription of $\mathcal{V}$ and $\mathcal{W}$ will, in general, produce also a rotation for $\mathcal{B}$, a possibility which is not considered in this paper, and this will be the object of future research.

To state the main results obtained above, it is convenient to introduce the following 6-dimensional subspace of $L^2(\Sigma)$

$$\mathcal{T}(\mathcal{B}) = \left\{ u \in L^2(\Sigma) : \ u = \alpha_i g^{(i)} + \beta_i G^{(i)}, \ \text{for some } \alpha, \beta \in \mathbb{R}^3 \right\}. \tag{3.17}$$

As we noticed, $\mathcal{T}(\mathcal{B})$ depends only on the geometric properties of $\mathcal{B}$ such as size or shape. In particular, it is independent of the orientation of $\mathcal{B}$ and on the fluid property.

Taking into account classical existence and uniqueness theorems for the exterior Stokes problem, see [7] Chapter V, we may then summarize the results obtained thus far in the following.

Theorem 3.1. *Let $\mathcal{B}$ have a locally lipschitzian boundary Σ. Then, for any $\xi \in \mathbb{R}^3$ there exists a unique solution v_0, p_0 to problem $(3.1)_{1,2,4,5,6}$ such that the restriction v_{0*} of v_0 to Σ belongs to $\mathcal{T}(\mathcal{B})$.*

[5] The matrix K is always invertible [9] Chapter 5.

4 Existence and Uniqueness for Problem $\mathcal{P}$

To solve Problem $\mathcal{P}$, we shall put it into an equivalent form. For a given $\xi \neq 0$, we set

$$v = u + \xi,$$

and find that $(2.1)_{1,2,4}$ is equivalent to

$$\left.\begin{aligned} \Delta u + \lambda\xi \cdot \nabla u = \lambda \operatorname{div} F(u) + \nabla p \\ \operatorname{div} u = 0 \end{aligned}\right\} \text{ in } \mathcal{D}$$
$$\lim_{|x|\to\infty} u(x) = 0 \tag{4.1}$$

while the self-propelling conditions (2.2), (2.3) become

$$\begin{aligned} \int_\Sigma u \cdot g^{(i)} &= \lambda \int_{\mathcal{D}} F : \nabla h^{(i)} + \lambda\xi \cdot \int_{\mathcal{D}} \nabla h^{(i)} \cdot u, \quad i = 1,2,3 \\ \int_\Sigma u \cdot G^{(i)} &= \lambda \int_{\mathcal{D}} F : \nabla H^{(i)} + \lambda\xi \cdot \int_{\mathcal{D}} \nabla H^{(i)} \cdot u, \quad i = 1,2,3. \end{aligned} \tag{4.2}$$

In (4.1), (4.2) we set

$$F(u) := u \otimes u, \tag{4.3}$$

and the vectors $g^{(i)}, G^{(i)}$ defined in (3.4). The identities (4.2) are obtained by multiplying first $(4.1)_1$ by $h^{(i)}$ and $H^{(i)}$, then (3.2), (3.3) by u, integrating by parts and proceeding as in the proof given for the case of the Stokes approximation (see (3.10), (3.12)). If the value of u at the boundary Σ is requested to be in the class $\mathcal{T}(\mathcal{B})$, (4.2) becomes

$$\begin{aligned} \alpha_j M_{ij} + \beta_j N_{ij} &= \lambda \int_{\mathcal{D}} F(u) : \nabla h^{(i)} + \lambda\xi \cdot \int_{\mathcal{D}} \nabla h^{(i)} \cdot u, \quad i = 1,2,3 \\ \alpha_j N_{ji} + \beta_j O_{ij} &= \lambda \int_{\mathcal{D}} F(u) : \nabla H^{(i)} + \lambda\xi \cdot \int_{\mathcal{D}} \nabla H^{(i)} \cdot u, \quad i = 1,2,3. \end{aligned} \tag{4.4}$$

for some $\alpha, \beta \in \mathbb{R}^3$.

A solution to (4.1), (4.3) and (4.4) will be obtained as a fixed point in a suitable Banach space. To this end, for $q \in (1, 3/2)$ we put

$$\begin{aligned} \langle\!\langle u \rangle\!\rangle_{\lambda,q} := (\lambda|\xi|)^{1/2} \|u\|_{2q/(2-q)} + (\lambda|\xi|)^{1/4} \|u\|_4 + \\ \|u\|_{3q/(3-2q)} + \|D^2 u\|_{2,q} + \|\nabla u\|_2, \end{aligned} \tag{4.5}$$

and set

$$\mathcal{X}^q = \left\{ \varphi \in L^1_{loc} : \ \langle\!\langle u \rangle\!\rangle_{\lambda,q} < \infty \right\}.$$

Clearly, $\mathcal{X}^q$ is a Banach space. We shall denote by $\mathcal{X}^q_\delta$ the ball of radius δ in $\mathcal{X}^q$.

We next consider the following map

$$\mathcal{N}: \ \varphi \in \mathcal{X}^q_\delta \to (\alpha, \beta) \to u$$

where α, β satisfy the following conditions

$$\begin{aligned} \alpha_j M_{ij} + \beta_j N_{ij} &= \lambda \int_{\mathcal{D}} F(\varphi) : \nabla h^{(i)} + \lambda \xi \cdot \int_{\mathcal{D}} \nabla h^{(i)} \cdot \varphi, \quad i = 1,2,3 \\ \alpha_j N_{ji} + \beta_j O_{ij} &= \lambda \int_{\mathcal{D}} F(\varphi) : \nabla H^{(i)} + \lambda \xi \cdot \int_{\mathcal{D}} \nabla H^{(i)} \cdot \varphi, \quad i = 1,2,3. \end{aligned} \tag{4.6}$$

while u satisfies

$$\left.\begin{aligned} \Delta u + \lambda \xi \cdot \nabla u &= \lambda \operatorname{div} F(\varphi) + \nabla p \\ \operatorname{div} u &= 0 \end{aligned}\right\} \ \text{in } \mathcal{D}$$
$$u = \alpha_j g^{(j)} + \beta_j G^{(j)} \ \text{ at } \Sigma \tag{4.7}$$
$$\lim_{|x| \to \infty} u(x) = 0.$$

We have the following.

Lemma 4.1. *There is a positive constant $C = C(\mathcal{B}, q)$ such that if $\lambda|\xi| < C$, the map $\mathcal{N}$ is a contraction on $\mathcal{X}^q_\delta$, for $\delta = |\xi|$.*

Proof. Applying Theorems VII.7.1 and VII.7.2 of [7] to (4.7), we obtain

$$\langle\!\langle u \rangle\!\rangle_{\lambda,q} + \|\nabla p\|_q \le c\lambda \left[\| \operatorname{div} F(\varphi)\|_q + \|F(\varphi)\|_2 + (|\alpha| + |\beta|)/\lambda \right]. \tag{4.8}$$

Using the Hölder and Sobolev inequalities, we readily deduce

$$\begin{aligned} \| \operatorname{div} F(\varphi)\|_q &\le c \|\varphi\|_{2q/(2-q)} \|\nabla \varphi\|_2 \le c(\lambda|\xi|)^{-1/2} \langle\!\langle \varphi \rangle\!\rangle^2_{\lambda,q} \\ \|F(\varphi)\|_2 &\le c\|\varphi\|^2_4 \le c(\lambda|\xi|)^{-1/2} \langle\!\langle \varphi \rangle\!\rangle^2_{\lambda,q}. \end{aligned} \tag{4.9}$$

The constant c in (4.9) depends only on $\mathcal{B}$, q and C_0 whenever $\lambda|\xi| \le C_0$. Replacing (4.9) into (4.8), we obtain

$$\langle\!\langle u \rangle\!\rangle_{\lambda,q} + \|\nabla p\|_q \le c \left(\lambda^{1/2} |\xi|^{-1/2} \langle\!\langle \varphi \rangle\!\rangle^2_{\lambda,q} + |\alpha| + |\beta| \right). \tag{4.10}$$

Furthermore, recalling that (see [7] Chapter V)

$$\nabla h^{(i)}, \nabla H^{(i)} \in L^{s'}(\mathcal{D}), \quad \text{all } s' > 3/2,$$

by the Hölder inequality we deduce, with $Z^{(i)}$ denoting either $h^{(i)}$ or $H^{(i)}$, $i = 1,2,3$,

$$\left| \int_D F(\varphi) : \nabla Z^{(i)} \right| + \left| \xi \cdot \int_D \nabla Z^{(i)} \cdot \varphi \right| \le c \left((\lambda|\xi|)^{-1/2} \langle\!\langle \varphi \rangle\!\rangle^2_{\lambda,q} + |\xi| \|\nabla Z^{(i)}\|_{s'} \|\varphi\|_s \right),$$

Assume $1 < q < 6/5$, and choose $s' = 2q/(3q-2)$. We then find

$$\left|\int_D F(\varphi) : \nabla Z^{(i)}\right| + \left|\xi \cdot \int_D \nabla Z^{(i)} \cdot \varphi\right| \leq c\left((\lambda|\xi|)^{-1/2}\langle\!\langle\varphi\rangle\!\rangle^2_{\lambda,q} + \lambda^{-1/2}|\xi|^{1/2}\langle\!\langle\varphi\rangle\!\rangle_{\lambda,q}\right). \quad (4.11)$$

Using this latter inequality into (4.6), and recalling that the matrix (3.16) is nonsingular (see Lemma 3.1), we infer for $\lambda|\xi| \leq c$

$$|\alpha| + |\beta| \leq C\left(\lambda^{1/2}|\xi|^{-1/2}\langle\!\langle\varphi\rangle\!\rangle^2_{\lambda,q} + \lambda^{1/2}|\xi|^{1/2}\langle\!\langle\varphi\rangle\!\rangle_{\lambda,q}\right), \quad (4.12)$$

where $C = C(\mathcal{B}, q)$. Collecting (4.8) and (4.12), we arrive at the following inequality

$$\langle\!\langle u\rangle\!\rangle_{\lambda,q} + \|\nabla p\|_q \leq c\left(\lambda^{1/2}|\xi|^{-1/2}\langle\!\langle\varphi\rangle\!\rangle^2_{\lambda,q} + \lambda^{1/2}|\xi|^{1/2}\langle\!\langle\varphi\rangle\!\rangle_{\lambda,q}\right). \quad (4.13)$$

If $\varphi \in X^q_\delta$, from (4.13) we obtain

$$\langle\!\langle u\rangle\!\rangle_{\lambda,q} + \|\nabla p\|_q \leq \delta c\left(\lambda^{1/2}|\xi|^{-1/2}\delta + \lambda^{1/2}|\xi|^{1/2}\right),$$

and so, if

$$\lambda|\xi| < (1/2c)^2, \quad (4.14)$$

we may choose

$$\delta = |\xi|, \quad (4.15)$$

and we prove that $\mathcal{N}$ transforms X^q_δ into itself. Once this has been established, it is easy to show that $\mathcal{N}$ is, in fact, a contraction. Actually, setting $\phi = \varphi_2 - \varphi_1$, $\varphi_1, \varphi_2 \in X^q_\delta$, from (4.3) and $(4.13)_2$ we obtain

$$\begin{aligned}\|\operatorname{div}(F(\varphi_2) - F(\varphi_1))\|_q &\leq c\left(\|\varphi_1\|_{2q/(2-q)}\|\nabla\phi\|_2 + \|\phi\|_{2q/(2-q)}\|\nabla\varphi_2\|_2\right)\\ &\leq c(\lambda|\xi|)^{-1/2}\delta\langle\!\langle\phi\rangle\!\rangle_{\lambda,q}\end{aligned}$$

$$\|F(\varphi_2) - F(\varphi_1)\|_2 \leq c\left((\|\varphi_2\|_4 + \|\varphi_1\|_4)\|\phi\|_4 \leq c(\lambda|\xi|)^{-1/2}\delta\langle\!\langle\phi\rangle\!\rangle_{\lambda,q}\right). \quad (4.16)$$

Thus, setting $w = u_2 - u_1 \equiv \mathcal{N}(\varphi_2) - \mathcal{N}(\varphi_2)$, $\pi = p_2 - p_1$, $A = \alpha_2 - \alpha_1$, $B = \beta_2 - \beta_1$ (with the obvious meaning for the symbols), from (4.7), (4.16) and Theorems VII.7.1 and VII.7.2 of [7], we find

$$\langle\!\langle w\rangle\!\rangle_{\lambda,q} + \|\nabla\pi\|_q \leq c\left(\lambda^{1/2}|\xi|^{-1/2}\delta\langle\!\langle\phi\rangle\!\rangle_{\lambda,q} + |A| + |B|\right). \quad (4.17)$$

Moreover, from (4.6), by an argument similar to that leading to (4.12), we obtain

$$|A| + |B| \leq c\lambda^{1/2}|\xi|^{-1/2}\delta\langle\!\langle\varphi\rangle\!\rangle^2_{\lambda,q} \quad (4.18)$$

which, once replaced into (4.17), furnishes

$$\langle\!\langle w \rangle\!\rangle_{\lambda,q} \leq c\lambda^{1/2}|\xi|^{-1/2}\delta\langle\!\langle w \rangle\!\rangle_{\lambda,q}.$$

Therefore, with the choice (4.15), if λ and ξ satisfy conditions of the type (4.14), this latter inequality implies that $\mathcal{N}$ is a contraction on $\mathcal{X}^q_\delta$, and the lemma is proved. □

From the previous lemma, we can obtain an existence result for Problem $\mathcal{P}$. To this end, for $w \in \mathcal{T}(\mathcal{B})$, we set

$$\|w\|_{\mathcal{T}} =: \sum_{i=1}^{3} \left(|\alpha_i| + |\beta_i|\right),$$

where

$$\alpha_i = \int_\Sigma w \cdot g^{(i)}, \;\; \beta_i = \int_\Sigma w \cdot G^{(i)}, \quad i = 1, 2, 3.$$

We have.

Theorem 4.1. *Let $\xi \neq 0$, be given and let $1 < q < 6/5$. Then, there exists $C = C(\mathcal{B}, q) > 0$, such that if $\lambda|\xi| \leq C$, Problem $\mathcal{P}$ admits at least one solution v, p, with $v_* \in \mathcal{T}(\mathcal{B})$ and such that $v, p \in C^\infty(\mathcal{D})$,*

$$(v - \xi) \in L^{2q/(2-q)}(\mathcal{D}), \quad \nabla v \in L^{4q/(4-q)}(\mathcal{D}) \cap L^2(\mathcal{D}), \quad D^2 v \in L^q(\mathcal{D})$$

$$p \in L^{3q/(3-q)}(\mathcal{D}), \nabla p \in L^q(\mathcal{D}).$$

Moreover, the following estimates hold

$$\begin{gathered} \langle\!\langle v - \xi \rangle\!\rangle_{\lambda,q} + \|\nabla p\|_q \leq c_1|\xi| \\ |\xi| \leq c_2\|v_*\|_{\mathcal{T}} \leq c_3|\xi| \end{gathered} \tag{4.19}$$

where $\langle\!\langle \, \rangle\!\rangle_{\lambda,q}$ is defined in (4.5), and $c_i = c_i(\mathcal{B}, q)$, $i = 1, 2, 3$.

Proof. The smoothness of v, p comes from known results on the regularity of solutions to the Navier-Stokes equations in exterior domains, [8] Theorem IX.1.1. Thus, in view of Lemma 4.1, to prove the result completely, it remains to show the second inequality in (4.19). Multiplying $(3.2)_1$ by $v - \xi$ and integrating by parts over $\mathcal{D}$, we find

$$\int_\Sigma (v_* - \xi) \cdot g^{(i)} = 2\int_{\mathcal{D}} D(h^{(i)}) : D(v), \quad i = 1, 2, 3. \tag{4.20}$$

The matrix (3.16) is not singular and so, from (4.20) we find

$$\|v_*\|_{\mathcal{T}} \leq c\left(|\xi| + \|\nabla v\|_2\right),$$

which in conjunction with $(4.19)_1$ allows us to conclude

$$\|v_*\|_{\mathcal{T}} \leq c|\xi|.$$

Let us now prove the reverse inequality. Since the matrix K defined in (3.13) is nonsingular, [9] Chapter 5, recalling that $u = v - \xi$, from (4.2) we find

$$|\xi| \le c \left(\|v_*\|_{\mathcal{T}} + \lambda \max_i \left| \int_D F(v-\xi) : \nabla h^{(i)} + \xi \cdot \int_D \nabla h^{(i)} \cdot (v-\xi) \right| \right).$$

Using (4.11) in this inequality we deduce

$$|\xi| \le c \left(\|v_*\|_{\mathcal{T}} + \lambda^{1/2} |\xi|^{-1/2} \langle\!\langle v-\xi \rangle\!\rangle^2_{\lambda,q} + \lambda^{1/2} |\xi|^{1/2} \langle\!\langle v-\xi \rangle\!\rangle_{\lambda,q} \right)$$

and so, with the help of $(4.19)_1$, from this latter relation we conclude

$$|\xi| \le c \left(\|v_*\|_{\mathcal{T}} + \lambda^{1/2} |\xi|^{3/2} \right).$$

Choosing $\lambda^{1/2} |\xi|^{1/2} c < 1$ we obtain

$$|\xi| \le c \|v_*\|_{\mathcal{T}}$$

and the proof of the theorem is completed. □

Our next objective is to investigate uniqueness for Problem $\mathcal{P}$. In this regard, we propose the following result whose proof is similar to Theorem 4.2 of [6] and therefore it will be omitted.

Lemma 4.2. *Let v, p be a solution to (2.1), with $\nabla v \in L^2(\mathcal{D})$, corresponding to $\xi \neq 0$ and $v_* \in W^{2-1/q_0, q_0}(\Sigma)$, $q_0 > 3$. Furthermore, let $q \in (1, 3/2)$. Then, there exists a positive constant $\lambda_1 = \lambda_1(\mathcal{B}, q, q_0)$ such that, if*

$$\lambda (\|v_*\|_{2-1/q_0, q_0(\Sigma)} + |\xi|) \le \lambda_1,$$

we have

$$(v-\xi) \in L^{2q/(2-q)}(\mathcal{D}), \quad (v-\xi)(1+|x|) \in L^\infty(\mathcal{D}), \quad \nabla v \in L^{4q/(4-q)}(\mathcal{D})$$

and the following estimate holds

$$(\lambda|\xi|)^{1/2} \|v-\xi\|_{2q/(2-q)} + (\lambda|\xi|)^{1/4} \|\nabla v\|_{4q/(4-q)} + \|\nabla v\|_2 + \|(v-\xi)(1+|x|)\|_\infty \le c(\|v_*\|_{2-1/q_0, q_0, \Sigma} + |\xi|).$$

with $c = c(\mathcal{B}, q, q_0)$.

We are now in a position to prove the following uniqueness result.

Theorem 4.2. *Let $\mathcal{S}_\xi$ be the class of solutions v, p to (2.1)–(2.3) corresponding to a given $\xi \neq 0$, such that*

(i) $\nabla v \in L^2(\mathcal{D})$;
(ii) $v_* \in \mathcal{T}(\mathcal{B})$;

(iii) $\|v_*\|_{\mathcal{T}} \leq C_0|\xi|$ *for some* $C_0 > 0$.

Then, there exists $C = C(\mathcal{B}, C_0) > 0$ *such that if* $\lambda|\xi| < C$, $\mathcal{S}_\xi$ *is constituted by at most one element.*

Proof. Let v, p and u, p_1 be two elements of $\mathcal{S}_\xi$ and let

$$U = u - v, \quad \pi = p_1 - p.$$

From (2.1) and (4.2) we obtain

$$\left.\begin{aligned} \Delta U - \lambda\xi\cdot\nabla U = \lambda\left[U\cdot\nabla u + (v-\xi)\cdot\nabla U\right] + \nabla\pi \\ \operatorname{div} U = 0 \end{aligned}\right\} \text{ in } \mathcal{D}$$
$$\lim_{|x|\to\infty} U(x) = 0 \tag{4.21}$$

and $(i = 1, 2, 3)$

$$\begin{aligned} \int_\Sigma U_*\cdot g^{(i)} &= \lambda\int_D \left[(v-\xi)\cdot\nabla h^{(i)}\cdot U + U\cdot\nabla h^{(i)}\cdot(u-\xi)\right] + \lambda\xi\cdot\int_D \nabla h^{(i)}\cdot U \\ \int_\Sigma U_*\cdot G^{(i)} &= \lambda\int_D \left[(v-\xi)\cdot\nabla H^{(i)}\cdot U + U\cdot\nabla H^{(i)}\cdot(u-\xi)\right] + \lambda\xi\cdot\int_D \nabla H^{(i)}\cdot U. \end{aligned} \tag{4.22}$$

where U_* is the restriction of U at Σ. Set

$$\langle U\rangle_{\lambda,q} \equiv (\lambda|\xi|)^{1/2}\|U\|_{2q/(2-q)} + \|U\|_{3q/(3-2q)} + \|\nabla U\|_2.$$

Applying Theorem VII.7.1 of [7] to (4.21), and using the Hölder inequality, Lemma 4.2 and the assumptions (i)–(iii), we find

$$\begin{aligned} \langle U\rangle_{\lambda,q} &\leq c(\lambda\|U\cdot\nabla u + (v-\xi)\cdot\nabla U\|_q + \|U_*\|_{\mathcal{T}}) \\ &\leq c\left(\lambda\|U\|_{2q/(2-q)}\|\nabla u\|_2 + \lambda\|v-\xi\|_{2q/(2-q)}\|\nabla U\|_2 + \|U_*\|_{\mathcal{T}}\right) \\ &\leq c\left(\lambda^{1/2}|\xi|^{1/2}\langle U\rangle_{\lambda,q} + \|U_*\|_{\mathcal{T}}\right). \end{aligned} \tag{4.23}$$

Furthermore, from (4.22) we obtain, with $Z^{(i)} = h^{(i)}, H^{(i)}$,

$$\begin{aligned} \left|\int_\Sigma U_*\cdot Z^{(i)}\right| \leq \lambda\|\nabla Z^{(i)}\|_{6q/(13q-12)}\|U\|_{3q/(3-2q)}(\|u-\xi\|_{2q/(2-q)} + \\ \|v-\xi\|_{2q/(2-q)}) + \lambda|\xi|\,\|\nabla Z^{(i)}\|_{2q/(3q-2)}\|U\|_{2q/(2-q)}. \end{aligned} \tag{4.24}$$

Recalling that $\nabla Z^{(i)} \in L^r(\mathcal{D})$ for all $r > 3/2$, [7], we choose $q \in (1, 12/9)$ and obtain from (4.24) the following inequality

$$\|U_*\|_{\mathcal{T}} \leq c\lambda^{1/2}|\xi|^{1/2}\langle U\rangle_{\lambda,q}. \tag{4.25}$$

Replacing (4.25) into (4.23), and taking $\lambda|\xi|$ less than a suitable constant depending only on $\mathcal{B}$ and C_0, we prove uniqueness. □

Remark 4.1. Theorem 4.1 ensures that for any $\xi \neq 0$, the class $\mathcal{S}_\xi$ defined in Theorem 4.2 is not empty.

Theorems 4.1 and 4.2 prove the existence of a map $\mathcal{M}$ from the space T of translational motions of $\mathcal{B}$ onto a subspace $\mathcal{T}'(\mathcal{B})$ of $\mathcal{T}(\mathcal{B})$. We know from the linear theory of Section 3 that $\mathcal{T}'(\mathcal{B})$ is expected to be *strictly* contained into $\mathcal{T}(\mathcal{B})$, due to the fact that a velocity distribution in $\mathcal{T}(\mathcal{B})/\mathcal{T}'(\mathcal{B})$ will produce, in general, also a rotation for $\mathcal{B}$. However, we shall show in the next theorem that the map $\mathcal{M}$ is in fact one-to-one on $\mathcal{T}'(\mathcal{B})$.

Theorem 4.3. *Let v, p and v_1, p_1 be two solutions to Problem $\mathcal{P}$, as given in Theorem 4.1, corresponding to ξ and ξ_1, respectively, with $\xi \neq \xi_1$. Let v_* and v_{1*} be their restrictions at Σ. Then, there exists a positive constant $C = C(\mathcal{B})$ such that if*

$$\lambda|\xi| < C, \tag{4.26}$$

necessarily $v_* \not\equiv v_{1*}$

Proof. Assume, by contradiction, $v_* \equiv v_{1*}$, and let

$$u = v - \xi, \quad u_1 = v_1 - \xi_1, \quad \mu = \xi - \xi_1$$
$$U = u - u_1, \quad \pi = p - p_1$$

We then obtain

$$\left.\begin{aligned} \Delta U - \lambda\xi \cdot \nabla U &= \lambda\left(\mu \cdot \nabla u_1 + U \cdot \nabla u + u_1 \cdot \nabla U\right) + \nabla\pi \\ \operatorname{div} U &= 0 \end{aligned}\right\} \quad \text{in } \mathcal{D}$$
$$U = -\mu \quad \text{on } \Sigma \tag{4.27}$$
$$\lim_{|x|\to\infty} U(x) = 0.$$

Moreover, from $(4.1)_1$, we find that

$$\mu \cdot \int_\Sigma g^{(i)} = \lambda \int_{\mathcal{D}} U \cdot \nabla h^{(i)} \cdot u + \lambda \int_{\mathcal{D}} u_1 \cdot \nabla h^{(i)} \cdot U + \lambda\xi \cdot \int_{\mathcal{D}} \nabla U \cdot h^{(i)} + \lambda\mu \cdot \int_{\mathcal{D}} \nabla h^{(i)} \cdot u_1, \quad i = 1, 2, 3. \tag{4.28}$$

Applying Theorem VII.7.1 of [7] to (4.27) we find for $q \in (1, 3/2)$

$$\begin{aligned} &\lambda\|\xi \cdot \nabla U\|_q + (\lambda|\xi|)^{1/2}\|U\|_{2q/(2-q)} + (\lambda|\xi|)^{1/4}\|\nabla U\|_{4q/(4-q)} + \|\nabla U\|_{3q/(3-q)} \\ &\leq c\lambda\left(|\mu|\|\nabla u_1\|_q + \|U\|_{2q/(2-q)}\|\nabla u\|_2 + \|u_1\|_4\|\nabla U\|_{4q/(4-q)} + \frac{|\mu|}{\lambda}\right). \end{aligned} \tag{4.29}$$

From Lemma 4.2 and $(4.19)_2$, we have

$$\|u_1\|_4 + \|\nabla u\|_2 \le c\|v_*\|_{\mathcal{T}},$$

and, for $q \in (4/3, 3/2)$,

$$\|\nabla u_1\|_q \le c(\lambda\|v_*\|_{\mathcal{T}})^{-1/4}\|v_*\|_{\mathcal{T}}.$$

Thus, from (4.29) and $(4.19)_2$ we recover the following inequality

$$\begin{aligned}\lambda\|\xi\cdot\nabla U\|_q + (\lambda\|v_*\|_{\mathcal{T}})^{1/2}\|U\|_{2q/(2-q)} + (\lambda\|v_*\|_{\mathcal{T}})^{1/4}\|\nabla U\|_{4q/(4-q)} + \\ \|\nabla U\|_{3q/(3-q)} \le c\lambda\Big[|\mu|(\lambda\|v_*\|_{\mathcal{T}})^{-1/4}\|v_*\|_{\mathcal{T}} + \\ \|v_*\|_{\mathcal{T}}\left(\|U\|_{2q/(2-q)} + \|\nabla U\|_{4q/(4-q)}\right) + |\mu|/\lambda\Big]. \quad (4.30)\end{aligned}$$

Moreover, from the Hölder inequality, we also obtain

$$\begin{aligned}\left|\int_{\mathcal{D}} U\cdot\nabla h^{(i)}\cdot u\right| + \left|\int_{\mathcal{D}} u_1\cdot\nabla h^{(i)}\cdot U\right| \le \\ \|U\|_{2q/(2-q)}\|\nabla h^{(i)}\|_2\left(\|u_1\|_{q/(q-1)} + \|u\|_{q/(q-1)}\right)\end{aligned}$$

and, since by Lemma 4.2 and $(4.19)_2$, for $q < 3/2$ it is

$$\|u_1\|_{q/(q-1)} + \|u\|_{q/(q-1)} \le c\left(\|u_1(1+|x|)\|_\infty + \|u(1+|x|)\|_\infty\right) \le c\|v_*\|_{\mathcal{T}},$$

we obtain

$$\left|\int_{\mathcal{D}} U\cdot\nabla h^{(i)}\cdot u\right| + \left|\int_{\mathcal{D}} u_1\cdot\nabla h^{(i)}\cdot U\right| \le c\|U\|_{2q/(2-q)}\|v_*\|_{\mathcal{T}} \quad (4.31)$$

Also, again from Lemma 4.2 and $(4.19)_2$, for $s \in (2,3)$ we find

$$\lambda\left|\mu\cdot\int_{\mathcal{D}}\nabla h^{(i)}\cdot u_1\right| \le c\lambda|\mu|\|u\|_s\|\nabla h^{(i)}\|_{s/(s-1)} \le c|\mu|(\lambda\|v_*\|_{\mathcal{T}})^{1/2}. \quad (4.32)$$

Finally, for $q < q_1 < 3/2$ we have

$$\left|\int_{\mathcal{D}}\xi\cdot\nabla U\cdot h^{(i)}\right| \le \|\xi\cdot\nabla U\|_{q_1}\|h^{(i)}\|_{q_1/(q_1-1)} \le C\|\xi\cdot\nabla U\|_{q_1},$$

and, by the convexity inequality,

$$\|\xi\cdot\nabla U\|_{q_1} \le \|\xi\cdot\nabla U\|_q^{\theta}\|\xi\cdot\nabla U\|_{3q/(3-q)}^{1-\theta}, \quad \theta\in(0,1).$$

Thus, by $(4.19)_2$ and Young's inequality we deduce

$$\begin{aligned}\lambda\left|\int_{\mathcal{D}}\xi\cdot\nabla U\cdot h^{(i)}\right| &\le c(\lambda|\xi|)^{1-\theta}\left(\lambda^{\theta}\|\xi\cdot\nabla U\|_q^{\theta}\|\nabla U\|_{3q/(3-q}^{1-\theta}\right) \\ &\le c(\lambda\|v_*\|_{\mathcal{T}})^{1-\theta}\left(\lambda\|\xi\cdot\nabla U\|_q + \|\nabla U\|_{3q/(3-q)}\right).\end{aligned} \quad (4.33)$$

Collecting (4.31)–(4.33) and using (4.28), we find that there exists a constant $C = C(\mathcal{B}) > 0$ such that if (4.26) holds then

$$|\mu| \le c\left[\lambda\|v_*\|_{\mathcal{T}}\|U\|_{2q/(2-q)} + (\lambda\|v_*\|_{\mathcal{T}})^{1-\theta}\left(\lambda\|\xi\cdot\nabla U\|_q + \|\nabla U\|_{3q/(3-q)}\right)\right].$$

Replacing this inequality into (4.30), it is immediate to show that there exists a positive constant C depending only on $\mathcal{B}$ such that if (4.26) is satisfied, we then get

$$(\lambda\|v_*\|_{\mathcal{T}})^{1/2}\|U\|_{2q/(2-q)} + (\lambda\|v_*\|_{\mathcal{T}})^{1/4}\|\nabla U\|_{4q/(4-q)} \le 0,$$

which implies, in particular, $\xi = \xi_1$, which contradicts the assumption. The theorem is completely proved. □

References

1. Adams, R. A. 1975, *Sobolev Spaces*, Academic Press, New York
2. Blake, J. R., and Otto, S. R., 1996, Ciliary Propulsion, Chaotic Filtration and a 'Blinking' Stokeslet, *J. Engineering Math.*, **30**, 151–168
3. Brennen, C., and Winet, H., 1977, Fluid Mechanics of Propulsion by Cilia and Flagella, *Ann. Rev. Fluid Mech.*, **9**, 339–398
4. Birkhoff, G., and Zarantonello, E. H., 1957, *Jets, Wakes, and Cavities,* Academic Press.
5. Finn, R., 1965, On the Exterior Stationary Problem for the Navier-Stokes Equations, and Associated Perturbation Problems, *Arch. Rational Mech. Anal.,* **19**, 363–406
6. Galdi, G. P., 1998, On the Steady, Translational Self-Propelled Motion of a Symmetric Body in a Navier-Stokes Fluid, *Quaderni di Matematica della II Università di Napoli*, Vol I, to be published
7. Galdi, G. P., 1994, *An Introduction to the Mathematical Theory of the Navier-Stokes Equations: Linearized Steady Problems*, Springer Tracts in Natural Philosophy, Vol. 38, Springer-Verlag
8. Galdi, G. P., 1994, *An Introduction to the Mathematical Theory of the Navier-Stokes Equations: Nonlinear Steady Problems*, Springer Tracts in Natural Philosophy, Vol. 39, Springer-Verlag
9. Happel, V., and Brenner, H., 1965, *Low Reynolds Number Hydrodynamics*, Prentice Hall.
10. Kozono, H., and Sohr, H., 1993, On Stationary Navier-Stokes Equations in Unbounded Domains, *Ricerche Mat.*, **42**, 69–86.
11. Kozono, H., Sohr, H., and Yamazaki, M., 1997, Representation Formula, Net Force and Energy Relation to the Stationary Navier-Stokes Equations in 3-Dimensional Exterior Domains, *Kyushu J. Math.*, **51**, 239–260.
12. Lugovtsov, A. A., and Lugovtsov, B. A., 1971, Example of a Viscous Incompressible Flow Past a Body with Moving Boundary, *Dynamics of Continuous Media*, Novosibirsk, **8**, 49–55 (in Russian)
13. Pukhnachev, V. V., 1989, Asymptotics of a Velocity Field at Considerable Distances From a Self-Propelled Body, *J. Appl. Mech. Tech. Phys.*, **30**, 52–60.
14. Pukhnachev, V. V., 1990, Stokes Approximation in a Problem of the Flow Around a Self-Propelled Body, *Boundary Value Problems in Mathematical Physics*, Naukova Dumka, Kiev, 65–73 (in Russian)

15. Pukhnachev, V. V., 1990, The Problem of Momentumless Flow for the Navier-Stokes Equations, *Springer Lecture Notes in Mathematics*, **1431**, Springer-Verlag, 87–94
16. Sennitskii, V. L., 1978, Liquid Flow Around a Self-Propelled Body, *J. Appl. Mech. Tech. Phys.*, **3**, 15–27
17. Sennitskii, V. L., 1984, An Example of Axisymmetric Fluid Flow Around a Self-Propelled Body, *J. Appl. Mech. Tech. Phys.*, **25**, 526–530
18. Sennitskii, V. L., 1990, Self-Propulsion of a Body in a Fluid, *J. Appl. Mech. Tech. Phys.*, **31**, 266–272
19. Shapere, A. and Wilczek, F., 1989, Geometry of Self-Propulsion at Low Reynolds Number, *J. Fluid Mech.*, **198**, 557–585
20. Taylor, G.I., 1951, Analysis of the Swimming of Microscopic Organisms, *Proc. Royal Soc. London A,* **209**, 447–461.

EQUADIFF 9 CD ROM, Brno 1997
Masaryk University

PROCEEDINGS
pp. 81–96

Weak Stabilization of Solutions to PDEs with Hysteresis in Thermovisco-Elastoplasticity

Pavel Krejčí and Jürgen Sprekels

Weierstrass Institute for Applied Analysis and Stochastics
Mohrenstr. 39
10117 Berlin, Germany
Email: krejci@wias-berlin.de
sprekels@wias-berlin.de
WWW: http://www.wias-berlin.de

Abstract. We present a thermodynamically consistent description of the uniaxial behavior of thermovisco-elastoplastic materials for which the total stress σ contains, in addition to elastic, viscous and thermic contributions, a plastic component $\sigma^p(x,t) = \mathcal{P}[\varepsilon(x,\cdot),\theta(x,t)](t)$. Here, ε and θ are the fields of strain and absolute temperature, respectively, and $\{\mathcal{P}[\cdot,\theta]\}_{\theta>0}$ denotes a family of (rate-independent) hysteresis operators of Prandtl-Ishlinskii type, parametrized by the absolute temperature. The momentum and energy balance equations governing the space-time evolution of the material form a system of two highly nonlinearly coupled partial differential equations involving partial derivatives of hysteretic nonlinearities at different places. It is shown that under no external forcing, the unique global strong solution of a corresponding initial-boundary value problem remains bounded in the energy norm and the velocity asymptotically vanishes for large times.

AMS Subject Classification. 73B30, 73E60, 35Q72, 35B40

Keywords. Thermoplasticity, viscoelasticity, hysteresis operators, nonlinear PDEs, asymptotic behavior

1 Introduction

For many materials the stress-strain (σ - ε) relations measured in uniaxial load-deformation experiments strongly depend on the absolute (Kelvin) temperature θ and, at the same time, exhibit a strong plastic behavior witnessed by the occurrence of rate-independent hysteresis loops. Figure 1 shows a typical diagram, where the elasticity modulus and the yield limit depend on temperature.

Among the materials exhibiting temperature-dependent, but rate-independent hysteretic effects are shape memory alloys (see, for instance, Chapter 5 in [1]) and even, although to a smaller extent, quite ordinary steels.

If the σ - ε relation exhibits a hysteresis, it can no longer be expressed in terms of simple single-valued functions since the latter are certainly not able to

give a correct account of the inherent memory structures that are responsible for the complicated loopings in the interior of experimentally observed hysteresis loops.

To avoid these difficulties, a different approach to thermoelastoplastic hysteresis based on the notion of *hysteresis operators* introduced by the Russian group around M. A. Krasnoselskii in the seventies (see [5]) has been proposed by the authors in [7]. The temperature-dependent plastic stress σ^p has been assumed in the form of an *operator equation* with a temperature-dependent hysteretic constitutive operator $\mathcal{P}$ of *Prandtl-Ishlinskii type*, namely

$$\sigma^p \;=\; \mathcal{P}[\varepsilon,\theta] \;:=\; \int_0^\infty \varphi(r,\theta)\,\mathfrak{s}_r[\varepsilon]\,dr\,. \tag{1.1}$$

In this connection, $\mathfrak{s}_r$ denotes the so-called *stop operator* or *elastic-plastic element* with threshold $r>0$ (to be defined in the next section), and $\varphi(\cdot,\theta)\ge 0$ is a density function with respect to $r>0$, parameterized by the absolute temperature θ. The integral formula (1.1) corresponds to an infinite rheological combination in parallel of elements $\mathfrak{s}_r$.

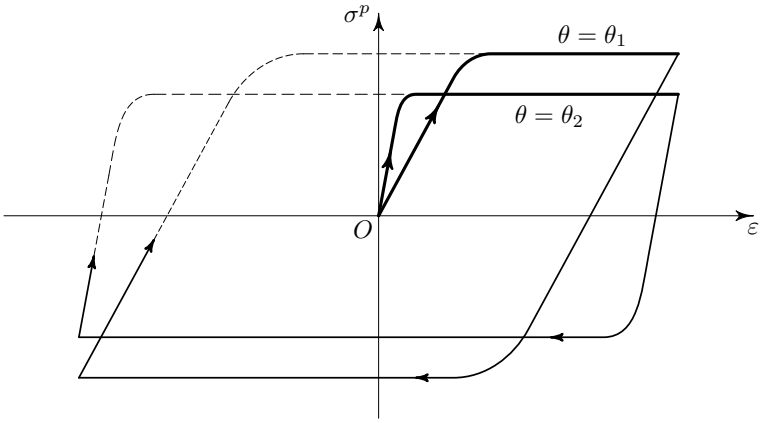


Fig. 1: *Strain – plastic stress diagrams at constant temperatures* $\theta_1\neq\theta_2$.

The advantage of this approach is that an operator equation like (1.1) is suited much better than a simple functional relation to keep track of *memory effects* imprinted on the material in the past history; in fact, the output at any time $t\in[0,T]$ may depend on the whole evolution of the input in the time interval $[0,t]$. Observe that the requirement of rate-independence implies that $\mathcal{P}$ cannot be expressed in terms of an integral operator of convolution type, i. e. we are not dealing with a model with fading memory.

In the isothermal case, i. e. if $\mathcal{P}$ is independent of θ, the space-time evolution is governed by the equation of motion which is of hyperbolic type, see e. g. [6]. In the temperature-dependent case, the equation of motion has to be complemented by a field equation representing the balance law of internal energy, and the second

principle of thermodynamics in form of the Clausius-Duhem inequality must be obeyed. The first problem then consists in a correct definition of thermodynamic state functions like the densities of free energy, internal energy and entropy. It is natural to expect that they will be given in the form of *operators* rather than of *functions*.

A corresponding construction has been carried out in [7,8]. It turns out, however, that we are no longer able to solve the hyperbolic case, and a further regularization is necessary. While [7] is devoted to the case when the total stress σ is composed of a plastic stress σ^p of the form (1.1) and a so-called *couple stress*, [7] deals with the situation when σ comprises, in addition to the plastic stress (1.1), (nonlinear) elastic, (linear) viscous, and (linear) thermic contributions σ^e, σ^v and σ^d, respectively; that is, we assume a constitutive law of the form

$$\sigma = \sigma^p + \sigma^e + \sigma^v + \sigma^d\,, \tag{1.2}$$

with σ^p given as in (1.1).

It should be mentioned at this place that hysteretic relations can usually not be described in an explicit form and, as a rule, enjoy only very restricted smoothness properties. Therefore, the classical techniques of one-dimensional thermovisco-elasticity developed for cases in which the stress-strain relation is given through a simple (possibly nonconvex, but differentiable) function (we only refer to the fundamental papers [2,3]) apply only partially, and new techniques tailored to deal with the specific behavior of hysteretic nonlinearities need to be employed.

The paper is organized as follows. In Section 2, the field equations governing the space-time evolution in thermovisco-elastoplastic materials with the constitutive law (1.2) are derived. We obtain a system of nonlinearly coupled partial differential equations involving partial derivatives of hysteretic nonlinearities at different places, even in derivatives of highest order. Section 3 contains a summary of results of [8] on existence, uniqueness and thermodynamic consistency of solutions and their continuous dependence on given data. In Section 4, we present a new result on weak asymptotic stabilization. Section 5 is an appendix, where we derive an auxiliary convergence theorem.

2 Thermoelastoplastic constitutive laws

The stop operator $\mathfrak{s}_r : W^{1,1}(0,T) \to W^{1,1}(0,T)$ in the equation (1.1) is defined as the solution operator $\sigma_r = \mathfrak{s}_r[\varepsilon]$ of the variational inequality

$$|\sigma_r(t)| \le r, \quad (\dot\varepsilon(t) - \dot\sigma_r(t))(\sigma_r(t) - \tilde\sigma) \ge 0 \quad \text{for a.e. } t \in]0,T[\,, \quad \forall \tilde\sigma \in [-r,r], \tag{2.1}$$

with initial condition

$$\sigma_r(0) = \operatorname{sign}(\varepsilon(0)) \min\{r, |\varepsilon(0)|\} \tag{2.2}$$

which describes the strain-stress law of Prandtl's model for elastic-perfectly plastic materials with a unit elasticity modulus and yield point r, see Fig. 2.

The density function φ in (1.1) is assumed to be given. It can be identified from the isothermal initial loading curves $\sigma = \Phi(\varepsilon, \theta)$ obtained experimentally by letting ε monotonically increase for each fixed temperature θ starting from the origin. The corresponding formula reads (see [6])

$$\Phi(\varepsilon, \theta) = \int_0^\varepsilon \int_s^\infty \varphi(r, \theta)\, dr\, ds. \tag{2.3}$$

We consider here only the case when φ is nonnegative, i.e. the initial loading curves at each constant temperature are concave and nondecreasing as on Fig. 1.

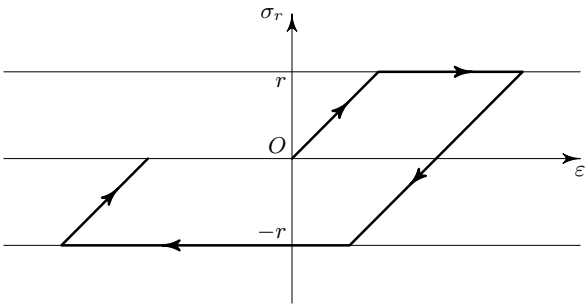


Fig. 2: *Prandtl's normalized elastic-perfectly plastic element*

The operator $\mathfrak{s}_r$ has following properties (for a proof, see [1], [6]).

Proposition 1. *Let $r > 0$ be given. Then it holds:*

(i) *For every $\varepsilon \in W^{1,1}(0,T)$, we have*

$$\left(\frac{d}{dt}\mathfrak{s}_r[\varepsilon]\right)^2 = \dot{\varepsilon}\frac{d}{dt}\mathfrak{s}_r[\varepsilon] \quad \text{a.e. in }]0,T[. \tag{2.4}$$

(ii) *For every $\varepsilon_1, \varepsilon_2 \in W^{1,1}(0,T)$, we have*

$$\frac{1}{2}\frac{d}{dt}\left(\mathfrak{s}_r[\varepsilon_1] - \mathfrak{s}_r[\varepsilon_2]\right)^2 \le (\dot{\varepsilon}_1 - \dot{\varepsilon}_2)(\mathfrak{s}_r[\varepsilon_1] - \mathfrak{s}_r[\varepsilon_2]) \quad \text{a.e. in }]0,T[, \tag{2.5}$$

$$\int_0^T \left|\frac{d}{dt}(\mathfrak{s}_r[\varepsilon_1] - \mathfrak{s}_r[\varepsilon_2])\right|(t)\, dt \le |\varepsilon_1(0) - \varepsilon_2(0)| + 2\int_0^T |\dot{\varepsilon}_1 - \dot{\varepsilon}_2|\,(t)\, dt, \tag{2.6}$$

$$|(\mathfrak{s}_r[\varepsilon_1] - \mathfrak{s}_r[\varepsilon_2])(t)| \le 2\max_{0\le\tau\le t}|\varepsilon_1(\tau) - \varepsilon_2(\tau)| \quad \forall t \in [0,T]. \tag{2.7}$$

(iii) *For every $r, q > 0$ and $\varepsilon \in W^{1,1}(0,T)$, we have*

$$|(\mathfrak{s}_r[\varepsilon] - \mathfrak{s}_q[\varepsilon])(t)| \le |r - q| \quad \forall t \in [0,T]. \tag{2.8}$$

The inequalities (2.6), (2.7) entail that the stop operator $\mathfrak{s}_r$ is Lipschitz continuous in $W^{1,1}(0,T)$ and admits a Lipschitz continuous extension onto $C([0,T])$. Moreover, we immediately see by definition that $\mathfrak{s}_r$ is a *causal* operator, that is, we have the implication

$$\varepsilon_1(\tau) = \varepsilon_2(\tau) \quad \forall \tau \in [0,t] \qquad \Rightarrow \qquad \mathfrak{s}_r[\varepsilon_1](t) = \mathfrak{s}_r[\varepsilon_2](t) \tag{2.9}$$

for every $t \in [0,T]$, which means that the output values at time t depend only on past values of the input. This enables us to consider $\mathfrak{s}_r$ as a family of operators acting in the spaces $C([0,t])$ for all $t \in]0,T]$.

From inequality (2.5) it immediately follows:

Corollary 2. *For all $\varepsilon, \varepsilon_1, \varepsilon_2 \in W^{1,1}(0,T)$, we have*

$$\mathfrak{s}_r[\varepsilon]\left(\dot{\varepsilon} - \frac{d}{dt}\mathfrak{s}_r[\varepsilon]\right) \geq 0 \quad \text{a.e. in }]0,T[\quad \textit{(energy inequality)}\,, \tag{2.10}$$

$$|(\mathfrak{s}_r[\varepsilon_1] - \mathfrak{s}_r[\varepsilon_2])(t)| \leq |\varepsilon_1(0) - \varepsilon_2(0)| + \int_0^t |\dot{\varepsilon}_1 - \dot{\varepsilon}_2|(\tau)\, d\tau \quad \forall t \in [0,T]. \tag{2.11}$$

In this paper we consider the one-dimensional equation of motion

$$\rho\, u_{tt} = \sigma_x + f, \tag{2.12}$$

where $\rho > 0$ is a constant referential density, u is the displacement, σ is the total unaxial stress and f is the volume force density.

We assume that σ can be decomposed into the sum

$$\sigma = \sigma^p + \sigma^e + \sigma^v + \sigma^d, \tag{2.13}$$

where

$$\sigma^e = \gamma(\varepsilon), \tag{2.14}$$

with a given nondecreasing Lipschitz continuous function $\gamma : \mathbb{R}^1 \to \mathbb{R}^1$, $\gamma(0) = 0$, is the (nonlinear) kinematic hardening component,

$$\sigma^v = \mu\dot{\varepsilon} \tag{2.15}$$

with a constant $\mu > 0$ is the viscous component,

$$\sigma^d = -\beta\theta \tag{2.16}$$

with a constant $\beta \in \mathbb{R}^1$ is the thermic dilation component and σ^p is the thermoplastic component given by (1.1). Equation (2.13) can be interpreted rheologically as a combination in parallel of the above components (see [9]). The stop operator $\mathfrak{s}_r$ is assumed to act on functions of x and t according to the formula

$$\mathfrak{s}_r[\varepsilon](x,t) := \mathfrak{s}_r[\varepsilon(x,\cdot)](t), \tag{2.17}$$

i.e. x plays the role of a parameter. The equation of motion (2.12) has to be coupled with the energy balance equation

$$U_t = \sigma\varepsilon_t - q_x + g, \tag{2.18}$$

where U is the total internal energy, q is the heat flux and g is the heat source density. The model is thermodynamically consistent provided the temperature θ and the entropy S satisfy the inequalities

$$\theta > 0, \tag{2.19}$$

$$S_t \geq \frac{g}{\theta} - \left(\frac{q}{\theta}\right)_x \qquad \text{(Clausius-Duhem inequality)}, \tag{2.20}$$

in an appropriate sense.

In [7] we derived the following expressions for thermoplastic parts of internal energy U^p and entropy S^p in operator form corresponding to the constitutive law (1.1),

$$U^p = \mathcal{V}[\varepsilon,\theta] := \frac{1}{2}\int_0^\infty (\varphi(r,\theta) - \theta\varphi_\theta(r,\theta))\,\mathfrak{s}_r^2[\varepsilon]\,dr, \tag{2.21}$$

$$S^p = \mathcal{S}[\varepsilon,\theta] := -\frac{1}{2}\int_0^\infty \varphi_\theta(r,\theta)\,\mathfrak{s}_r^2[\varepsilon]\,dr. \tag{2.22}$$

In accordance with (2.13), (2.21), (2.22), we put

$$U := C_V\theta + \mathcal{V}[\varepsilon,\theta] + \Gamma(\varepsilon) + V_0, \tag{2.23}$$

$$S := C_V \log\theta + \mathcal{S}[\varepsilon,\theta] + \beta\varepsilon, \tag{2.24}$$

where $C_V > 0$, the purely caloric part of the *specific heat*, is a constant, $V_0 > 0$ is a constant which is chosen in order to ensure that $U \geq 0$ according to Hypothesis (H2) below, and $\Gamma(\varepsilon) := \int_0^\varepsilon \gamma(s)ds$. For the heat flux we assume Fourier's law

$$q = -\kappa\theta_x \tag{2.25}$$

with a constant heat conduction coefficient $\kappa > 0$. We complete the system (2.12), (2.18) with the small deformation hypothesis

$$\varepsilon = u_x \tag{2.26}$$

and rewrite it in the form

$$\rho u_{tt} - (\gamma(u_x) + \mathcal{P}[u_x,\theta] + \mu u_{xt} - \beta\theta)_x = f, \tag{2.27}$$

$$(C_V\theta + \mathcal{V}[u_x,\theta])_t - \kappa\theta_{xx} = (\mathcal{P}[u_x,\theta] + \mu u_{xt} - \beta\theta)\,u_{xt} + g. \tag{2.28}$$

3 Existence, uniqueness and thermodynamic consistency

We consider a model problem for a system of the form (2.27), (2.28), namely

$$u_{tt} - \big(\gamma(u_x)\big)_x - (\mathcal{P}[u_x,\theta])_x - \mu u_{xxt} + \beta\theta_x = f(\theta,x,t), \tag{3.1}$$

$$(C_V\theta + \mathcal{V}[u_x,\theta])_t - \theta_{xx} = \mathcal{P}[u_x,\theta]u_{xt} + \mu u_{xt}^2 - \beta\theta u_{xt} + g(\theta,x,t), \tag{3.2}$$

for $x \in]0,1[$, $t \in [0,T]$, where $T > 0$, $\mu > 0$, $C_V > 0$, $\beta \in \mathbb{R}^1$ are fixed constants, $\gamma\colon \mathbb{R}^1 \to \mathbb{R}^1$, $f, g\colon]0,\infty[\times]0,1[\times[0,T] \to \mathbb{R}^1$ are given functions, and $\mathcal{P}$, $\mathcal{V}$ are the operators defined by (1.1), (2.21) with a given distribution function $\varphi\colon (]0,\infty[)^2 \to [0,\infty[$ satisfying Hypothesis (H2) below.

In other words, we assume in (2.27), (2.28) that the volume force and heat source densities are given functions of x and t which may also depend on the instantaneous value of θ, and we rescale the units in such a way that $\rho \equiv \kappa \equiv 1$. The system (3.1), (3.2) is coupled with boundary and initial conditions which are chosen in the following simple form.

$$u(0,t) = u(1,t) = \theta_x(0,t) = \theta_x(1,t) = 0, \tag{3.3}$$

$$u(x,0) = u^0(x),\ u_t(x,0) = u^1(x),\ \theta(x,0) = \theta^0(x). \tag{3.4}$$

The data are assumed to satisfy the following hypotheses.

Hypothesis (H1).

(i) $u^0, u^1 \in W^{2,2}(0,1) \cap \overset{\circ}{W}{}^{1,2}(0,1)$, $\theta^0 \in W^{1,2}(0,1)$, and there exists a constant $\delta > 0$ such that

$$\theta^0(x) \geq \delta \qquad \forall x \in [0,1]. \tag{3.5}$$

(ii) $\gamma\colon \mathbb{R}^1 \to \mathbb{R}^1$ is an absolutely continuous function, $\gamma(0) = 0$, and there exists a constant $\gamma_0 > 0$ such that

$$0 \leq \frac{d\gamma(\varepsilon)}{d\varepsilon} \leq \gamma_0 \qquad \text{a.e. in } \mathbb{R}^1. \tag{3.6}$$

(iii) The functions f, g are measurable, $f(\cdot,x,t)$, $g(\cdot,x,t)$ are absolutely continuous in $[0,\infty[$ for a.e. $(x,t) \in]0,1[\times]0,T[$. Moreover, there exist a constant $K > 0$ and functions $f_0, g_0 \in L^2(]0,1[\times]0,T[)$ such that

$$g(0,x,t) = g_0(x,t) \geq 0 \qquad a.e., \tag{3.7}$$

$$|f(\theta,x,t)| + |f_t(\theta,x,t)| \leq f_0(x,t) \quad a.e., \tag{3.8}$$

$$|f_\theta(\theta,x,t)| + |g_\theta(\theta,x,t)| \leq K \quad a.e.\ . \tag{3.9}$$

Hypothesis (H2).

The function $\varphi\colon\ ]0,\infty[^2 \to [0,\infty[$ is measurable, $\varphi(r,\cdot)$, $\varphi_\theta(r,\cdot)$ are absolutely continuous for a.e. $r>0$, and there exist constants $L>0$, $V_0>0$ such that for a.e. $\theta>0$ the following inequalities hold.

$$\int_0^\infty \varphi(r,\theta)\,dr \le L, \tag{3.10}$$

$$\int_0^\infty |\varphi_\theta(r,\theta)|\,r\,dr \le L, \tag{3.11}$$

$$\int_0^\infty \theta\,|\varphi_{\theta\theta}(r,\theta)|\,r^2\,dr \le C_V, \tag{3.12}$$

where C_V is the constant introduced in (2.23),

$$\frac{1}{2}\int_0^\infty |\varphi(r,\theta)-\theta\varphi_\theta(r,\theta)|\,(1+r^2)\,dr \ \le\ V_0. \tag{3.13}$$

Example 3. A typical function φ satisfying Hypothesis (H2) can be chosen as

$$\varphi(r,\theta) = \bar E(\theta)\ c(r-\bar r(\theta)), \tag{3.14}$$

where $c\in\mathcal{D}(]-m,m[)$ is a mollifier such that

$$\int_{-m}^m c(s)\,ds = 1, \qquad c\ge 0, \tag{3.15}$$

with a (small) constant $m>0$, and $\bar E,\bar r$ are given functions such that $\bar E(\theta)\le L$, $m\le\bar r(\theta)\le R$, for some constant $R\ge m$, with $(1+\theta)\left(|\bar E'(\theta)|+|\bar r'(\theta)|\right)$ bounded and $\theta\left(|\bar E''(\theta)|+|\bar r''(\theta)|+\bar E'^2(\theta)+\bar r'^2(\theta)\right)$ small, uniformly with respect to θ.

The existence result in [8] is stated as follows.

Theorem 4. *Let Hypotheses (H1), (H2) hold. Then there exists a unique solution (u,θ) to the problem (3.1)–(3.4) such that*

$$u_{tt}, u_{xx}, u_{xxt}, \theta_x \in L^\infty(0,T;\,L^2(0,1)), \tag{3.16}$$

$$u_{xtt}, \theta_t, \theta_{xx} \in L^2(]0,1[\times]0,T[), \tag{3.17}$$

$$\theta, u, u_x, u_t, u_{xt} \in C([0,1]\times[0,T]). \tag{3.18}$$

In addition, there exists a constant $c_0>0$ depending only on the given data such that for all $t\in[0,T]$ and $x\in[0,1]$ we have

$$\theta(x,t) \ge \delta e^{-c_0 t} > 0, \tag{3.19}$$

and (3.1)–(3.4) are satisfied almost everywhere.

We first check that the model is thermodynamically consistent according to (2.19), (2.20).

Corollary 5. *The solution from Theorem 4 satisfies the Clausius-Duhem inequality (2.20) with S defined by (2.24), (2.22) almost everywhere in $]0,1[\times]0,T[$.*

Proof of Corollary 5. For a.e. x and t we have

$$\begin{aligned} &\theta S_t + \theta\left(\frac{q}{\theta}\right)_x - g \qquad (3.20)\\ &= C_V\theta_t + \theta\left(\mathcal{S}[u_x,\theta]\right)_t + \beta\theta u_{xt} - \theta_{xx} - g + \frac{1}{\theta}\theta_x^2\\ &= -\left(\mathcal{V}[u_x,\theta]\right)_t + \theta\left(\mathcal{S}[u_x,\theta]\right)_t + \mathcal{P}[u_x,\theta]\,u_{xt} + \mu u_{xt}^2 + \frac{1}{\theta}\theta_x^2\\ &= \int_0^\infty \varphi(r,\theta)\,\mathfrak{s}_r[u_x]\left(u_x - \mathfrak{s}_r[u_x]\right)_t dr + \mu u_{xt}^2 + \frac{1}{\theta}\theta_x^2, \end{aligned}$$

and the assertion follows from (2.10). □

The solutions to (3.1)–(3.4) are unique and depend continuously on the data in the following way, see [8].

Theorem 6. *Let Hypotheses (H1) (ii), (H2) hold, let (u^0,u^1,θ^0,f,g), $(u'^0,u'^1,\theta'^0,f',g')$ be two sets of given functions satisfying Hypothesis (H1), and let (u,θ), (u',θ') be solutions of (3.1) – (3.4) corresponding to these data, respectively, which satisfy (3.16) – (3.19). Assume moreover that there exist a constant $\tilde{K}>0$ and functions $d_f,d_g\in L^2(]0,1[\times]0,T[)$ such that*

$$|f(\theta_1,x,t) - f'(\theta_2,x,t)| \le \tilde{K}|\theta_1-\theta_2| + d_f(x,t), \qquad (3.21)$$

$$|g(\theta_1,x,t) - g'(\theta_2,x,t)| \le \tilde{K}|\theta_1-\theta_2| + d_g(x,t), \qquad (3.22)$$

holds for all $\theta_1,\theta_2\in\mathbb{R}^+$ and a.e. $(x,t)\in]0,1[\times]0,T[$.

Then there exists a constant C depending only on the size of the data in their respective spaces such that for all $t\in[0,T]$ the differences $\bar{u}=u-u'$, $\bar{\theta}=\theta-\theta'$, satisfy

$$\begin{aligned} &\|\bar{u}_t(t)\|^2 + \int_0^t \left(\|\bar{\theta}\|^2 + \|\bar{u}_{xt}\|^2\right)(\tau)\,d\tau \qquad (3.23)\\ &\le C\left(\|\bar{u}_t(0)\|^2 + \|\bar{u}_x(0)\|^2 + \|\bar{\theta}(0)\|^2 + \int_0^t\int_0^1 (d_f^2+d_g^2)\,dx\,dt\right). \end{aligned}$$

The proofs of the above theorems depend substantially on the following properties of the hysteresis operators $\mathcal{P}$ and $\mathcal{V}$.

Proposition 7. *Let Hypothesis (H2) hold. Then the operators $\mathcal{P},\mathcal{V}$ are causal and have the following properties.*

(i) *For every* $\varepsilon, \theta \in W^{1,1}(0,T)$, $\theta > 0$, *we have*

$$|\mathcal{P}[\varepsilon,\theta](t)| \le V_0, \quad |\mathcal{V}[\varepsilon,\theta](t)| \le V_0, \tag{3.24}$$

$$\left|\frac{d}{dt}\mathcal{P}[\varepsilon,\theta](t)\right| \le L\left(|\dot\varepsilon(t)| + |\dot\theta(t)|\right), \quad \text{a.e. in }]0,T[. \tag{3.25}$$

(ii) *For every* $\varepsilon_2, \theta_1, \theta_2 \in W^{1,1}(0,T)$, $\theta_1 > 0$, $\theta_2 > 0$ *and for every* $t \in [0,T]$, *we have*

$$|\mathcal{P}[\varepsilon_1,\theta_1] - \mathcal{P}[\varepsilon_2,\theta_2]|\,(t) \le L\left(|\theta_1-\theta_2|(t) + 2\max_{0\le\tau\le t}|\varepsilon_1-\varepsilon_2|(\tau)\right), \tag{3.26}$$

$$|\mathcal{V}[\varepsilon_1,\theta_1] - \mathcal{V}[\varepsilon_2,\theta_2]|\,(t) \le \frac{C_V}{2}|\theta_1-\theta_2|(t) + 2V_0\max_{0\le\tau\le t}|\varepsilon_1-\varepsilon_2|(\tau). \tag{3.27}$$

4 Asymptotic behavior

The system (3.1)–(3.4) exhibits multiple sources of energy dissipation (plasticity, viscosity, heat conduction) and it is quite justified to expect that under vanishing external forcing, that is $f \equiv g \equiv 0$, the velocity and the temperature gradient should asymptotically vanish as $t \to \infty$. This will certainly not be true for the strain because of the existence of remanent plastic deformations, cf. Section III.2 of [6] for the temperature-independent case. It turns out however that, here again, the problem is more difficult than in the case without hysteresis, due to the lack of smoothness of hysteresis operators. Below we prove that in fact, the velocity tends to 0 in L^2 as $t \to \infty$, but no asymptotics is known for the velocity gradient and for the temperature. The exact result reads as follows.

Theorem 8. *Let the hypotheses of Theorem 4 be satisfied. Assume moreover that* $\gamma(\varepsilon) = \gamma_0\,\varepsilon$ *for some* $\gamma_0 \ge 0$ *and that* $f(\theta,x,t) = g(\theta,x,t) = 0$ *for all* $\theta > 0$ *and (a.e.)* $x \in]0,1[$, $t > 0$. *Then the solution* (u,θ) *of* (3.1)–(3.4) *satisfies*

$$\lim_{t\to\infty}\int_0^1 u_t^2(x,t)\,dx = 0. \tag{4.1}$$

Proof. In the sequel, we denote by C_1, C_2, ... constants depending only on the initial conditions.

Step 1. Multiply (3.1) by u_t, add the result to (3.2) and integrate with respect to x over $]0,1[$. This yields the global energy balance identity

$$\frac{d}{dt}\int_0^1 \Big(\frac{1}{2}u_t^2 + \frac{\gamma_0}{2}u_x^2 + C_V\,\theta + \mathcal{V}[u_x,\theta]\Big)(x,t)\,dx = 0. \tag{4.2}$$

Step 2. Let us multiply (3.2) by $-1/\theta$. We rewrite the result in the form

$$\left(-C_V\,\log\,\theta + \frac{1}{2}\int_0^\infty \varphi_\theta(r,\theta)\,\mathfrak{s}_r^2[u_x]\,dr\right)_t + \frac{1}{\theta}\big(\theta_{xx} + \mu\,u_{xt}^2\big)$$

$$+\frac{1}{\theta}\int_0^\infty \varphi(r,\theta)\,\mathfrak{s}_r[u_x]\big(u_x-\mathfrak{s}_r[u_x]\big)_t\,dr+\beta\,u_{xt}\;=0, \tag{4.3}$$

and integrating with respect to x and t we obtain from (2.10) that

$$\int_0^1\left(-C_V\,\log\,\theta+\frac{1}{2}\int_0^\infty \varphi_\theta(r,\theta)\,\mathfrak{s}_r^2[u_x]\,dr\right)(x,t)\,dx$$

$$+\int_0^t\int_0^1\left(\frac{\theta_x^2}{\theta^2}+\mu\frac{u_{xt}^2}{\theta}\right)dx\,dt\;\le\;C_1. \tag{4.4}$$

The left-hand side can be estimated using the relations

$$\begin{aligned}\int_0^\infty \varphi_\theta(r,\theta)\,\mathfrak{s}_r^2[u_x]\,dr\;&=\;\int_0^\infty\Big[\big(\varphi_\theta(r,\theta)-\varphi_\theta(r,1)\big)\\&\quad+\big(\varphi_\theta(r,1)-\varphi(r,1)\big)+\varphi(r,1)\Big]\,\mathfrak{s}_r^2[u_x]\,dr\\&\ge\;-\int_0^\infty\Big[\big|\varphi_\theta(r,\theta)-\varphi_\theta(r,1)\big|+\big|\varphi_\theta(r,1)-\varphi(r,1)\big|\Big]\,r^2\,dr.\end{aligned} \tag{4.5}$$

From Hypothesis (H2) it follows that

$$\int_0^\infty\big|\varphi_\theta(r,1)-\varphi(r,1)\big|\,r^2\,dr\;\le\;2V_0, \tag{4.6}$$

$$\int_0^\infty\big|\varphi_\theta(r,\theta)-\varphi_\theta(r,1)\big|\,r^2\,dr\;\le\;\left|\int_1^\theta\int_0^\infty|\varphi_{\theta\theta}(r,\theta')|\,dr\,d\theta'\right|\;\le\;C_V\,|\log\,\theta|. \tag{4.7}$$

Using the trivial inequality

$$|\log\,\theta|\;\le\;\max\{\theta,-\log\,\theta\}, \tag{4.8}$$

and the estimate

$$\int_0^1\theta(x,t)\,dx\;\le\;C_2, \tag{4.9}$$

which follows from (4.2), we conclude that

$$\int_0^1\big|\log\,\theta(x,t)\big|\,dx\;+\;\int_0^t\int_0^1\left(\frac{\theta_x^2}{\theta^2}+\frac{u_{xt}^2}{\theta}\right)dx\,dt\;\le\;C_3. \tag{4.10}$$

Step 3. For every x and t we have

$$|u_t(x,t)|\;\le\;\int_0^1|u_{xt}(\xi,t)|\,d\xi\;\le\;\int_0^1\frac{|u_{xt}|}{\sqrt{\theta}}\sqrt{\theta}\,d\xi\;\le\;\sqrt{C_2}\left(\int_0^1\frac{u_{xt}^2}{\theta}\,d\xi\right)^{1/2}, \tag{4.11}$$

hence

$$\int_0^t \max_x |u_t(x,\tau)|^2 \, d\tau \ \le \ C_4. \tag{4.12}$$

Step 4. Analogously, for every x, y and t we have

$$\begin{aligned} \sqrt{\theta(x,t)} \ &\le \ \sqrt{\theta(y,t)} + \frac{1}{2} \int_0^1 \frac{|\theta_x|}{\sqrt{\theta}}(\xi,t) \, d\xi \\ &\le \ \sqrt{\theta(y,t)} + \frac{1}{2} \left(C_2 \int_0^1 \frac{\theta_x^2}{\theta^2}(\xi,t) \, d\xi \right)^{1/2}, \end{aligned} \tag{4.13}$$

hence

$$\max_x \theta(x,t) \ \le \ C_5 \left(1 + \int_0^1 \frac{\theta_x^2}{\theta^2}(\xi,t) \, d\xi \right). \tag{4.14}$$

Step 5. Multiply (3.1) by u_t and integrate over x. We obtain from (3.24)

$$\frac{1}{2} \frac{d}{dt} \int_0^1 u_t^2(x,t) \, dx + \mu \int_0^1 u_{xt}^2(x,t) \, dx \ \le \ \int_0^1 \left(V_0 + |\beta| \, \theta + \gamma_0 \, |u_x| \right) |u_{xt}|(x,t) \, dx, \tag{4.15}$$

and Hölder's inequality together with (4.2), (4.14) leads to the estimate

$$\begin{aligned} \frac{d}{dt} \int_0^1 u_t^2(x,t) \, dx + \int_0^1 u_{xt}^2(x,t) \, dx \ &\le \ C_6 \left(1 + \int_0^1 \theta^2(x,t) \, dx \right) \\ &\le \ C_7 \left(1 + \max_x \theta(x,t) \right) \\ &\le \ C_8 \left(1 + \int_0^1 \frac{\theta_x^2}{\theta^2}(x,t) \, dx \right). \end{aligned} \tag{4.16}$$

Step 6. For $t > 0$ put

$$y(t) := \int_0^1 u_t^2(x,t) \, dx, \quad h(t) := C_8 \int_0^1 \frac{\theta_x^2}{\theta^2}(x,t) \, dx. \tag{4.17}$$

By (4.12), (4.10) we have

$$\int_0^t y(\tau) \, d\tau \ \le \ C_4, \quad \int_0^t h(\tau) \, d\tau \ \le \ C_8 \, C_3, \tag{4.18}$$

and (4.16) can be rewritten in the form

$$\dot{y}(t) \ + \ y(t) \ \le \ C_8 + h(t) \quad \text{a.e.} \tag{4.19}$$

To complete the proof of Theorem 8, it suffices to use Theorem 9 below. □

5 Appendix: A differential inequality

We prove here the following general convergence result for differential inequalities which extends Lemma 3.1 of [12].

Theorem 9. *Let a nondecreasing function $f : [0,\infty[\to]0,\infty[$, an absolutely continuous function $y : [0,\infty[\to [0,\infty[$ and a function $h \in L^1(0,\infty)$, $h \geq 0$ a.e. be given. Assume that*

$$\int_0^\infty y(t)\,dt = Y < \infty, \quad \int_0^\infty h(t)\,dt = H < \infty, \tag{5.1}$$

$$\dot{y}(t) \leq f(y(t)) + h(t) \quad \text{a.e.}, \tag{5.2}$$

where the dot denotes derivative with respect to t. Then $\lim_{t\to\infty} y(t) = 0$.

If moreover there exist constants $A, B \geq 0$ such that $f(y) \leq Ay^2 + B$ for every $y \geq 0$, then

$$y(t) \leq \begin{cases} e^{AY}(y(0) + H + B) & \text{for } \ t < 1, \\ e^{AY}(Y + H + B/2) & \text{for } \ t \geq 1. \end{cases} \tag{5.3}$$

The following example shows that we cannot expect any a priori pointwise bound for $y(t)$ if $f(y)$ grows faster than y^2.

Example 10. Let $\varepsilon \in]0,1[$ be given. For $n > 1$ put

$$y_n(t) = \begin{cases} |t-1|^{\varepsilon-1} & \text{for } \ t \in [0,2] \setminus]1-1/n, 1+1/n[, \\ n^{1-\varepsilon} & \text{for } \ t \in [1-1/n, 1+1/n], \\ e^{2-t} & \text{for } \ t > 2. \end{cases} \tag{5.4}$$

Then y_n are absolutely continuous, $\int_0^\infty y_n(t)\,dt \leq 1 + 2/\varepsilon$, $y_n(0) = 1$, $\dot{y}_n(t) \leq (1-\varepsilon) y_n^{2+\varepsilon/(1-\varepsilon)}(t)$ a.e., and the sequence $\{y_n(1)\}$ is unbounded.

Proof of Theorem 9. Assume that there exists $\alpha > 0$ and a sequence $t_n \uparrow \infty$ such that

$$y(t_n) \geq 2\alpha \quad \forall n \in \mathbb{N}. \tag{5.5}$$

We may assume (selecting a subsequence, if necessary) that the inequality

$$t_{n+1} - t_n > \frac{2Y}{\alpha} + \beta, \tag{5.6}$$

holds for every $n \in \mathbb{N}$, where

$$\beta := \frac{\alpha}{2f(2\alpha)}, \quad t_1 > \frac{Y}{\alpha}. \tag{5.7}$$

By (5.1), the sets

$$A_n := \{t \in]t_n - \frac{Y}{\alpha}, t_n[: y(t) < \alpha\} \tag{5.8}$$

are nonempty and we may put for all $n \in \mathbb{N}$

$$a_n := \sup A_n, \tag{5.9}$$

and similarly

$$b_n := \inf\{t \in]t_n, t_n + \frac{Y}{\alpha}[: y(t) < \alpha\}, \tag{5.10}$$

$$s_n := \min\{t \in [a_n, b_n] : y(t) \geq 2\alpha\}. \tag{5.11}$$

By construction we have for all $n \in \mathbb{N}$

$$a_n < s_n \leq t_n < b_n < a_{n+1}, \tag{5.12}$$

$$a_{n+1} - b_n > \beta, \tag{5.13}$$

$$y(a_n) = y(b_n) = \alpha, \quad y(s_n) = 2\alpha, \quad y(t) \geq \alpha \quad \forall t \in [a_n, b_n]. \tag{5.14}$$

We now define an auxiliary function z by the formula

$$z(t) := \begin{cases} y(t) - \alpha & \text{for } \ t \in \bigcup_{n=1}^{\infty} [a_n, b_n], \\ 0 & \text{otherwise.} \end{cases} \tag{5.15}$$

Then z is nonnegative, absolutely continuous, and for a.e. $t > 0$ we have

$$\dot{z}(t) \leq f(z(t) + \alpha) + h(t), \quad z(t) \leq y(t). \tag{5.16}$$

Moreover, for $t \in [s_n - \beta, s_n]$ we have

$$z(t) \leq \alpha, \tag{5.17}$$

and integrating (5.16) from t to s_n we obtain

$$\begin{aligned} \alpha - z(t) &\leq \int_t^{s_n} \big(f(z(\tau) + \alpha) + h(\tau)\big) d\tau \\ &\leq \beta f(2\alpha) + \int_{s_n - \beta}^{s_n} h(\tau)\, d\tau. \end{aligned} \tag{5.18}$$

For all $t \in [s_n - \beta, s_n]$ we therefore have

$$\frac{\alpha}{2} \leq z(t) + \int_{s_n - \beta}^{s_n} h(\tau)\, d\tau, \tag{5.19}$$

and integrating once more we conclude that

$$\frac{1}{2}\alpha\beta \le \int_{s_n-\beta}^{s_n} \big(z(\tau)+\beta h(\tau)\big)\,d\tau \quad \forall n\in\mathbb{N}, \tag{5.20}$$

which is a contradiction, since both z and h are integrable and the intervals $]s_n-\beta, s_n[$ are pairwise disjoint.

To prove (5.3), it suffices to rewrite (5.2) in the form

$$\frac{d}{dt}\Big(y(t)e^{-A\int_0^t y(\tau)\,d\tau}\Big) \le \big(B+h(t)\big)e^{-A\int_0^t y(\tau)\,d\tau}, \tag{5.21}$$

hence for every $0\le s<t$ we have

$$\begin{aligned} y(t) &\le y(s)e^{A\int_s^t y(\tau)\,d\tau} + \int_s^t \big(B+h(\tau)\big)e^{A\int_\tau^t y(\sigma)\,d\sigma}d\tau \\ &\le e^{AY}\big(y(s)+H+B(t-s)\big). \end{aligned} \tag{5.22}$$

For $t\le 1$ we simply put $s=0$, for $t\ge 1$ we integrate (5.22) with respect to s from $t-1$ to t. □

References

1. M. Brokate, J. Sprekels: *Hysteresis and Phase Transitions*. Springer-Verlag, Heidelberg, 1996.
2. C. M. Dafermos: Global smooth solutions to the initial-boundary value problem for the equations of one-dimensional thermoviscoelasticity. *SIAM J. Math. Anal.* **13** (1982), 397–408.
3. C. M. Dafermos, L. Hsiao: Global smooth thermomechanical processes in one-dimensional thermoviscoelasticity. *Nonlin. Anal. TMA* **6** (1982), 435–454.
4. A. Yu. Ishlinskii: Some applications of statistical methods to describing deformations of bodies. *Izv. AN SSSR, Techn. Ser.* **9** (1944), 583–590.
5. M. A. Krasnosel'skii, A. V. Pokrovskii: *Systems with Hysteresis*. Springer-Verlag, Heidelberg, 1989 (Russian edition: Nauka, Moscow, 1983).
6. P. Krejčí: *Hysteresis, convexity and dissipation in hyperbolic equations*, Gakuto Int. Series Math. Sci. & Appl., Vol. 8, Gakkōtosho, Tokyo, 1996.
7. P. Krejčí, J. Sprekels: On a system of nonlinear PDEs with temperature-dependent hysteresis in one-dimensional thermoplasticity. *J. Math. Anal. Appl.* **209** (1997), 25–46.
8. P. Krejčí, J. Sprekels: Temperature-dependent hysteresis in one-dimensional thermovisco-elastoplasticity. *Appl. Math.* **43** (1998) (in print).
9. J. Lemaitre, J.-L. Chaboche: *Mechanics of solid materials*. Cambridge Univ. Press, 1990 (French edition: Bordas, Paris, 1985).
10. I. Müller: *Thermodynamics*. Pitman, New York, 1985.
11. L. Prandtl: Ein Gedankenmodell zur kinetischen Theorie der festen Körper. *Z. Ang. Math. Mech.* **8** (1928), 85–106.
12. W. Shen, S. Zheng: On the coupled Cahn-Hilliard equations. *Comm. PDE* **18** (1993), 701–727.

13. J. Sprekels, S. Zheng, P. Zhu: Asymptotic behavior of the solutions to a Landau-Ginzburg system with viscosity for martensitic phase transitions in shape memory alloys *(to appear in Siam. J. Math. Anal.)*.
14. M. Šilhavý: *The mechanics and thermodynamics of continuous media*. Springer, Berlin – Heidelberg, 1996.

EQUADIFF 9 CD ROM, Brno 1997
Masaryk University

PROCEEDINGS
pp. 97–113

Some Global Bifurcation Problems for Variational Inequalities

Vy Khoi Le[1] and Klaus Schmitt[2]

[1] Department of Mathematics and Statistics
University of Missouri - Rolla
Rolla, Missouri, 65409, USA
Email: vy@umr.edu
[2] Department of Mathematics
University of Utah
Salt Lake City, Utah 84112, USA
Email: schmitt@math.utah.edu

Abstract. The paper presents several examples of bifurcation problems for variational inequalities and discusses an abstract framework for treating such problems. This abstract framework is applied to analyze some of the problems stated.

AMS Subject Classification. 35J85, 35R35, 49J40, 49R99, 73V25

Keywords. Variational inequalities, unilateral problems, topological degree, bifurcation problems

1 Introduction

This paper is based on a lecture presented by the second author at *Equadiff 9* held during the last week of August, 1997 in Brno, Czech Republic. The purpose of the lecture was to present several illustrations of global bifurcation phenomena in variational inequalities and to present some general framework for the analysis of such problems. Thus we present and discuss several examples and show how the global bifurcation results derived in [9] may be applied.

We first present examples of bifurcation problems which may be formulated as variational inequalities, then provide an abstract setting for these problems and state and sketch a proof of a global bifurcation theorem which will apply in these situations and finally provide a (partial) bifurcation analysis for the examples given.

When studying buckling phenomena of constrained elastic systems, one is led in a very natural way to bifurcation problems for variational inequalities. For example, the problem of the buckling of a slender column (beam) that is constrained by some obstacles leads to a problem for variational inequalities, simply because one searches for extremal points of an energy functional in a space of possible displacements determined by the obstacles, and hence these extremal points, which in the absence of constraints result in the Euler-Lagrange differential equations, now will be characterized as solutions of inequalities.

2 Some examples

In this section we present several examples of bifurcation problems which may be formulated as bifurcation problems for variational inequalities.

2.1 A unilateral problem

Consider the following ordinary differential equation

$$-u'' + u = \lambda(u + u^3),\ t \in (0, \pi), \tag{2.1}$$

subject to the unilateral constraints

$$\begin{cases} 0 \le u(0),\ 0 \le u(\pi) \\ u'(0) \le 0 \le u'(\pi) \\ u(0)u'(0) = 0 = u(\pi)u'(\pi). \end{cases} \tag{2.2}$$

Since nontrivial solutions of (2.1) may not have multiple zeros, we see that the above problem includes four different types of boundary value problems, namely problems subject to the following conditions:

1. *Dirichlet boundary conditions:*

$$u(0) = 0 = u(\pi), \tag{2.3}$$

where, however λ must be restricted so that the second of the unilateral conditions (2.2) hold, i.e.

$$u'(0) < 0 < u'(\pi). \tag{2.4}$$

Thus, for example, the problem may not have any solutions u, with $u(t) > 0$, $t \in (0, \pi)$, nor any solutions u with $u(t) > 0$ for t in a neighborhood of 0 and $u(t) < 0$ for t in a neighborhood of π. Thus, imitating the bifurcation analysis for nonlinear Sturm-Liouville problems, we would surmise that the values

$$\lambda = n^2 + 1,\ n = 1, 3, \cdots \tag{2.5}$$

are bifurcation points, whereas the values

$$\lambda = n^2 + 1,\ n = 2, 4, \cdots \tag{2.6}$$

are not. Furthermore, changing the sign of a solution will no longer yield a solution. Solutions must have an even number of zeros interior to $(0, \pi)$.

2. *Neumann boundary conditions:*

$$u'(0) = 0 = u'(\pi), \tag{2.7}$$

where, however λ must be restricted so that the first of the unilateral conditions (2.2) hold, i.e.

$$u(0), u(\pi) > 0. \tag{2.8}$$

Again, using the bifurcation analysis for nonlinear Sturm-Liouville problems, we find that the values

$$\lambda = n^2 + 1, \ n = 0, 2, 4, \cdots \tag{2.9}$$

are bifurcation points, whereas the values

$$\lambda = n^2 + 1, \ n = 1, 3, \cdots \tag{2.10}$$

are not. Again, changing the sign of a solution will no longer yield a solution and solutions must have an even number of zeros interior to $(0, \pi)$.

3. *Mixed Dirichlet and Neumann boundary conditions:*

$$u(0) = 0 = u'(\pi), \tag{2.11}$$

where, however λ must be restricted so that the first and the second of the unilateral conditions (2.2) hold, i.e.

$$u'(0) < 0, \ u(\pi) > 0. \tag{2.12}$$

As above, we compute that the values

$$\lambda = \left(\frac{2n-1}{2}\right)^2 + 1, \ n = 1, 3, 5, \cdots \tag{2.13}$$

are bifurcation points, whereas the values

$$\lambda = \left(\frac{2n-1}{2}\right)^2 + 1, \ n = 2, 4, 6, \cdots \tag{2.14}$$

are not. Changing the sign of a solution will no longer yield a solution and these solutions must have an odd number of simple zeros interior to $(0, \pi)$.

4. *Mixed Neumann and Dirichlet boundary conditions:*

$$u'(0) = 0 = u(\pi), \tag{2.15}$$

where, however λ must be restricted so that the first and the second of the unilateral conditions (2.2) hold, i.e.

$$u(0) > 0, \ u'(\pi) > 0. \tag{2.16}$$

In this case we obtain the set of bifurcation points as for the other set of mixed boundary conditions considered above.

Let us formulate the above problem as an equivalent bifurcation problem for a variational inequality. To this end we consider the Sobolev space $H^1(0,\pi)$ (of all $L^2(0,\pi)$ functions with a square integrable first distributional derivative) and let the (closed and convex) set K be defined by

$$K = \{u \in H^1(0,\pi) : u(0) \geq 0,\ u(\pi) \geq 0\}.$$

Then, if u solves (2.2), u will be in $H^2(0,\pi)$ and hence $u \in C^1[0,\pi]$. Therefore if u also satisfies the boundary constraints (2.3), we may multiply (2.1) by an arbitrary $v \in K$ and an integration by parts and the boundary constraints yield

$$\begin{cases} \displaystyle\int_0^\pi u'(v-u)' + u(v-u) - \lambda(u+u^3)(v-u) \geq 0,\ \forall v \in K, \\ u \in K, \end{cases} \tag{2.17}$$

which is a variational inequality. Conversely, if u solves the variational inequality (2.17), using the density of $C_0^\infty(0,\pi)$ in K, we easily conclude that u actually solves (2.1), (2.2).

If we denote by I_K, the indicator function of the set K, i.e.

$$I_K(u) = \begin{cases} 0, & u \in K \\ \infty, & u \notin K, \end{cases}$$

then we see that the variational inequality (2.17) is equivalent to the variational inequality

$$\begin{cases} \displaystyle\int_0^\pi u'(v-u)' + u(v-u) - \lambda(u+u^3)(v-u) + I_K(v) - I_K(u) \geq 0 \\ \forall v \in H^1(0,\pi) \\ u \in H^1(0,\pi). \end{cases} \tag{2.18}$$

We note here, that because of the convexity and closedness of K, the functional I_K is a lower semicontinuous convex functional on the (Hilbert) space $H^1(0,\pi)$.

2.2 A unilateral problem for a semilinear elliptic equation

A higher dimensional analogue of the problem discussed above in section 2.1 is the following unilateral problem. Let Ω be a bounded smooth domain in $\mathbb{R}^N$, $N \geq 2$, and consider the semilinear elliptic equation

$$-\Delta u + u = \lambda(u + g(u)),\ x \in \Omega, \tag{2.19}$$

where $g : \mathbb{R} \to \mathbb{R}$ is a smooth odd function with $g'(0) = 0$ and $|g(u)| \leq a + b|u|^s$, $1 \leq s < \frac{N+2}{N-2}$. Let the following unilateral constraints be imposed

$$\begin{cases} u(x) \geq 0,\ \frac{\partial u}{\partial \nu} \geq 0,\ x \in \partial\Omega \\ u(x)\frac{\partial u}{\partial \nu} = 0,\ x \in \partial\Omega, \end{cases} \tag{2.20}$$

where ν is the unit normal vector field to $\partial\Omega$.

In this case, if we consider the Sobolev space $H^1(\Omega)$ and let

$$K = \{u \in H^1(\Omega) : u(x) \geq 0,\ x \in \partial\Omega \text{ (in the sense of traces)}\},$$

then the above unilateral problem is equivalent to the variational inequality

$$\begin{cases} \displaystyle\int_\Omega \nabla u \nabla(v-u) + u(v-u) - \lambda(u+g(u))(v-u) + I_K(v) - I_K(u) \geq 0, \\ \forall v \in H^1(\Omega), \\ u \in H^1(\Omega). \end{cases} \tag{2.21}$$

It is again apparent that the special Dirichlet problem, i.e. equation (2.19) subject to the boundary condition

$$u = 0,\ x \in \partial\Omega,$$

and the Neumann problem, i.e. equation (2.19) subject to

$$\frac{\partial u}{\partial \nu} = 0,\ x \in \partial\Omega$$

will yield some of the bifurcation points for problem (2.21). However, one very quickly sees that much more is needed to detect other bifurcation points.

2.3 A simply supported, or clamped, slender beam subject to elastic obstacles

In this example, we consider a bifurcation problem for a beam resting between two foundations (one above and one below, with partial contact along its length) with nonlinear elastic laws. This problem can be modeled by the following variational inequality:

$$\begin{cases} \displaystyle\int_0^a u''(v-u)'' - \lambda \int_0^a \frac{u'}{\sqrt{1+u'^2}}(v-u)' \\ \displaystyle + \left[\int_{I_1} k_1 (v^-)^\gamma + \int_{I_2} k_2 (v^+)^\beta\right] \\ \displaystyle - \left[\int_{I_1} k_1 (u^-)^\gamma + \int_{I_2} k_2 (u^+)^\beta\right] \geq 0,\ \forall v \in E, \\ u \in E. \end{cases} \tag{2.22}$$

Here, $[0,a]$ $(a > 0)$ is the interval occupied by the beam, and $E = H_0^2(0,a)$, or $E = H^2(0,a) \cap H_0^1(0,a)$ depending on whether the beam is clamped or is simply supported at the ends 0 and a. $I_1, I_2 \subset (0,a), |I_1|, |I_2| > 0$ are closed,

disjoint sets representing the domain of possible contact between the beam and the foundations.

We refer to [12], [13], and [14], for the physical motivation in deriving such a model.

Because $u \mapsto u^+, u^-,\ u \in \mathbb{R}$ are nonnegative and convex, we see that the functional j, given by

$$j(u) = \int_{I_1} k_1(u^-)^\gamma + \int_{I_2} k_2(u^+)^\beta,$$

is well defined, with values in $[0, \infty]$. Moreover, j is convex and nonnegative, and $j(0) = 0$. Using Fatou's lemma, we find that j is lower semicontinuous on V.

2.4 Bifurcation problems for Navier-Stokes flows

We consider here bifurcation problems for some (nonlinear) variational inequalities associated with the Navier-Stokes equation, subject to different types of unilateral constraints (cf. [10]). Let Ω be a bounded domain in $\mathbb{R}^3$ with smooth boundary. We are concerned with variational inequalities of the form:

$$\begin{cases} \nu \displaystyle\int_\Omega Du : D(v-u) + b(u,u,v-u) + j(v) - j(u) \\ \geq \displaystyle\int_\Omega g(x,u,\lambda)\cdot(v-u),\ \forall v \in E \\ u \in E. \end{cases} \tag{2.23}$$

Here $E = \{v \in [H_0^1(\Omega)]^3 : \text{div} v = 0 \text{ a.e. in } \Omega\}$. E is a (Hilbert) subspace of $[H_0^1(\Omega)]^3$ with the restricted norm and scalar product. We also denote $Du = [\partial_i u_j]_{1\leq i,j\leq 3}$ and assume that $\nu > 0$ is the viscosity constant.

Let b be the trilinear form defined on $[H_0^1(\Omega)]^3$ by

$$\begin{aligned} b(u,v,w) &= \int_\Omega \sum_{i,j=1}^3 u_i(\partial_i v_j) w_j dx \\ &= \int_\Omega u^T (Du) w dx, \end{aligned}$$

for all $u, v, w \in [H_0^1(\Omega)]^3$.

We also assume that $j : V \to [0,\infty]$ is a convex, lower semicontinuous functional such that $j(0) = 0$, and $g : \Omega \times \mathbb{R}^3 \times \mathbb{R} \to \mathbb{R}^3,\ (x,u,\lambda) \mapsto g(x,u,\lambda)$ satisfies the Carathéodory condition (i.e. g_i satisfies this condition for each $i = 1,2,3$). We assume that g is differentiable with respect to u and $g, D_u g$ satisfies the usual growth condition:

$$\begin{cases} |g(x,u,\lambda)| \leq A(\lambda) + B(\lambda)|u|^{s-1} \\ |D_u g(x,u,\lambda)| \leq A(\lambda) + B(\lambda)|u|^{s-2}, \end{cases} \tag{2.24}$$

for a.e. $x \in \Omega$, all $u, \lambda \in \mathbb{R}$, with $A, B \in L^{\infty}_{loc}(\mathbb{R})$, $1 < s < 6(= 2^*)$.

Here u is the velocity of the fluid, b is the usual trilinear form in the Navier-Stokes equation, and g is the outer force acting on the fluid. g depends on u (in a nonlinear manner) and on λ, which usually represents the magnitude of the force. We assume that

$$g(x, 0, \lambda) = 0 \quad \text{for a.e. } x \in \Omega, \quad \text{all } \lambda \in \mathbb{R},$$

i.e., we have no external force at points with zero velocity. Here j is some kind of constraint imposed on the velocity. In many cases, j is of the form $j = I_K$, where K is a closed, convex subset of V, representing the set of admissible velocity fields of the fluid. For example, interesting choices of K are the following:

$$K = \{u \in E : u_1(x) \geq -c, \ u_2(x) \geq -d, \ c, d \geq 0\},$$

$$K = \{u \in E : |\nabla \times u| \leq c, \ c \geq 0\},$$

$$K = \{u \in E : |\int_S u \cdot n dS| \leq c, \ c \geq 0\}.$$

In the case $j = 0$, the variational inequality (2.23) becomes the equation:

$$\begin{cases} \nu \int_\Omega Du : Dv + b(u, u, v) = \int_\Omega g(x, u, \lambda) \cdot v, \ \forall v \in E \\ u \in E, \end{cases} \tag{2.25}$$

which is the usual variational form of the Navier-Stokes equation (cf. [11], [16], or [17]).

Other interesting choices for the functional j (the case of visco plastic Bingham fluids, cf. [11]) are:

$$j(u) = \int_\Omega \mu(x)|Du|^\gamma,$$

$$j(u) = \int_\Omega \mu(x)|\sum \epsilon_{ij}^2(u)|^\gamma,$$

where

$$\epsilon_{ij}(u) = \frac{1}{2}(\partial_i u_j + \partial_j u_i)$$

and μ is a nonnegative locally integrable function.

2.5 Bifurcation problems associated with the p-Laplace operator

In this example, we consider bifurcation problems for the following variational inequality:

$$\begin{cases} \int_\Omega |\nabla u|^{p-2} \nabla u \nabla (v - u) - \int_\Omega [\lambda |u|^{p-2} u + g(x, u, \lambda)](v - u) + j(v) - j(u) \\ \quad \geq 0, \ \forall v \in E, \\ u \in E. \end{cases} \tag{2.26}$$

Here $p > 1$, Ω is a bounded domain in $\mathbb{R}^N$ $(N \geq 1)$ with a smooth boundary,

$$E = \{u \in W^{1,p}(\Omega) : v = 0 \text{ on } \Gamma\},$$

where Γ is a (relatively) open subset of $\partial\Omega$ with positive measure. $W^{1,p}(\Omega)$ is the usual Sobolev space, equipped with the norm,

$$\|u\|_{W^{1,p}(\Omega)} = \left[\int_\Omega (|u|^p + |\nabla u|^p)\right]^{1/p}, \ u \in W^{1,p}(\Omega).$$

$(E, \|\cdot\|_{W^{1,p}(\Omega)})$ is a closed (Banach) subspace of $W^{1,p}(\Omega)$. By Poincaré's inequality, we know that

$$\|u\| = \left(\int_\Omega |\nabla u|^p\right)^{1/p}, \ u \in E,$$

defines a norm on E, equivalent to $\|\cdot\|_{W^{1,p}(\Omega)}$. In the sequel, we will always consider E with this norm. We also define the pairing between E and E^* by $\langle\cdot,\cdot\rangle$. We assume that

$$g : \Omega \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$$

is a Carathéodory function, such that

$$g(x, u, \lambda) = o(|u|^{p-1}), \tag{2.27}$$

as $u \to 0$, uniformly a.e. with respect to $x \in \Omega$ and uniformly with respect to λ on bounded intervals, and, moreover, g satisfies the growth condition

$$|g(x, u, \lambda)| \leq C(\lambda)[m(x) + M|u|^{p-1}], \tag{2.28}$$

for a.e. $x \in \Omega$, all $u, \lambda \in \mathbb{R}$, where $C(\lambda) \geq 0$ is bounded on bounded sets, $m \in L^{\frac{p}{p-1}}(\Omega)$, and $M > 0$ is a constant.

As a particular choice for the functional j we shall take

$$j(u) = \int_{\partial\Omega} |u| dS, \ u \in V. \tag{2.29}$$

Other choices of j will also be considered.

3 The abstract setting

In this section we shall provide an abstract framework for a bifurcation analysis for the types of problems introduced in the previous section, section 2. The setting will be variational inequalities in reflexive Banach spaces.

3.1 Notation and definitions

Throughout we shall denote by E a reflexive Banach space and by E^* its dual. The norm in E will be denoted by $\|\cdot\|$ and that in E^* by $\|\cdot\|_*$. The pairing between E^* and E shall be given by $\langle\cdot,\cdot\rangle$, i.e. if $f \in E^*$ and $u \in E$, then $f(u) = \langle f, u\rangle$.

We shall assume that:

–

$$j, J : E \to \mathbb{R}_+ \cup \{\infty\}$$

are convex and lower semicontinuous functionals with

$$j(0) = J(0) = 0.$$

–

$$A, \alpha : E \to E^*$$

are continuous and bounded operators with

$$A(0) = \alpha(0) = 0,$$

which are strictly monotone, coercive and belong to class (S), i.e:

- *A is strictly monotone:*

$$\langle A(u) - A(v), u - v\rangle > 0, \text{ whenever } u \neq v.$$

- *A is coercive:* There exist constants $c > 0$ and $p > 1$ such that

$$\langle A(u), u\rangle \geq c\|u\|^p, \ \forall u \in E.$$

- *A belongs to class (S)* : For all weakly convergent sequences $\{v_n\}$, $v_n \rightharpoonup v$, with

$$\lim\langle A(v_n), v_n - v\rangle = 0,$$

it must hold that

$$v_n \to v.$$

–

$$B, f : \mathbb{R} \times E \to E^*$$

are completely continuous operators with

$$B(\lambda, 0) = 0 = f(\lambda, 0), \ \forall\lambda \in \mathbb{R}.$$

3.2 Homogenizations

The following relationships between the operators introduced above (section 3.1) will be assumed:

– For all sequences $\{v_n\}$, $v_n \to v$, and all sequences of positive numbers σ_n, $\sigma_n \to 0+$,

$$\lim \frac{1}{\sigma_n^{p-1}} A(\sigma_n v_n) = \alpha(v).$$

– For all weakly convergent sequences $\{v_n\}$, $v_n \rightharpoonup v$, and all sequences of positive numbers σ_n, $\sigma_n \to 0+$, all sequences $\{\lambda_n\}$, $\lambda_n \to \lambda$,

$$\lim \frac{1}{\sigma_n^{p-1}} B(\lambda_n, \sigma_n v_n) = f(\lambda, v).$$

– For all weakly convergent sequences $\{v_n\}$, $v_n \rightharpoonup v$, and all sequences of positive numbers σ_n, $\sigma_n \to 0+$,

$$\liminf \frac{1}{\sigma_n^{p}} j(\sigma_n v_n) \geq J(v),$$

further, for all $v \in E$, and all sequences of positive numbers σ_n, $\sigma_n \to 0+$, there exists a sequence $\{v_n\}$, $v_n \to v$, such that

$$\lim \frac{1}{\sigma_n^{p}} j(\sigma_n v_n) = J(v).$$

3.3 Equivalent operator equations

Consider, for $g \in E^*$, the variational inequality

$$\begin{cases} \langle A(u) - g, v - u\rangle + j(v) - j(u) \geq 0, \ \forall v \in E \\ u \in E. \end{cases} \tag{3.1}$$

It follows from classical results (see e.g. [7], [11]), that this problem is uniquely solvable, hence defines an operator

$$T_{A,j} : E^* \to E \tag{3.2}$$

by

$$T_{A,j}(g) = u,$$

where u is the unique solution of (3.1). This operator is also continuous (cf. [9]). Therefore, if we consider the variational inequality

$$\begin{cases} \langle A(u) - B(\lambda, u), v - u\rangle + j(v) - j(u) \geq 0, \ \forall v \in E, \\ u \in E, \end{cases} \tag{3.3}$$

then u solves (3.3) if and only if u solves

$$T_{A,j}B(\lambda, u) = u. \tag{3.4}$$

And similarly if we consider the variational inequality

$$\begin{cases} \langle \alpha(u) - f(\lambda, u), v - u\rangle + J(v) - J(u) \geq 0, \ \forall v \in E, \\ u \in E, \end{cases} \tag{3.5}$$

then u solves (3.5) if and only if u solves

$$T_{\alpha,J}f(\lambda, u) = u. \tag{3.6}$$

It follows from the relationships between A and α, B and f and j and J, that if u solves (3.5) then so does σu for any $\sigma(> 0) \in \mathbb{R}$.

3.4 Global bifurcation

Let us assume that $(\lambda_0, 0) \in \mathbb{R} \times E$ is a bifurcation point for (3.3), then it follows that (3.5) and hence also (3.6) will have a nontrivial solution for $\lambda = \lambda_0$. Therefore, if $a \in \mathbb{R}$ is such that (3.5) has only the trivial solution for $\lambda = a$, it will follow that for $r > 0$, sufficiently small, the Leray-Schauder degree

$$d(\mathrm{id} - T_{\alpha,J}f(a, \cdot), B_r(0), 0)$$

is defined (here $B_r(0)$ is the open ball of radius r in E centered at 0) and we obtain

$$d(\mathrm{id} - T_{\alpha,J}f(a, \cdot), B_r(0), 0) = d(\mathrm{id} - T_{A,j}B(a, \cdot), B_r(0), 0)$$

(see e.g. [9]). We hence may employ the homotopy invariance principle of the Leray-Schauder degree, to conclude that if $a, b \in \mathbb{R}$, $a < b$ are such that (3.5) has only the trivial solution for $\lambda = a, b$ and if

$$d(\mathrm{id} - T_{\alpha,J}f(a, \cdot), B_r(0), 0) \neq d(\mathrm{id} - T_{\alpha,J}f(b, \cdot), B_r(0), 0) \tag{3.7}$$

then $[a, b] \times \{0\}$ will contain a bifurcation point for (3.4) and hence for (3.3) (cf. [8]). In fact, we may employ the global bifurcation result of Rabinowitz [15] to conclude that global bifurcation takes place in the sense of that theorem.

Thus in bifurcation problems of the type (3.3), in order to be able to apply the above considerations we need to compute the operators α and f, the functional J. Further one needs to find values $a, b \in \mathbb{R}$, $a < b$ such that (3.7) holds for λ values a and b (by no means an easy task, in general). This we shall do for some of the examples considered in section 2 and refer the interested reader to many additional examples in [9].

4 Examples revisited

In this section we shall employ the abstract setting discussed in section 3 to discuss the existence of some bifurcation points for examples related to those introduced in section 2. We shall not dwell on the first example, since this problem is equivalent to the existence of bifurcation branches (in K) of four different nonlinear Sturm-Liouville problems, those problems being completely understood.

Before turning to the discussion of some of the other examples, we present some other abstract features common to some of them.

4.1 Semilinear problems

Let us assume

$$a : E \times E \to \mathbb{R}$$

is a continuous, coercive and bilinear form and let

$$A : E \to E^*$$

be defined by

$$\langle A(u), v\rangle = a(u, v).$$

Furthermore assume that

$$B(\lambda, u) = \lambda Bu + R(u),\ \ R(u) = o(\|u\|),\ \text{ as } u \to 0,$$

with B compact linear and that

$$j = I_K,$$

where K is a closed convex subset of E with $0 \in K$.

In this case one easily computes that $p = 2$, $\alpha = A$, $f(\lambda u) = \lambda Bu$ and $J = I_{K_0}$, where K_0 is the support cone of K, i.e

$$K_0 = \overline{\cup_{t>0} tK}.$$

If it is the case that K_0 is a subspace of E, then the variational inequality (3.5) becomes

$$\begin{cases} \langle \alpha(u) - f(\lambda, u), v - u\rangle + I_{K_0}(v) - I_{K_0}(u) \geq 0,\ \forall v \in E, \\ u \in E, \end{cases} \tag{4.1}$$

which is equivalent to

$$\begin{cases} \langle \alpha(u) - f(\lambda, u), v - u\rangle \geq 0,\ \forall v \in K_0 \\ u \in K_0, \end{cases} \tag{4.2}$$

and, since K_0 is a subspace, the latter is equivalent to

$$\begin{cases} \langle \alpha(u) - f(\lambda, u), v \rangle = 0, \ \forall v \in K_0 \\ u \in K_0. \end{cases} \tag{4.3}$$

From this we see (recall the comment at the end of section 3.3) that the solution operator $T_{\alpha,J}$ is a bounded linear operator and equation (3.6) becomes

$$u = \lambda T_{\alpha,J} B u. \tag{4.4}$$

Hence the possible bifurcation points for (3.3) are to be sought among the countable set $\{(\lambda_i, 0)\}$, where λ_i is a characteristic value of the compact linear operator $T_{\alpha,J}B$. And each characteristic value of odd multiplicity will yield a bifurcation point. We note here that what has just been said is true as long as J is the indicator function of a subspace, irregardless whether $j = I_K$ for some closed convex set K.

4.2 A semilinear elliptic problem

Let $\Omega \subset \mathbb{R}^N$ be a bounded domain with smooth boundary $\partial\Omega$, and let $\Gamma \subset \partial\Omega$ be a relatively open subset of positive measure.

Let

$$E = \{u \in H^1(\Omega) : u = 0, \text{ a.e. on } \Gamma\}.$$

Let

$$a : E \times E \to \mathbb{R}$$

be given by

$$a(u, v) = \int_\Omega \nabla u \cdot \nabla v,$$

then (because of Poincaré's inequality) a is a continuous, coercive and bilinear form. Let $g : \mathbb{R} \to \mathbb{R}$ be a continuous function with $g(u) = o(|u|)$ as $u \to 0$, and define $B(\lambda, u)$ by

$$\langle B(\lambda, u), v \rangle = \int_\Omega \lambda uv + g(u)v,$$

then

$$\langle f(\lambda, u), v \rangle = \int_\Omega \lambda uv.$$

Let us define the functional j by

$$j(u) = \int_{\partial\Omega} \mu |u|^\gamma,$$

where μ, γ are positive constants with $1 \leq \gamma < 2$.

Embedding theorems (see [1], [6]) tell us that the mapping

$$H^1(\Omega) \hookrightarrow L^q(\partial\Omega)$$
$$u \mapsto u|_{\partial\Omega}$$

are compact for

$$1 \leq q < \bar{p} = \begin{cases} \frac{2(N-1)}{N-2}, & N > 2 \\ \infty, & N = 1, 2. \end{cases}$$

It hence will follow that j is convex and lower semicontinuous and (since $p = 2$ and $1 \leq \gamma < 2$) that

$$J(u) = I_{H_0^1(\Omega)}.$$

It hence follows from the results above, i.e. the results in section 4.1, that (3.5) is equivalent to the problem

$$\int_\Omega \nabla u \cdot \nabla v - \lambda \int_\Omega uv = 0, \ \forall v \in H_0^1(\Omega), \ u \in H_0^1(\Omega), \tag{4.5}$$

which is equivalent to the eigenvalue problem

$$\Delta u + \lambda u = 0, \ u \in H_0^1(\Omega). \tag{4.6}$$

We hence conclude that all eigenvalues of (4.6) which are of odd multiplicity yield bifurcation points.

4.3 An inequality involving the p-Laplacian

A situation, similar to the above, arises, if we consider the example presented in section 2.5. There we let

$$E = \{u \in W^{1,p}(\Omega) : u = 0, \text{ a.e. on } \Gamma\}$$

and let

$$A : E \to E^*$$

be given by

$$\langle A(u), v\rangle = \int_\Omega |\nabla u|^{p-2}\nabla u \cdot \nabla v.$$

Let $g : \mathbb{R} \to \mathbb{R}$ be a continuous function with $g(u) = o(|u|^{p-1})$ as $u \to 0$, and define $B(\lambda, u)$ by

$$\langle B(\lambda, u), v\rangle = \int_\Omega \lambda|u|^{p-2}uv + g(u)v,$$

then

$$\langle f(\lambda, u), v\rangle = \int_\Omega \lambda|u|^{p-2}uv.$$

Let us define the functional j by

$$j(u) = \int_{\partial\Omega} \mu|u|,$$

where μ is a positive constant.

Again, using embedding theorems (see [1]) we see that the mapping

$$W^{1,p}(\Omega) \hookrightarrow L^1(\partial\Omega)$$
$$u \mapsto u|_{\partial\Omega}$$

is compact.

It hence will follow that j is convex and lower semicontinuous and that

$$J(u) = I_{W_0^{1,p}(\Omega)}.$$

It hence follows from the results above, i.e. the results in section 4.1, that (3.5) is equivalent to the problem

$$\int_\Omega |\nabla u|^{p-2}\nabla u \cdot \nabla v - \lambda \int_\Omega |u|^{p-2}uv = 0, \ \forall v \in W_0^{1,p}\Omega, \tag{4.7}$$

which is equivalent to the eigenvalue problem

$$\operatorname{div}(|\nabla u|^{p-2}\nabla u) + \lambda|u|^{p-2}u = 0, \ u \in W_0^{1,p}(\Omega). \tag{4.8}$$

This eigenvalue problem has received much attention during recent years and several results about eigenvalues and the the computation of the Leray-Schauder degree of the associated completely continuous perturbation of the identity in a neighborhood of such eigenvalues have become available (see e.g. [2], [3], [4], [5]).

4.4 Stationary Navier-Stokes flows

In this section we consider the example discussed in section 2.4 and refer to this section for the statement of the problem and the notation.

Again the operator A is given by a continuous, coercive and bilinear form, hence $A = \alpha$. Also it easily follows that

$$\langle f(\lambda, u), v\rangle = \lambda \int_\Omega D_u g(x, 0)u \cdot v.$$

To hence obtain the homogeneous variational inequality (3.5) we must compute the functional J. To this end, we observe that if $j = I_K$, where K is any of the choices given in section 2.4, then $J = I_E$, since the support cone of K in any of the cases is the whole space.

We hence obtain that, in these cases, (3.5) is given by

$$\begin{cases} \nu \int_\Omega Du : Dv + \lambda \int_\Omega D_u g(x,0)u \cdot v = 0, \ \forall v \in E, \\ u \in E, \end{cases} \tag{4.9}$$

which is the eigenvalue problem for the Stokes equation. Its eigenvalues of odd multiplicity hence yield global bifurcation points for (2.23).

Let us now consider the case that j is given by

$$j(u) = \int_\Omega \mu(x)|Du|^\gamma,$$

where $\mu \in L^\infty(\Omega)$ and $\gamma \geq 1$. We observe that the effective domain of j is given by

$$D(j) = \{u : j(u) < \infty\} = \begin{cases} E, \ 1 \leq \gamma \leq 2 \\ \{u \in E : \mu|Du|^\gamma \in L^1(\Omega)\}, \ \gamma > 2 \end{cases}$$

Using these facts one may now compute

$$J = \begin{cases} I_W, 1 \leq \gamma < 2 \\ j, \gamma = 2 \\ I_E, \gamma > 2, \end{cases}$$

where

$$W = \{u \in E : Du = 0, \text{ a.e. on } \Omega \setminus \Omega_0\},$$

and

$$\Omega_0 = \{x \in \Omega : \mu(x) = 0\}.$$

References

1. R. Adams: *Sobolev Spaces*, Academic Press, New York, 1975.
2. A. Anane: *Simplicité et isolation de la première valeur propre du p-Laplacien*, C. R. Acad. Sci Paris, 305 (1987), 725–728.
3. F. de Thélin: *Sur l'espace propre associé à la première valeur propre du pseudo-Laplacien*, C. R. Acad. Sci. Paris, 303 (1986), pp. 355–358.
4. M. A. del Pino and R. F. Manásevich: *Global bifurcation from the eigenvalues of the p-Laplacian*, J. Diff. Equa., 92 (1991), pp. 226–251.
5. M. García-Huidobro, R. Manásevich and K. Schmitt: *On principal eigenvalues of p-Laplacian like operators,* J. Diff. Equations, 130 (1996), 235–246.
6. D. Gilbarg and N. Trudinger: *Elliptic Partial Differential Equations of Second Order*, Springer, Berlin, 1983.
7. D. Kinderlehrer and G. Stampacchia: *An Introduction to Variational Inequalities*, Academic Press, New York, 1980.

8. M. A. Krasnosels'kii: *Topological Methods in the Theory of Nonlinear Integral Equations*, Pergamon Press, Oxford, 1963.
9. V. Le and K. Schmitt: *Global Global Bifurcation in Variational Inequalities: Applications to Obstacle and Unilateral Problems*, vol. 123 Applied Math. Sciences, Springer, New York, 1997.
10. V. Le and K. Schmitt: *On variational inequalities associated with the Navier-Stokes equation: Some bifurcation problems,* Elect. J. Differential Equations, to appear 1998.
11. J. L. Lions: *Quelques Méthodes de Résolution des Problèmes aux Limites non Linéaires*, Dunod, Paris, 1969.
12. J. Oden and J. Martins: *Models and computational methods for dynamics friction phenomena,* Comp. Methods Appl. Mech. Eng., 52 (1985), pp. 527–634.
13. P. Rabier and J. Oden: *Solution to Signorini-like contact problems through interface models, I, preliminaries and formulation of a variational inequality,* Nonl. Anal., TMA, 11 (1987), pp. 1325–1350.
14. P. Rabier and J. Oden: *Solution to Signorini-like contact problems through interface models, II, existence and uniqueness theorems,* Nonl. Anal., TMA, 12 (1988), pp. 1–17.
15. P. Rabinowitz: *Some aspects of nonlinear eigenvalue problems,* Rocky Mtn. J. Math., 3 (1973), 162–202.
16. R. Temam: *Navier-Stokes Equations,* North-Holland, New York, 1977.
17. E. Zeidler: *Nonlinear Functional Analysis and its Applications, Vol.4: Applications to Mathematical Physics*, Springer, Berlin, 1988.

EQUADIFF 9 CD ROM, Brno 1997
Masaryk University

PROCEEDINGS
pp. 115–145

Seventy-Five Years of Global Analysis around the Forced Pendulum Equation

Jean Mawhin

Département de mathématique, Université Catholique de Louvain
B-1348 Louvain-la-Neuve, Belgium
Email: mawhin@amm.ucl.ac.be

Es dürfte nützlich sein, zuweilen die allgemeinen Forschritte,
welche die Analysis in den letzen Jahren gemacht hat,
an einem bestimmten Beispiel zu prüfen.(...)
Wir wollen die Differentialgleichung des Herrn Duffing
in der ursprünglichen Form hier etwa genauer untersuchen.
G. Hamel, 1922

Description of the set P of f for which the equation $u'' + \sin u = f(t)$
has a T-periodic solution seems to remain a terra incognita.
S. Fučik, 1979

Abstract. We survey the recent progress made in the study of harmonic, subharmonic and other solutions of the forced pendulum equation

$$u'' + cu' + a \sin u = h(t)$$

when the forcing term h is periodic, almost periodic or bounded. The results depend upon various methods of nonlinear functional analysis, critical point theory and dynamical systems.

AMS Subject Classification. 34C25, 34C11, 34C27, 70K40

Keywords. Forced pendulum equation, periodic solutions, almost periodic solutions, Lagrange stability

1 Introduction

Seventy-five year ago, in a paper published in 1922 in the special issue of the *Mathematische Annalen* dedicated to Hilbert's sixtieth birthday anniversary [86], Hamel, one of his former students, has provided the first general existence results for the periodic solutions of the periodically forced pendulum equation

$$y'' + a \sin y = b \sin t. \tag{1.1}$$

This equation had been the central topics of a monograph published four years earlier by Duffing [55], who had restricted his study to the approximate determination of the periodic solutions for the following approximation of equation (1.1)

$$y'' + ay - cy^3 = b \sin t,$$

which still bears his name today.

Hamel's paper starts by an existence result for a 2π-periodic solution of equation (1.1) by using the *direct method of the calculus of variations* made rigorous by Hilbert at the beginning of the century. He shows indeed that the *action integral*

$$A(y) := \int_0^{2\pi} \left(\frac{y'^2(t)}{2} + a \cos y(t) + by(t) \sin t \right) dt$$

has a minimum on the space of 2π-periodic functions of class C^1. His argument extends easily to the more general case where $b \sin t$ is replaced by any 2π-periodic function with mean value zero, a fact rediscovered independently, in the easier framework of Sobolev spaces, some sixty years later [179,47], and rapidly followed by the proof of the existence of a *second* 2π-periodic solution [124] through the use of more sophisticated tools of *critical point theory.* In the second section of [86], Hamel uses the *Ritz method* to find a first approximation of the amplitude of the periodic solution found in the previous section. In Section 4, Hamel observes that the symmetries of the equation imply that any solution of equation (1.1) such that

$$y(0) = y(\pi) = 0 \tag{1.2}$$

can be extended as an *odd* 2π-periodic solution. He then reduces the problem (1.1)–(1.2) to the integral equation

$$y(t) = -a \int_0^{2\pi} K(t, \tau) \sin y(\tau)\, d\tau - b \sin t := F(y)(t), \tag{1.3}$$

where $K(t, \tau)$ is the Green function of its linear part, and shows that the corresponding *method of successive approximations*

$$y_{n+1} = F(y_n), \quad y_0(t) = -b \sin t,$$

converges. His argument is equivalent to showing the existence of a sufficiently large integer m, for which the m^{th} iterate F^m of F is a contraction in the space of $C[0, \pi]$. Observe that this is published the very same year where Banach publishes his version of the *contraction mapping theorem*! Notice also that this part of Hamel's paper will inspire Hammerstein's famous researches on *nonlinear integral equations.* Hamel uses not only the equivalent integral equation for existence and uniqueness conclusions, but also to obtain approximations to the solutions. For this, Hamel relies upon Schmidt's version of the *Lyapunov-Schmidt's method.*

This short description clearly shows that Hamel's paper anticipates or uses several of the fundamental methods of nonlinear analysis, and that the opening sentence of his paper, recalled in exergue of this work, is fully justified.

The survey papers [110,111] describe the contributions to the forced pendulum equation in the fifty-five years following Hamel's work. An important role in renewing the interest to the forced pendulum equation was played in the late seventies by Fučik, who wrote, in the Introduction of Chapter 26 of his monograph [72] : *Finally we shall present here one attempt to obtain the existence of a T-periodic solution of the mathematical pendulum equation*

$$-u''(x) + \sin u(x) = f(x). \tag{1.4}$$

The result is not final since the necessary and sufficient condition obtained for T-periodic solvability of (1.4) *is not useful.* After describing very partial results in this direction and mentioning extensions personally communicated by Dancer, Fučik concluded with the sentence mentioned in exergue of this paper.

Motivated by Fučik's remarks, but unaware of the existence of Hamel's paper, Castro [38], Dancer [47] and Willem [179], reintroduced in the early eighties the use of variational methods in the study of the forced pendulum. The time was ripe for the obtention, more than sixty years after the first one, of a *second periodic solution,* using a version of the *mountain pass lemma* [124]. The survey papers [110,111,114,117,118,121,183], as well as to the monographs [112,126,40,79,98,151] provide a description of the state of the art till the early nineties for the global results on the existence and multiplicity of periodic solutions.

At the same time, the forced pendulum equation also became a paradigm for the *theory of chaos,* and appeared in the description of *Josephson type junctions.* According to Baker and Gollub [13]: *Now 400 years after Galileo's initial work, the pendulum has again become an object of research as a chaotic system.* We shall not develope this viewpoint here and refer to a nice survey of Chenciner [42] and to the papers or monographs [13,22,25,26,29,49,54,80,81,82,87,88,89,94,95] [96,97,100,101,104,127,152,160,168,150,170] and their references.

Despite its fundamental role in the development of the qualitative theory of nonlinear differential equations and its applications to engineering, we shall not discuss here the special case of the *pendulum equation with a constant torque*

$$y'' + cy' + a\sin y = b,$$

initiated by Tricomi [175,176] and widely developed since (see e.g. [10,15,85,105] [154,153,163,164] and their references).

Moreover, to keep the size of the paper reasonable and make more easy the comparaison between results obtained through different methods, we shall only state the theorems for the special case of the standard forced pendulum equation

$$y'' + cy' + a\sin y = h(t).$$

Most of the assertions remain valid if $a\sin y$ is replaced by an arbitrary continuous function $g(y)$ which is S-periodic for some $S > 0$ and of mean value zero.

There are even recent results which depend upon the fact that $a \sin y$ is replaced by a S-periodic function whose Fourier series contains higher harmonics [93]. Also, some conclusions survive when the friction term cy' is replaced by a more general one of Liénard type $f(y)y'$ or of Rayleigh type $f(y')$ (see e.g. [124,83]).

For the same reason, we shall not describe the possible generalizations to *systems of equations of the pendulum-type,* and in particular to the equations of the *forced multiple pendulum,* and to *higher order pendulum-type equations.* The reader can consult the original papers [125,36,53,66,149,64,119,40,62,171,173] [174,56,57,128,63,58,67,21]. Some of those results are related to the famous solution by Conley and Zehnder [44] of a *conjecture of Arnold* in symplectic geometry. See also, for example, [45,189,190,61,106]. We shall not describe the results dealing with *symmetric forcing terms* $h(t)$, which have been recently considered in [161,162,155,16,136]. Also we shall leave aside the existence of forced oscillations for the *spherical pendulum* (which depend upon methods of a quite different nature, and have been the object of a sequence of papers by Furi, Pera and Spadini [73,74,75,76,77]), for some pendulum-type equations describing the *libration of satellites* (see [17,99,113,116,147,84]), and for delay-differential equations of the pendulum type like the *sunflower equation* [37].

Let us mention also that the corresponding problem for the case of *Dirichlet boundary conditions,* namely

$$y'' + y + a \sin y = h(t), \quad y(0) = 0 = y(\pi),$$

and its analog for partial differential equations, has been the object, since the pioneering paper of Ward [178], of a number of studies based upon various methods. See [169,107,156,157,158,159,46,11,31,32,33,34]. This problem has both deep analogies and strong differences with the periodic boundary value problem for the forced pendulum.

Finally, let us warn the reader that, despite of its substantial size, the given bibliography is undoubtly far to be complete, but its size is sufficient to show how stimulating has been the study of the forced pendulum equation in the recent development of nonlinear and global analysis, and of the theory of dynamical systems.

2 Periodic forcing

2.1 The problems

We consider the (possibly dissipative) *periodically forced pendulum equation*

$$y'' + cy' + a \sin y = h(t), \tag{2.1}$$

where, without loss of generality, $c \geq 0$, $a > 0$, and h is T-periodic, for some period $T > 0$, and corresponding frequency $\omega := \frac{2\pi}{T}$. For the simplicity of exposition, we shall assume that h is continuous. Most results hold under weaker regularity conditions.

A T-*periodic* solution of equation (2.1) is a solution $y : \mathbb{R} \to \mathbb{R}$ such that $y(t+T) = y(t)$ for all $t \in \mathbb{R}$. By integrating equation (2.1) over $[0, T]$, we immediately see that a *necessary condition* for the existence of a T-periodic solution to equation (2.1) is that

$$\left| \frac{1}{T} \int_0^T h(t)\, dt \right| \leq a.$$

The main questions which can be raised about the T-periodic problem for equation (2.1) are the following ones:

1. Determine the nature and the properties of the set

$$\mathcal{R} = \mathcal{R}(c, a, T)$$

of T-periodic forcings h such that equation (2.1) has at least one T-periodic solution, i.e. the *range* of the nonlinear operator

$$\frac{d^2}{dt^2} + c\frac{d}{dt} + a\sin(\cdot)$$

over the space of T-periodic functions of class C^2.
2. For $h \in \mathcal{R}$, discuss the *multiplicity* of the T-periodic solutions.
3. For $h \in \mathcal{R}$, discuss the *stability* of the T-periodic solutions.
4. Discuss the existence of *other solutions* and *properties of the set of all solutions.*

Concerning the multiplicity, it is clear that if y is a T-periodic solution of equation (2.1), then the same is true for $y + 2k\pi$, $k \in \mathbb{Z}$. Consequently, we shall say that y_1 and y_2 are *distinct* T-*periodic solutions of* (2.1) if they do not differ by a multiple of 2π.

In the sequel of the paper, we shall use the following notations.

$$L_T^p = \{h \in L_{loc}^p(\mathbb{R}) : h(t+T) = h(t) \text{ for a.e. } t \in \mathbb{R}\}$$

$$C_T = \{h \in C(\mathbb{R}) : h(t+T) = h(t) \text{ for all } t \in \mathbb{R}\}$$

$$H_T^1 = \{h \in AC_{loc}(\mathbb{R}) : h' \in L^2\}$$

$$\|h\|_p = \left(\frac{1}{T} \int_0^T |h(t)|^p\, dt \right)^{1/p}, \quad \|h\|_\infty = \max_{t \in [0,T]} |h(t)|,$$

$$\|h\|_{H^1} = \left(\|h\|_2^2 + \|h'\|_2^2 \right)^{1/2}$$

$$\overline{h} = \frac{1}{T} \int_0^T h(t)\, dt, \quad \widetilde{h}(t) = h(t) - \overline{h} \qquad \left(\int_0^T \widetilde{h}(t)\, dt = 0 \right)$$

$$\widetilde{L_T^p} = \{h \in L_T^p : \overline{h} = 0\}, \quad \widetilde{C_T} = \{h \in C_T : \overline{h} = 0\}$$

Consequently,

$$L_T^p = \mathbb{R} \oplus \widetilde{L_T^p}, \quad C_T = \mathbb{R} \oplus \widetilde{C_T},$$

with the corresponding decomposition $y = \overline{y} + \widetilde{y}$.

We shall also use an interesting *equivalent formulation* of the problem of T-periodic solutions for equation (2.1).

Lemma 1. *If $\widetilde{H}(t) = \widetilde{H}_{c,T}(t)$ denotes the unique T-periodic solution in $\widetilde{C_T}$ of*

$$y'' + cy' = \widetilde{h},$$

then $y(t)$ is a T-periodic solution of equation (2.1) if and only $x(t) = y(t) - \widetilde{H}(t)$ is a T-periodic solution of equation

$$x'' + cx' + a\sin(x + \widetilde{H}(t)) = \overline{h}. \tag{2.2}$$

2.2 The methods

Various methods have been used in the study of the T-periodic solutions of equation (2.1) or (2.2). For the reader's convenience, we shall give a brief survey of the ones directly involved in the results described in this survey.

2.2.1 Poincaré's method

Let $y(t;u)$ be the solution of equation (2.1) such that

$$y(0,u) = u_1,\ y'(0;u) = u_2,$$

and let

$$P : \mathbb{R}^2 \to \mathbb{R}^2, u \mapsto [y(T;u), y'(T;u)].$$

Then $y(t;u)$ is a T-periodic solution of equation (2.1) if and only if u is a fixed point of P. P is called the *Poincaré's operator*.

If $c = 0$, P is area-preserving, and one can then use various *twist theorems*. Take polar coordinates (r,θ) in the plane, and denote by A the annulus $[a,b]\times S^1$. A first useful result is *Poincaré-Birkhoff's twist theorem* [148,23].

Lemma 2. *Every area-preserving homeomorphism $\phi : A \to A$ with lift*

$$(r,\theta) \mapsto (f(r,\theta), \theta + g(r,\theta)), \tag{2.3}$$

rotating the two boundaries in opposite directions, i.e. such that

$$g(a,\theta)g(b,\theta) < 0, \theta \in \mathbb{R},$$

possesses at least two fixed points in the interior .

A second one is *Moser's twist theorem* [129].

Lemma 3. *Let $l \geq 5$, $\alpha \in C^5(\mathbb{R})$ be such that $|\alpha'(r)| \geq \nu > 0$ for all $r \in [a,b]$, and let $\varepsilon > 0$. Then there exists $\delta = \delta(\varepsilon, l, \alpha) > 0$ such that any area-preserving mapping (2.3) of A into $\mathbb{R}^2$ with $f, g \in C^l$ such that*

$$|f - r|_{C^l} + |g - \alpha|_{C^l} < \nu\delta,$$

possesses an invariant curve of the form

$$r = c + u(\xi),\ \theta = \xi + v(\xi),$$

in A, where u, v are of class C^1, 2π-periodic, such that $|u|_{C^1} + |v|_{C^1} < \varepsilon$, and $c \in]a,b[$ is constant. Moreover, the induced mapping on this curve is given by $\xi \to \xi + \omega$, where ω is incommensurable with 2π, and satisfies infinitely many conditions

$$\left| \frac{\omega}{2\pi} - \frac{p}{q} \right| \geq \gamma q^{-\tau},$$

with some positive γ, τ, for all integers $q > 0$, and p. In fact, each choice of ω in the range of α satisfying the above Diophantine inequalities gives rise to such an invariant curve.

Call $\phi : A \to A$ a *monotone twist homeomorphism* if it preserves orientation, preserves boundary components of A and if for a lift $F(r,\theta) = (f(r,\theta), g(r,\theta))$, the function $g(\cdot,\theta)$ is a strictly monotone function for each θ. For definiteness, we assume this function to be strictly increasing. Let $F^j(r,\theta) = (r_j, \theta_j)$, and

$$\alpha_r(\phi) = \lim_{j\to\infty} \frac{\theta_j}{j}$$

be its *rotation number*. The *twist interval* of ϕ is the interval $[\alpha_a(\phi), \alpha_b(\phi)]$. It is defined up to an integral translation. If $\phi^q(z) = z$, then $F^q(r,\theta) = T^p(r,\theta)$, for some integer p determined up to a multiple of q, and $T(r,\theta) = (r, \theta + 2\pi)$. $\frac{p}{q}$ is called the *rotation number* of z. One calls such a point $z = (r,\theta)$ a *Birkhoff point of type (p,q)* if there exists a sequence $(r_n, \theta_n)_{n\in\mathbb{Z}}$ such that $(r_0, \theta_0) = (r, \theta)$, $\theta_{n+1} > \theta_n$, $(n \in \mathbb{N})$, $(r_{n+q}, \theta_{n+q}) = (r_{n+q}, \theta_n + 2\pi)$, $(r_{n+q}, \theta_{n+q}) = F(r_n, \theta_n)$.

One then has the *Birkhoff's twist theorem* [24].

Lemma 4. *Let $\phi : A \to A$ be an area-preserving monotone twist homeomorphism and*

$$\frac{p}{q} \in [\alpha_a(\phi), \alpha_b(\phi)]$$

be a rational number with p, q relatively prime. Then there exist two Birkhoff periodic orbits of type (p,q) for ϕ.

A *Mather set* of rotation number α for F is a closed invariant set for F with representation $u = u(\theta), v = v(\theta)$ where u is monotone increasing, $u - Id$ and v are 2π-periodic (not necessarily continuous!), and $u(\theta+\alpha) = \phi_1(u,v),\ v(\theta+\alpha) = \phi_2(u,v)$.

The following result is the *Aubry-Mather's twist theorem* [12,109].

Lemma 5. *Let $\phi : A \to A$ be an area-preserving monotone twist homeomorphism and let $\alpha \in [\alpha_a(\phi), \alpha_b(\phi)]$. Then there exists an invariant Mather set Γ_α with rotation number α. Furthermore, Γ_α is a subset of a closed curve $y = w(x)$ where w is 2π-periodic and Lipschitz continuous, i.e. $v(\theta) = w(u(\theta)$. For rational $\alpha = \frac{p}{q}$, this theorem provides orbits (r_j, θ_j) satisfying $\theta_{j+q} = \theta_j + 2p\pi$, $r_{j+q} = r_0$ for $j \in \mathbb{Z}$.*

For some surveys on the Aubry-Mather's twist theorem, see [14,43,92].

2.2.2 Lyapunov-Schmidt's method

The *Lyapunov-Schmidt's method* (see e.g. [78]) is based upon the following elementary fact.

Lemma 6. *$y = \overline{y} + \widetilde{y}$ is a T-periodic solution of equation (2.1) if and only if it is a solution of the system*

$$\widetilde{y}'' + c\widetilde{y}' + a\sin(\overline{y} + \widetilde{y}) = \overline{a\sin(\overline{y} + \widetilde{y})} + \widetilde{h}(t), \quad \overline{a\sin(\overline{y} + \widetilde{y})} = \overline{h} \tag{2.4}$$

In the classical Lyapunov-Schmidt's method, the first equation in (2.4) is solved with respect to $\widetilde{y}$ for fixed $\overline{y}$ (using a fixed point or implicit function theorem, or critical point theory) and this solution is introduced in the second equation, which then becomes the (one-dimensional) *bifurcation equation*. One can also study directly the equivalent system (2.4) by *degree theory* or *critical point theory*.

2.2.3 Upper and lower solutions

The *method of upper and lower solutions* for the periodic solutions of equation (5) (see e.g. [112]) consists in the following statement.

Lemma 7. *If α and β are of class C^2, T-periodic and such that, for all $t \in \mathbb{R}$,*
i) $\alpha(t) \le \beta(t)$
ii) $\alpha''(t) + c\alpha'(t) + a\sin\alpha(t) \ge h(t) \ge \beta''(t) + c\beta'(t) + a\sin\beta(t)$,
then (2.1) has at least one T-periodic solution y such that $\alpha(t) \le y(t) \le \beta(t)$.

The reader will easily state the analogous statement for the periodic solutions of (2.2).

2.2.4 Critical point theory

The starting point of the use of a *variational method* or of *critical point theory* to the periodic solutions of the forced pendulum equation without dissipation is the following classical observation.

Lemma 8. *y is a T-periodic solution of*

$$y'' + a\sin y = h(t) \tag{2.5}$$

if and only if y is a critical point of the action functional

$$A_h : H^1_T \to \mathbb{R}, y \mapsto \int_0^T \left(\frac{y'^2(t)}{2} + a\cos y(t) + h(t)y(t) \right) dt. \tag{2.6}$$

Various tools of critical point theory like *minimization, mountain pass lemma, Lyusternik-Schnirelmann theory, Morse theory* (see e.g. [124]) can be applied to (2.5) or to its equivalent form (2.2). Notice that a semi-variational method has been used in [1] to study the dissipative forced pendulum.

2.3 Results valid for all c, a, T, h

Rewrite equation (2.1) as

$$y'' + cy' + a\sin y = \overline{h} + \widetilde{h}(t) \tag{2.7}$$

The following results are now classical and can be found in [124,68,112]. Their proof uses Lyapunov-Schmidt's argument, topological degree, upper and lower solutions. Some of them can already been found in [47] and some have been reobtained in [91].

Theorem 1. *For each $\widetilde{h} \in L^1_T$, there exists*

$$m_{\widetilde{h}} = m_{\widetilde{h}}(c, a, T) \le M_{\widetilde{h}} = M_{\widetilde{h}}(c, a, T)$$

such that the following hold.

1. $-a \le m_{\widetilde{h}} \le M_{\widetilde{h}} \le a$ *et* $-a = m_0 < M_0 = a$.
2. $m_{\widetilde{h}_k} \to m_{\widetilde{h}}$ *and* $M_{\widetilde{h}_k} \to M_{\widetilde{h}}$ *if* $\widetilde{H}_k \to \widetilde{H}$ *uniformly on* $\mathbb{R}$.
3. *Equation* (2.7) *has at least one T-periodic solution if and only if* $\overline{h} \in [m_{\widetilde{h}}, M_{\widetilde{h}}]$.
4. *Equation* (2.7) *has at least two distinct T-periodic solutions if* $\overline{h} \in]m_{\widetilde{h}}, M_{\widetilde{h}}[$.
5. *If* $m_{\widetilde{h}} = M_{\widetilde{h}}$, *equation* (2.7) *has, for each* $\xi \in \mathbb{R}$, *at least one T-periodic solution y with* $\overline{y} = \xi$.

In particular, $\mathcal{R}(c, a, T)$ is closed and

$$\mathcal{R}(c, a, T) = \bigcup_{\widetilde{h} \in \widetilde{C_T}} [m_{\widetilde{h}}, M_{\widetilde{h}}] \times \{\widetilde{h}\} \subset [-a, a] \times \widetilde{C_T}.$$

2.4 Open problems and partial solutions.

Some important questions are left open by the results of Theorem 1, and are only partially solved.

2.4.1 Find an explicit element in $[m_{\widetilde{h}}, M_{\widetilde{h}}]$

Theorem 2. *When $c = 0$, then $0 \in [m_{\widetilde{h}}, M_{\widetilde{h}}]$.*

This is shown by proving the existence of a global minimum for the action functional A_h ([86,179,180,47]). The reason of the success of the minimization method is that $A_h(y + 2\pi) = A_h(y)$ if and only if $\overline{h} = 0$. This property together with the coercivity of A_h with respect to $\widetilde{y}$ allows easily to obtain a bounded minimizing sequence for A_h. The periodicity property of A_h when $\overline{h} = 0$ allows also the use of a Lusternik-Schnirelmann type argument to prove directly that A_h has two distinct critical points (see [119,40,149]). Another proof of this fact has been given in [71] using a generalized Poincaré-Birkhoff theorem. No proof based upon degree theory is known at this day.

Theorem 3. *When $\frac{c}{T} > \frac{1}{\pi\sqrt{3}}\|\widetilde{h}\|_2$, then $0 \in]m_{\widetilde{h}}, M_{\widetilde{h}}[$.*

This is proved by topological degree arguments ([124]).

The question was then raised to know if $0 \in [m_{\widetilde{h}}, M_{\widetilde{h}}]$ for each $c > 0$. A *negative answer* was first given by a counterexample of Ortega [140], recently improved by another one of Alonso [2] showing that *for each $c > 0$, there exists $T_0 = T_0(a, c)$ such that for each $T > T_0$, $0 \notin [m_{\widetilde{h}}, M_{\widetilde{h}}]$*. The idea of Alonso's counterexample consists in constructing a forcing term close to a piecewise constant function $h(t)$ taking a large positive value p in the interval $[0, \tau]$ and a small negative value $-q$ in the interval $[\tau, T]$, where $p\tau - q(T - \tau) = 0$.

2.4.2 Prove or disprove the existence of some $\widetilde{h}$ such that $m_{\widetilde{h}} = M_{\widetilde{h}}$

This problem remains open. Here is some known partial information.

Theorem 4. *The set $\{\widetilde{h} \in \widetilde{C_T} : m_{\widetilde{h}} < M_{\widetilde{h}}\}$ is open and dense.*

This has been proved using various arguments [124,112,108], and in particular a *generalized Sard-Smale's theorem.* Thus, *generically,* $[m_{\widetilde{h}}, M_{\widetilde{h}}]$ is a non degenerate interval.

Theorem 5. *For $c = 0$,*

$$\{\widetilde{h} \in \widetilde{C_T} : \lim_{|\lambda|\to\infty} m(\lambda\widetilde{h}) = \lim_{|\lambda|\to\infty} M(\lambda\widetilde{h}) = 0\}$$

contains an open and dense subset of $\widetilde{C_T}$.

This has been proved by Kannan and Ortega [91], who also gave an example showing that this set is not open. The proof makes use of some Riemann-Lebesgue lemma and asymptotic analysis techniques.

2.5 The conservative case $c = 0$

We shall now concentrate on some results which hold for the conservative case $c = 0$. Recall the a *regular value* for a continuously differentiable mapping f between two smooth Banach manifolds is the image by f of a point c such that f'_c is onto.

Theorem 6. *The set $\mathcal{G}$ of regular values for $y'' + a \sin y$ is open and dense in $\widetilde{C_T}$, and, for every $g \in \mathcal{G}$, there exists $\varepsilon > 0$ such that, if $\|h - g\|_\infty \leq \varepsilon$, then equation (2.5) has a T-periodic solution.*

This has been proved [108] using a generalized Sard-Smale lemma.

Recently, using techniques of critical point theory (a suitable *minimax method*), Serra, Tarallo and Terracini [167] have introduced a new condition in order that $m_{\widetilde{h}} < M_{\widetilde{h}}$.

Theorem 7. *If $\overline{h} = 0$, and if $c_0 = \inf_{H^1_T} A_h$, then $m_{\widetilde{h}} < M_{\widetilde{h}}$ if and only if the following condition*

$$(K_0) \qquad \mathcal{K}(\xi) := \{y \in H^1_T : A_h(y) = c_0,\ \overline{y} = \xi\} = \emptyset \text{ for some } \xi \in \mathbb{R}$$

holds. Moreover, if (K_0) does not hold, then, for each $\xi \in \mathbb{R}$, $\mathcal{K}(\xi) = \{y_\xi\}$, with $\xi \to y_\xi$ continuous and $y_{\xi_1}(t) < y_{\xi_2}(t)$ for all $t \in \mathbb{R}$ whenever $\xi_1 < \xi_2$, and equation (2.5) has no other periodic solutions.

In a subsequent paper [166], Serra and Tarallo have introduced a new *reduction method of Lyapunov-Schmidt's type,* which sheds some light on some of the unsolved problems for the conservative forced pendulum equation.

Theorem 8. *For each $\xi \in \mathbb{R}$, let*

$$\varphi_h(\xi) := \min_{\overline{y}=\xi} A_h(y), \qquad M_h(\xi) = \{y \in H^1_T : \overline{y} = \xi,\ A_h(y) = \varphi_h(\overline{y})\},$$

and let

$$M_h = \bigcup_{\xi \in \mathbb{R}} M_h(\xi) = \{u \in H^1_T : A_h(u) = \varphi_h(\overline{u})\}.$$

Then the following results hold.

1. *φ_h is defined and locally Lipschitz continuous on $\mathbb{R}$.*
2. *$M_h(\xi) \neq \emptyset$ and compact for each $\xi \in \mathbb{R}$ and $M_h : \mathbb{R} \to 2^{H^1_T}$ upper semi-continuous.*
3. *If $y \in M$ and $\overline{y}$ is a local minimum for φ_h, then y is a local minimum for A_h.*
4. *φ_h is differentiable at ξ if and only if $y \mapsto \int_0^T (a \sin y(t) - h(t))\, dt$ is constant on $M_h(\xi)$.*
5. *If φ_h has a critical point, then A_h has a critical point.*
6. *If φ_h is not strictly monotone, then A_h has a critical point.*

It is interesting to compare this approach to the classical method of Lyapunov-Schmidt. In this case, one proves (by critical point theory if $c = 0$ and Schauder's fixed point theorem in all cases) that, for each $\xi \in \mathbb{R}$, the set

$$K_h(\xi) = \{y \in C_T : \overline{y} = \xi \text{ and } \widetilde{y} \text{ solves the first equation in (2.4)}\}$$

is not empty, and then the problem is reduced to find the elements of the set $K_h = \bigcup_{\xi \in \mathbb{R}} K_h(\xi)$ such that $\overline{a \sin y} = \overline{h}$. In the Serra-Tarallo's approach, on each *slice* $\xi + \widetilde{H_T^1}$ of H_T^1, one considers only the elements of $K_h(\xi)$ which minimize the restriction of A_h on this slice, which provides the subset $M_h(\xi) \subset K_h(\xi)$, and then, instead of trying to solve the second equation of (2.4) on this set, one concentrates on the reduced functional φ_h and relates its critical points to those of A_h. Hence the spirit is more variational than in the earlier approaches combining a Lyapunov-Schmidt argument with some variational method, in that the emphasis, at each step, remains on the functional instead of on its gradient. Because the minimization is made on each slice on the function space, one can imitate the type of humor which has led from the name *Klein-Gordon equation* for $u_{tt} - \Delta u + u = 0$ to the name *Sine-Gordon equation* for $u_{tt} - \Delta u + \sin u = 0$, and call the Serra-Tarallo's approach a *Lyapunov-Schnitt's method.*

Notice that one of the main features of this approach is that, in contrast to most other ones, it applies when $a \sin y$ is replaced by a more general almost periodic function.

2.6 The case where $c = 0$ and $a < \omega^2$

In the conservative case, more precise results can be obtained when the following condition

$$a < \omega^2 \tag{2.8}$$

holds.

Using global analysis and *singularity theory,* Donati [51] has proved the following result about the multiplicity of solutions.

Theorem 9. *If (2.8) holds and $\overline{h} \in [m_{\widetilde{h}}, M_{\widetilde{h}}]$, then equation (2.5) has at most finitely many distinct T-periodic solutions when $[m_{\widetilde{h}}, M_{\widetilde{h}}] \neq \{0\}$. Otherwise, equation (2.5) has an analytic unbounded curve of solutions.*

Serra and Tarallo [166] have used their *Lyapunov-Schnitt's method* to obtain more precise information.

Theorem 10. *Assume that (2.8) holds. Then*

1. *If φ_h is constant, then $M_h(\xi) = \{y_\xi\}$, and if y is a periodic solution of equation (2.5), then $\overline{h} = 0$ and $y = y_\xi$ for some $\xi \in \mathbb{R}$.*
2. *φ_h is not constant if and only if there exists $\varepsilon_0 > 0$ such that equation (2.5) has at least one T-periodic solution for each $|\overline{h}| < \varepsilon_0$.*

3. $[m_{\widetilde{h}}, M_{\widetilde{h}}] = \{0\}$ *if and only if* φ_h *is constant.*
4. $\{\widetilde{h} \in \widetilde{C_T} : \varphi_{\widetilde{h}}$ *is not constant*$\}$ *is open and dense in* $\widetilde{C_T}$.
5. *If* φ_h *is constant and* $\overline{h} \neq 0$, *then equation* (2.5) *has no bounded solution.*

The same approach has also been used by Calanchi and Tarallo [30] to show the following result.

Theorem 11. *There exists* $K = K(a,T) > 0$ *such that if* $\|h\|_2 < K$, *each critical point of* A_h *over* H^1_T *is a local minimum or a point of mountain pass type.*

2.7 Stability of the T-periodic solutions

2.7.1 The dissipative case $c > 0$

By imposing some restrictions upon c, a, and T, it is possible to obtain on one hand *exact multiplicity results* for the T-periodic solutions, and, on the other hand, informations upon their *Lyapunov stability*. The pioneering work in the first direction is due to Tarantello [172] (using a Lyapunov-Schmidt approach) and, in the second direction, to Ortega [141,142,143] (using some relations between stability and the Brouwer degree of Poincaré's operator). A recent paper of Čepička, Drábek and Jenšiková [39] provides the sharpest known conditions.

Theorem 12. *If*

$$c > 0, \quad a < \max\left\{\frac{c^2}{4} + \omega^2, \omega\sqrt{c^2+\omega^2}\right\}$$

then equation (2.7) *has :*

1. *exactly one* T*-periodic solution if either* $\overline{h} = m_{\widetilde{h}}$ *or* $\overline{h} = M_{\widetilde{h}}$.
2. *exactly two* T*-periodic solutions if* $\overline{h} \in]m_{\widetilde{h}}, M_{\widetilde{h}}[$.

If

$$c > 0, \quad a < \max\left\{\frac{c^2+\omega^2}{4}, \frac{\omega}{2}\sqrt{c^2+\frac{\omega^2}{4}}\right\},$$

then the conclusions (1.-2.) remain true and the periodic solution obtained in (1.) is unstable while one solution obtained in (2.) is asymptotically stable and the other unstable.

The proof of the exact multiplicity results in Theorem 5 is based upon the Lyapunov-Schmidt's reduction method together with the *real analytic version of the implicit function theorem* to analyze the bifurcation equation. The uniqueness in the solution of the first equation in (2.4) is deduced from some preliminary assertions on the T-periodic solutions of linear equations of the type

$$y'' + cy' + g(t)y = 0,$$

with g T-periodic. The stability conclusion is obtained in the same way as in Ortega's papers.

2.7.2 The conservative case $c = 0$

The difficulty in analyzing the stability in the conservative case is that asymptotic stability can no more be expected. In a recent paper, Dancer and Ortega [48] have proved the following proposition.

Lemma 9. *A stable isolated fixed point of an orientation preserving local homeomorphism on $\mathbb{R}^2$ has fixed point index equal to one.*

The proof of this result depends upon a variant of Brouwer's lemma on translation arcs. One of the given applications is the following result.

Lemma 10. *If y is an isolated T-periodic solution of the second order equation, with continuous right-hand member T-periodic with respect to t,*

$$y'' = \frac{\partial V}{\partial y}(t, y), \tag{2.9}$$

and y reaches a local minimum on H^1_T of the action functional

$$f(y) = \int_0^T \left(\frac{y'^2(t)}{2} + V(t, y(t)) \right) dt,$$

then y is unstable.

This result is proved by showing first, through a result of Amann on the computation of degree of gradient mappings and a relatedness principle of Krasnosel'skii-Zabreiko, that the index of y is equal to minus one. The result then follows from the previous one.

An immediate consequence for the pendulum equation is the following one.

Theorem 13. *If $\overline{h} = 0$, and if a T-periodic solution minimizing $A_{\widetilde{h}}$ is isolated, then it is unstable.*

One can then raise the question to known if the above results still hold without the assumption that the T-periodic solution is isolated. Ortega [145] has proved the following interesting result.

Lemma 11. *If $D \subset \mathbb{R}$ is a domain and $F : D \subset \mathbb{R}^2 \to \mathbb{R}^2$ is real analytical and not the identity on D, its Jacobian is equal to 1 on D, and if p is a stable fixed point of F, then p is isolated in the fixed points set of F.*

The delicate proof of this result uses Brouwer's plane translation theorem.

As an application, the following unstability result is proved in [145].

Lemma 12. *If V is T-periodic with respect to t and real analytic, and y is a T-periodic solution of equation (2.9) such that y reaches a local minimal of f on H^1_T, then y is unstable.*

An immediate consequence for the forced pendulum equation is the following one.

Theorem 14. *If h is analytical and $\overline{h} = 0$, then, given $N \in \mathbb{Z}$, the number of T-periodic solutions of equation (2.5) that are stable and geometrically different is finite.*

2.8 Existence of more than two T-periodic solutions

In [50], Donati proved that *given $a > 0$ and $T > 0$, there exists some $h^* \in C^T$ with $\overline{h^*} = 0$ and a neighborhoud V of h^* such that for each $h \in V$, equation (2.5) has at least four distinct T-periodic solutions.* The proof is based upon a *classification of singularities of the nonlinear Fredholm operator* $\frac{d^2}{dt^2} + a\sin(\cdot)$. Applying to (2.2) a classical *perturbation method* as used for example by Loud for Duffing's equation, Ortega [146] has recently improved this result by replacing 4 by any even number.

Theorem 15. *Given $a > 0$ and an integer $N \geq 1$, there exists $h^* \in C^T$ satisfying $\overline{h^*} = 0$ and such that equation (2.5) with h replaced by h^* has at least $2N$ distinct T-periodic solutions. In addition, there exists $\delta > 0$ such that if h satisfies $\overline{h} = 0$ and $\|h - h^*\|_{L^1} < \delta$, then the conclusion also holds for equation (2.5).*

The idea of the proof consists in considering the equation

$$y'' + a\sin(y + P_0(t)) = 0, \tag{2.10}$$

where

$$P_0(t) = 2\pi\left(\frac{t}{T} - \left[\frac{t}{T}\right]\right),$$

which has a continuum $(y_c)_{c\in\mathbb{R}}$ of T-periodic solutions, and in considering a perturbation of equation (2.10)

$$y'' + a\sin(y + P_0(t) + \Psi(t,\varepsilon)) = 0, \tag{2.11}$$

with conditions upon Ψ insuring that $P_0(t) + \Psi(t,\varepsilon)$ is smooth and that one has at least 2N periodic simultaneous bifurcations for $\varepsilon = 0$.

To motivate a further multiplicity result of perturbation type, let us recall that for the undamped free pendulum equation

$$y'' + a\sin y = 0 \tag{2.12}$$

it is known that the period $T(A)$ of the periodic solutions of (2.12) as a function of their amplitude $A > 0$ is an increasing function such that

$$\lim_{A\to 0+} T(A) = \frac{2\pi}{\sqrt{a}}, \quad \lim_{A\to\pi-} T(A) = +\infty.$$

Consequently, given any positive integer N, then, if $a > \frac{4\pi^2 N^2}{T^2}$, equation (2.12) has a closed orbit with least period $\frac{T}{k}$ for each $k = 1, 2, \ldots, N$. Using a perturbation argument and W. Ding's generalization of the Poincaré-Birkhoff fixed point theorem for area-preserving twist mappings of an annulus, Fonda and Zanolin [65] have proved the following result for the forced case.

Theorem 16. *Given any positive integer N, there exists a constant $a_0 > 0$ such that, for any $a \geq a_0$, equation (2.5) has at least N periodic solutions with minimal period T, which can be chosen to have exactly $2j$ simple crossings with 0 in the interval $[0, T[$, with $j = 1, 2, \ldots, N$.*

2.9 Subharmonic solutions in the conservative case $c = 0$

Let us first recall that, if $k \geq 2$ is an integer, a *subharmonic solution* of order k of (2.1) is a periodic solution of equation (2.1) with minimal period kT. The first existence results for the subharmonic solutions of equation (2.5) with $\overline{h} = 0$ have been obtained by Fonda and Willem [64] (see also Offin [137] for a close result based upon an index theory for periodic extremals and a variant of the mountain pass lemma).

Theorem 17. *Suppose that the T-periodic solutions of equation (2.5) are isolated and that every T-periodic solution of equation (2.5) having Morse index equal to zero is nondegenerate. Then there exists $k_0 \geq 2$ such that, for every prime integer $k \geq k_0$, there is a periodic solution of equation (2.5) with minimal period kT. If moreover the kT-periodic solutions of equation (2.5) are nondegenerate for $k = 1$ and every prime integer k, then there exists a $k_0 \geq 3$ such that, for every prime integer $k \geq k_0$, there are two periodic solutions of equation (2.5) with minimal period kT.*

To prove this result, Fonda and Willem consider the critical points of the functional

$$A_{h,k} = \int_0^{kT} \left(\frac{y'^2(t)}{2} + a \cos y(t) + h(t) y(t) \right) dt,$$

over the Sobolev space H^1_{kT}. Then, by assumption and an easy reasoning, $A_{h,1} = A_h$ has a finite number of critical points $y_0, y_1, \ldots, y_n$, which, of course, are also critical points of $A_{h,k}$ for any $k \geq 2$. The first ingredient of the proof consists in showing the existence of some integer k_0 such that, for $k \geq k_0$ and $0 \leq i \leq n$, either the Morse index $J(y_i, kT, 1)$ of y_i is equal to 0 and y_i is nondegenerate, or $J(y_i, kT, 1) \geq 2$. This is done using an iteration formula for the Morse index due to Bott. Now, let $k \geq k_0$ be a prime number, so that the critical points of $A_{h,k}$ have minimal period T or kT. Assuming by contradiction that $y_0, \ldots, y_n$ are the only critical points of $A_{h,k}$, one is led to a contradiction in the Morse inequalities of *Morse theory* (see e.g. [126]) applied to $A_{h,k}$. The proof of the second part of Theorem 17 is similar.

Combining the Fonda-Willem's theorem with the generic results of [108], one gets the *generic existence of subharmonic solutions.*

Theorem 18. *There exists an open dense subset $\mathcal{G}$ of $\widetilde{C_T}$ such that for every $h \in \mathcal{G}$, there exists a $k_0 \geq 2$ such that, for every prime integer $k \geq k_0$, equation (2.5) has a periodic solution with minimal period kT.*

As shown in [167], the *Lyapunov-Schnitt's reduction method* also provides some information about subharmonic solutions, by relating their existence to the properties of φ_h.

Theorem 19. *Equation (2.5) with $\overline{h} = 0$ has subharmonics of infinitely many distinct levels if and only if φ_h is not constant. If $c_0^T := \min_{H_T^1} A_h$ is isolated in the set of critical levels of A_h, then equation (2.5) with $\overline{h} = 0$ admits subharmonics of arbitrary large minimal period if and only if φ_h is not constant. Finally, the isolatedness assumption in the previous statement can be dropped if $a < \omega^2$.*

Finally, the Fonda-Zanolin multiplicity result [65] has a counterpart for subharmonic solutions, proved using the same technique.

Theorem 20. *Given any two positive integers M, N, there exists a constant $a_0 > 0$ such that, for any $a \geq a_0$, equation (2.5) has, for each $k = 1, 2, \ldots, M$, at least N periodic solutions with minimal period kT.*

2.10 Rotating solutions in the conservative case $c = 0$

Besides periodic solutions, the free pendulum has also *rotating solutions* which are the sum of a linear function of t and of a periodic term. Under some conditions, the conservative forced pendulum (2.5) can also admit such solutions. Most of the results in this case are obtained via combination of Poincaré's method and some theorem for twist mappings.

The following results have been proved by Levi [103] using Moser's twist theorem. The basic idea is that, for large velocities $x = y'$, the forced pendulum equation has solutions which are close to those of the integrable system $y'' = 0$.

Theorem 21. *For any $\omega \in]0, 2\pi[$ satisfying, for some $c_0 > 0$ and $\mu > 0$, the set of inequalities*

$$\left| \frac{\omega}{2\pi} - \frac{m}{n} \right| > \frac{c_0}{n^{2+\mu}},$$

for all $m, n \in \mathbb{Z}$ with $n \neq 0$, there exists an integer $k_0 = k_0(c_0, \mu)$ such that the Poincaré's mapping associated to (2.5) possesses, for all integers k with $|k| \geq k_0$, a countable set of invariant curves $y = f_{\frac{\omega}{2\pi}+k}(x) \equiv f_{\frac{\omega}{2\pi}+k}(x+1)$. For any real number α, equation (2.5) has a Birkhoff orbit with that rotation number. For any rational $\alpha = \frac{p}{q}$ there exists at least two solutions satisfying $y(t+qT) = y(t)+2p\pi$.

A similar result was proved independently by Moser [130], using a variational method which can be traced to Percival and Mather (see [131]).

Theorem 22. *If $\overline{h} = 0$, then, for some sufficiently large irrational α (satisfying a Diophantine condition), equation (2.5) has solutions of the form $y(t) = U(t, \alpha t)$ such that $U(t, \theta) - \theta$ is continuous, T-periodic in t and 2π-periodic in θ, and $\partial_\theta U > 0$.*

Physically, the above result means that there exists a motion with any average angular velocity (see also Dovbysh [52]).

The following result of You [185] is also proved using Moser's twist theorem.

Theorem 23. *Equation (2.5) admits an infinite number of invariant tori, and thus an infinite number of almost periodic solutions, when* $\overline{h} = 0$, *and no invariant torus when* $\overline{h} \neq 0$.

In the case of an analytic h, Ortega's approach described in Section 2.7.2 provides some information about the number and stability of rotating solutions [145].

Theorem 24. *If* h *is analytic and* $\overline{h} = 0$, *then, given* $N \in \mathbb{Z}$, *the number of stable and distinct* T*-periodic solutions with winding number* N *(i.e. solution such that* $y(t+T) = y(t) + 2N\pi$*) of equation (2.5) is finite.*

Finally, the change of variable $y(t) = k\omega t + v(t)$, and the use of direct methods of the calculus of variations to the transformed equation allows a very simple proof of the following special case of Theorem 13 [121].

Theorem 25. *For each* $a > 0$, $T > 0$, $k \in \mathbb{Z} \setminus \{0\}$, *and each* h *with* $\overline{h} = 0$, *equation (2.5) has at least one solution of the form* $y(t) = k\omega t + v(t)$ *with* v T*-periodic.*

2.11 Lagrange stability

2.11.1 The conservative case $c = 0$

Equation (2.5) is called *Lagrange stable* if any solution of (2.1) is bounded over $\mathbb{R}$ in the phase cylinder $\{(y \bmod 2\pi, y')\}$. Physically, this means that any solution of (2.1) has angular velocity bounded over $\mathbb{R}$.

The problem of the Lagrange stability of equation (2.5) was raised by Moser in the Introduction of [129]. Its positive solution is a consequence of the results of Levi, Moser and You described in the previous section.

Theorem 26. *If* $\overline{h} = 0$, *then for any sufficiently large* $N > 0$, *there exists* $M = M(N)$ *such that any solution* $y(t)$ *of equation (2.5) with* $|y'(0)| \leq M$ *satisfies* $|y'(t)| \leq N$ *for all* $t \in \mathbb{R}$.

As shown by You [185], the conditions that the mean value of h is zero is necessary and sufficient for the Lagrange stability.

Theorem 27. *If* $a > 0$, *then equation (2.5) is Lagrange stable if and only if* $\overline{h} = 0$.

2.11.2 The dissipative case $c > 0$

Some results for Lagrange stability in the dissipative case have been obtained by Andres [4,5] and Andres-Staněk [9] using Lyapunov function techniques. See also [6,7,8] for further discussions and problems.

Here, *Lagrange stability* of (2.1) has to be understood as the boundedness *over* $\mathbb{R}_+$ of any solution in the phase cylinder $\{y \bmod 2\pi, y'\}$. Physically, that means that any solution of (2.1) has angular velocity bounded in the future.

Theorem 28. *The equation* (2.1) *is Lagrange stable provided* $\overline{h} = 0$ *and*

$$c > \frac{(a + \|h\|_\infty)\left\{\|H\|_\infty + \left[\|H\|_\infty^2 + 4(2a + \pi(a + \|h\|_\infty))\right]^{1/2}\right\}}{2(2a + \pi(c + \|h\|_\infty))},$$

where $H(t) = \int_0^t h(s)\,ds$.

3 Bounded or almost periodic forcing

3.1 Bounded forcing

Using a version of the *method of upper and lower solutions* for solutions bounded over $\mathbb{R}$ going back to to Opial [138] (see also [122]), one can prove the following result, which is the one dimensional case of a result for elliptic partial differential equations due to Fournier, Szulkin et Willem [69]. Consider the dissipative forced pendulum-type equation

$$y'' + cy' + a \sin y = h(t), \tag{3.1}$$

where $a > 0$, $c \geq 0$, and $h : \mathbb{R} \to \mathbb{R}$ is continuous and bounded.

Theorem 29. *If* $c \geq 0$ *and if* $h : \mathbb{R} \to \mathbb{R}$ *continuous is such that*

$$-a \leq h(t) \leq a, \tag{3.2}$$

for all $t \in \mathbb{R}$, *then equation* (3.1) *has at least one solution* y *such that*

$$\frac{\pi}{2} \leq y(t) \leq \frac{3\pi}{2}$$

for all $t \in \mathbb{R}$. *If condition* (3.2) *is restricted to*

$$\|h\|_\infty < a, \tag{3.3}$$

then there exists $\varepsilon > 0$ *such that equation* (3.1) *has a unique solution* y *such that*

$$\frac{\pi}{2} + \varepsilon \leq y(t) \leq \frac{3\pi}{2} - \varepsilon \tag{3.4}$$

for all $t \in \mathbb{R}$.

Using the equivalent formulation for the forced pendulum problem together with some results of Ortega on bounded solutions of second order linear equations [144] (see also [122]), one can prove in a similar way the following existence and uniqueness theorem [123].

Theorem 30. *If $c > 0$, $h = h^* + h^{**}$ where h^{**} is bounded and h^* has a bounded primitive over $\mathbb{R}$, and if inequalities*

$$osc_{\mathbb{R}} H_c^* \leq \pi,$$

and

$$\|h^{**}\|_\infty \leq a \cos\left(\frac{osc_{\mathbb{R}} H_c^*}{2}\right),$$

hold, where H_c^* is the unique bounded solution of $y'' + cy' = h^*(t)$, then equation (3.1) has at least one solution y such that

$$\frac{\pi}{2} + H_c^*(t) \leq y(t) \leq \frac{3\pi}{2} + H_c^*(t),$$

for all $t \in \mathbb{R}$. If the inequalities above are strenghtened to

$$osc_{\mathbb{R}} H_c^* < \frac{\pi}{2}, \tag{3.5}$$

$$\|h^{**}\|_\infty \leq \frac{a\sqrt{2}}{2}\left[\sin\left(\frac{osc_{\mathbb{R}} H_c^*}{2}\right) + \cos\left(\frac{osc_{\mathbb{R}} H_c^*}{2}\right)\right], \tag{3.6}$$

then there exists $\varepsilon > 0$ such that equation (3.1) has a unique solution y satisfying the inequality

$$\frac{\pi}{2} + \varepsilon \leq y(t) \leq \frac{3\pi}{2} - \varepsilon, \tag{3.7}$$

for all $t \in \mathbb{R}$. When $c = 0$, the above results hold if $h^{**} = 0$, $h = h^*$ has a second primitive H^1 bounded over $\mathbb{R}$ and H_c^* is replaced by H^1 in (3.5).

3.2 Particular almost periodic forcings

3.2.1 A class of almost periodic functions

The following classes of almost periodic functions was introduced by Belley, Fournier and Saadi Drissi [19,20,21]. Given a countable set $\Gamma \subset \mathbb{R}$, symmetric with respect to the origin, put

$$C_\Gamma = \left(\sum_{\lambda \in \Gamma \setminus \{0\}} \frac{1}{\lambda^2}\right)^{1/2}.$$

Let $P_\Gamma(\mathbb{R})$ denote the class of all (real-valued) trigonometric polynomials $p(t) = \sum_{\lambda \in \Gamma} \alpha_\lambda e^{i\lambda t}$ where all but finitely many of the coefficients α_λ vanish, and $\alpha_{-\lambda}$

is the complex conjugate of α_λ. On $P_\Gamma(\mathbb{R})$ one can put the uniform norm $\|\cdot\|_\infty$ and the norm $\|\cdot\|_2$ associated with the inner product

$$\langle p, q\rangle = \lim_{T\to\infty} \frac{1}{T}\int_0^T p(t)q(t)\,dt.$$

Let $AP_\Gamma(\mathbb{R})$ and $B^2_\Gamma(\mathbb{R})$ denote the completion of $P_\Gamma(\mathbb{R})$ with respect to the norms $\|\cdot\|_\infty$ and $\|\cdot\|_2$ respectively. The operation $\langle\cdot,\cdot\rangle$ can be extended to $B^2_{\mathbb{R}}(\mathbb{R})$ by defining it to be the inner product on $B^2_{\mathbb{R}}(\mathbb{R})$ associated with the norm $\|\cdot\|_2$.

For any $x \in AP_{\mathbb{R}}(\mathbb{R})$, define $\widehat{x} : \mathbb{R} \to \mathbb{C}$ by

$$\widehat{x}(\lambda) = \langle x(t), e^{-i\lambda t}\rangle = \lim_{T\to\infty} \frac{1}{T}\int_0^T x(t)e^{-i\lambda}\,dt.$$

This notation can be extended to $x \in B^2_{\mathbb{R}}(\mathbb{R})$ by $\widehat{x}(\lambda) = \lim_{n\to\infty} \widehat{p_n}(\lambda)$ for any sequence $\{p_n\}$ in $P_{\mathbb{R}}(\mathbb{R})$ such that $\|p_n - x\|_2 \to 0$. For any subset X of $B^2_\Gamma(\mathbb{R})$, let $\widetilde{X} : \{x \in X : \widehat{x}(0) = 0.\}$ One often writes

$$\overline{x} = \widehat{x}(0) = \lim_{T\to\infty} \frac{1}{T}\int_0^T x(t)\,dt,$$

and $\widetilde{x}(t) = x(t) - \overline{x}$.

If $x \in B^2_\Gamma(\mathbb{R})$ and $y \in \widetilde{B}^2_\Gamma(\mathbb{R})$ are such that

$$\langle x, p'\rangle = -\langle y, p\rangle$$

for all $p \in P_\Gamma(\mathbb{R})$, then y is said to be the *weak derivative* of x. Note that y is necessarily unique in $\widetilde{B}^2_\Gamma(\mathbb{R})$, and we write $y = x'$.

3.2.2 The results

Consider first the dissipative forced pendulum-type equation

$$y'' + cy' + a\sin y = h(t), \tag{3.8}$$

where $a > 0$, $c \geq 0$, and h almost periodic.

The following result is due to Belley-Fournier-Saadi Drissi [20], and proved using a Lyapunov-Schmidt's argument modeled on that of [62].

Theorem 31. *Let $e \in B^2_\Gamma(\mathbb{R})$ be fixed and assume that the following conditions hold.*

1. $C_\Gamma < +\infty$.
2. $c > 0$ *and* $a < \frac{c}{C_\Gamma}$.
3. $\beta := C_\Gamma(C_\Gamma^{-2} + c^2)^{-1/2}a \leq \delta(\widetilde{e})$, *where*

$$\delta(\widetilde{e}) = \left[\left(\overline{\cos\widetilde{E}}\right)^2 + \left(\overline{\sin\widetilde{E}}\right)^2\right]^{1/2},$$

and $\widetilde{E}(t)$ is the unique weak almost periodic solution of equation $y'' + cy' = \widetilde{e}(t)$.

4. $|\overline{e}| \leq (\delta(\widetilde{e}) - \beta)A$.

Then there exists some $\widetilde{\mathcal{Y}}$ *in the orthogonal supplement of* $B^2(\Gamma)$ *in* $B^2(\mathbb{R})$ *such that equation* (3.8) *with* $h = e + \widetilde{\mathcal{Y}}$ *has at least one weak almost periodic solution* $y \in AP_\Gamma(\mathbb{R})$ *such that* $y' \in \widetilde{AP}_\Gamma(\mathbb{R})$ *and* $y'' \in \widetilde{B}^2_\Gamma(\mathbb{R})$.

Consider now the conservative forced pendulum equation

$$y'' + a \sin y = h(t), \tag{3.9}$$

where $a > 0$, and h almost periodic.

The following result was proved by Belley-Fournier-Saadi Drissi [19] and Belley-Fournier-Hayes [18] using a Lyapunov-Schmidt's argument modeled on that of [62].

Theorem 32. *If* $C_\Gamma < \infty$, *then given* $\xi \in \mathbb{R}$, *and* $\widetilde{e} \in \widetilde{B}^2_\Gamma(\mathbb{R})$, *there exists a function* $\gamma \in B^2_{\mathbb{R}}(\mathbb{R}) \ominus \widetilde{B}^2_\Gamma(\mathbb{R})$ *such that the equation*

$$z'' + a \sin(\xi + z) = \gamma(t) + \widetilde{e}(t),$$

holds in $\widetilde{B}^2_\Gamma(\mathbb{R})$ *for some* $z \in \widetilde{AP_\Gamma}(\mathbb{R})$ *for which the weak derivative* $z' \in \widetilde{B}^2_\Gamma(\mathbb{R})$ *exists and admits a weak derivative* $z'' \in \widetilde{B}^2_\Gamma(\mathbb{R})$. *Furthermore, if* $a < C_\Gamma^{-2}$, *this solution* z *is unique.*

3.3 General almost periodic forcing

Combining some results on the existence and uniqueness of bounded solutions over $\mathbb{R}$ with Amerio's criterion on the existence of almost periodic solutions (see e.g. [60]), Fink [59] has given in 1968 some partial extension of the *method of upper and lower solutions* to almost periodic solutions. A special case of his results is the following proposition.

Lemma 13. *Let* $c \in \mathbb{R}$, $g \in C^1(\mathbb{R}, \mathbb{R})$ *and* h *continuous and almost periodic. Assume that there exist* $a < b$ *and* $\lambda \in \mathbb{R}$ *such that* $g'(x) > 0$ *for all* $x \in [a, b]$, *and*

$$g(a) + h(t) \leq 0 \leq g(b) + h(t)$$

for all $t \in \mathbb{R}$. *Then equation*

$$y'' + cy' = g(y) + h(t)$$

has a unique almost periodic solution y *such that* $a \leq y(t) \leq b$ *for all* $t \in \mathbb{R}$.

This result implies the following existence theorem, also proved independently by Fournier-Szulkin-Willem [69] as a special case of a more general result for elliptic partial differential equations.

Theorem 33. *For each* $c \geq 0$ *and each* $h \in AP(\mathbb{R})$ *such that* $\|h\|_\infty < a$, *equation* (3.8) *has a unique solution* $y \in AP(\mathbb{R})$ *such that* $\pi/2 < y(t) < 3\pi/2$.

Indeed, the condition upon $\|h\|_\infty$ implies the existence of $\varepsilon > 0$ such that $a = \frac{\pi}{2} + \varepsilon$ and $b = \frac{3\pi}{2} - \varepsilon$ satisfy the conditions of Lemma 13. The result with $c = 0$ generalizes an earlier approximate solvability result of Blot [27] for equation (3.9), based upon variational techniques and convex analysis, which provides the existence for a dense subset of forcing functions h only.

Similar arguments applied to the equivalent formulation of the forced pendulum equation provide the following existence theorem [123].

Theorem 34. *If $c > 0$, $h = h^* + h^{**}$ where h^{**} is almost periodic and h^* has an almost periodic primitive, and if conditions (3.5) and (3.6) are satisfied, then there exists $\varepsilon > 0$ such that equation (3.8) has a unique almost periodic solution verifying inequality (3.7). If $c = 0$, and $h \in C$ has an almost periodic second primitive H^1 satisfying (3.5) with H_c^* replaced by H^1, then the same conclusion holds.*

This result when $c = 0$ generalizes an earlier approximate solvability result of Blot [28] for equation (3.9), based upon variational techniques and convex analysis, which gives existence for a dense subset of forcing functions h only.

References

1. A. Aït Ouassarad, J. M. Belley, and G. Fournier, *Note sur une méthode semi-variationnelle pour l'équation du pendule avec friction et forcing périodique de moyenne nulle*, Ann. Sci. Math. Québec **20** (1996), 109-117
2. J. M. Alonso, *Nonexistence of periodic solutions for a damped pendulum equation*, preprint
3. Z. Amine, and R. Ortega, *Existence of asymptotically stable periodic solutions of a forced equation of Liénard type*, Nonlinear Anal. **22** (1994), 993-1003
4. J. Andres, *Note on the asymptotic behavior of solutions of damped pendulum equations under forcing*, Nonlinear Anal. **18** (1992), 705–712
5. J. Andres, *Lagrange stability of higher-order analogy of damped pendulum equations*, Acta UPO 106, Phys., **31** (1992)
6. J. Andres, *Several remarks to problem of Moser and conjecture of Mawhin*, Boll. Un. Mat. Ital. (7) **7A** (1993) 377–386
7. J. Andres, *Further remarks to problem of Moser and conjecture of Mawhin*, Topological Methods in Nonlinear Anal. **6** (1995), 163–174
8. J. Andres, *Concluding remarks to problem of Moser and conjecture of Mawhin*, Ann. Math. Silesianae **10** (1996), 57–65
9. J. Andres, and S. Stanek, *Note to the Lagrange stability of excited pendulum-type equations*, Math. Slovaca **43** (1993) 617–630
10. A. A. Andronov, A. A. Vitt and S. E. Khaikin, *Theory of Oscillators*, Dover, New York, 1987
11. D. Arcoya, and A. Canada, *Critical point theorems and applications to nonlinear boundary value problems*, Nonlinear Anal. **14** (1990), 393–411
12. S. Aubry and P. Y. Le Daeron, *The discrete Fraenkel-Kontorova model and its extensions I*, Phys. D **8** (1983), 381–422

13. G. L. Baker and J. P. Gollub, *Chaotic Dynamics, an Introduction*, Cambridge University Press, Cambridge, 1990
14. V. Bangert, *Mather sets for twist maps and geoderics on tori*, Dynamics Reported **1** (1988), 1–56
15. E. A. Barbashin, and V. A. Tabueva, *Dynamical systems with a cylindrical phase space* (Russian), Nauka, Moscow, 1969
16. T. Bartsch, and Z. Q. Wang, *Periodic solutions of even Hamiltonian systems on the torus*, Math. Z. **224** (1997), 65–76
17. V. Beletski, *Essais sur le mouvement des corps cosmiques*, Mir, Moscou, 1986
18. J. M. Belley, G. Fournier, and J. Hayes, *Existence of almost periodic weak type solutions for the conservative forced pendulum equation*, J. Differential Equations **124** (1996), 205–224
19. J. M. Belley, G. Fournier, and K. Saadi Drissi, *Almost periodic weak solutions to forced pendulum type equations without friction*, Aequationes Math. **44** (1992), 100–108
20. J. M. Belley, G. Fournier, and K. Saadi Drissi, *Solutions faibles presque périodiques d'équation différentielle du type du pendule forcé*, Bull. Cl. Sci. Acad. R. Belgique **3** (1992), 173–186
21. J. M. Belley, G. Fournier, and K. Saadi Drissi, *Solutions presque périodiques du système différentiel du type pendule forcé couplé*, Bull. Cl. Sci. Acad. R. Belgique **3** (1992), 265–278
22. V. N. Belykh, N. F. Pedersen, and O. H. Soerenson, *Shunted-Josephson-junction model*, Phys. Rev. B **16** (1977), 4853–4871
23. G. D. Birkhoff, *Proof of Poincaré's last geometric theorem*, Trans. Amer. Math. Soc. **14** (1913), 14–22
24. G. D. Birkhoff, *On the periodic motions of dynamical systems*, Acta Math. **50** (1927), 359–379
25. J. A. Blackburn, Z. J. Yang, S. Vik, H. T. J. Smith, and M. A. H. Nerenberg, *Experimental study of chaos in a driven pendulum*, Physica D **26** (1987), 385–395
26. J. A. Blackburn, S. Vik, B. R. Wu, and H. T. J. Smith, *Driven pendulum for studying chaos*, Rev. Sci. Instruments **60** (1989), 422–426
27. J. Blot, *Une méthode hilbertienne pour les trajectoires presque périodiques*, C. R. Acad. Sci. Paris, Ser. I Math. **313** (1991), 487–490
28. J. Blot, *Almost periodically forced pendulum*, Funkcialaj Ekvacioj **36** (1993), 235–250
29. K. Briggs, *Simple experiments in chaotic dynamics*, Amer. J. Phys. **55** (1987), 1083–1089
30. M. Calanchi, and M. Tarallo, *On the count and the classification of periodic solutions to forced pendulum type equations*, Differential and Integral Equations (to appear)
31. A. Canada, *A note on the existence of global minimum for a noncoercive functional*, Nonlinear Anal. **21** (1993), 161–166
32. A. Canada, and F. Roca, *Some qualitative properties of the range of conservative pendulum-like operators with Dirichlet boundary conditions*, Equadiff 95, Lisbon (to appear)
33. A. Canada, and F. Roca, *Existence and multiplicity of solutions of some conservative pendulum-type equations with homogeneous Dirichlet conditions*, Differential and Integral Equations (to appear)
34. A. Canada, and F. Roca, *Qualitative properties of some noncoercive functionals arising from nonlinear boundary value problems*, preprint

35. A. Capozzi, D. Fortunato, and A. Salvatore, *Periodic solutions of Lagrangian systems with bounded potential*, J. Math. Anal. Appl. **124** (1987), 482–494
36. G. Caristi, *Periodic solutions of bounded perturbations of linear second order ordinary differential equations*, Riv. Mat. Pura Appl. **2** (1987), 53–60
37. A. Casal, and A. Somolinos, *Forced oscillations for the sunflower equation. Entrainment*, Nonlinear Anal. **6** (1982), 397–414
38. A. Castro, *Periodic solutions of the forced pendulum equation*, in Differential Equations, Ahmad, Keener, Lazer ed., Academic Press, New York, 1980 149–160
39. J. Čepička, P. Drábek, and J. Jenšiková, *On the stability of periodic solutions of the damped pendulum equation*, J. Math. Anal. Appl. **209** (1997), 712–723
40. K. C. Chang, *On the periodic nonlinearity and the multiplicity of solutions*, Nonlinear Anal. **13** (1989), 527–537
41. K. C. Chang, Y. Long and E. Zehnder, *Forced oscillations for the triple pendulum*, in Analysis, et cetera, Rabinowitz and Zehnder ed., Academic Press, New York, 1990, 177–208
42. A. Chenciner, *Systèmes dynamiques différentiables*, in Encyclopaedia Universalis, Universalia, Paris, 1978, 594–630
43. A. Chenciner, *La dynamique au voisinage d'un point fixe elliptique conservatif: de Poincaré et Birkhoff à Aubry et Mather*, Séminaire Bourbaki No. 622, Astérisque, vol. 121–122, 1985, 147–170
44. C. Conley and E. Zehnder, *The Birkhoff-Lewis fixed point theorem and a conjecture of V. I. Arnold*, Invent. Math. **73** (1983), 33–49
45. C. Conley and E. Zehnder, *A global fixed point theorem for symplectic maps and subharmonic solutions of Hamiltonian equations on tori*, in Proc. Symp. Pure Math., vol. **45**, part 1 (1986), 283–299
46. D. Costa, H. Jeggle, R. Schaaf, and K. Schmitt, *Oscillatory perturbations of linear problems at resonance*, Results in Math. **14** (1988), 257–287
47. E. N. Dancer, *On the use of asymptotics in nonlinear boundary value problems*, Ann. Mat. Pura Appl. (4) **131** (1982), 167–185
48. E. N. Dancer and R. Ortega, *The index of Lyapunov stable fixed points in two dimensions*, J. Dynamics and Differential Equations **6** (1994), 631–637.
49. D. D'Humières, M. R. Beasley, B. A. Huberman, and A. Libchaber, *Chaotic states and routes to chaos in the forced pendulum*, Phys. Rev. A **26** (1982), 3483–3496
50. F. Donati, *Sur l'existence de quatre solutions périodiques pour l'équation du pendule forcé*, C. R. Acad. Sci. Paris I – **317** (1993), 667–672
51. F. Donati, *Some remarks about periodic solutions to the forced pendulum equation*, Differential and Integral Equations **8** (1995), 141–149
52. S. A. Dovbysh, *Kolmogorov stability, the impossibility of Fermi acceleration and the existence of periodic solutions in some Hamiltonian-type systems*, J. Appl. Math. Mech. **56** (1992), 218–229
53. P. Drábek and S. Invernizzi, *Periodic solutions for systems of forced coupled pendulum-like equations*, J. Differential Equations **70** (1987), 390–402
54. B. Duchesne,C. W. Fischer, C. G. Gray, and K. R. Jeffrey, *Chaos in the motion of an inverted pendulum; an undergraduate laboratory experiment*, Amer. J. Phys. **59** (1991), 987–992
55. G. Duffing, *Erzwungen Schwingungen bei veränderlicher Eigenfrequenz und ihre technisch Bedeutung*, Sammlung Vieweg Heft 41/42, Vieweg, Braunschweig, 1918
56. P. L. Felmer, *Multiple periodic solutions for Lagrangian systems in T^n*, Nonlinear Anal. **15** (1990), 815–831
57. P. L. Felmer, *Periodic solutions of spatially periodic Hamiltonian systems*, J. Differential Equations **98** (1992), 143–168

58. P. L. Felmer, *Rotation type solutions for spatially periodic Hamiltonian systems*, Nonlinear Anal. **19** (1992), 409–425
59. A. M. Fink, *Uniqueness theorems and almost periodic solutions to second order differential equations*, J. Differential Equations **4** (1968), 543–548
60. A. M. Fink, *Almost Periodic Differential Equations*, Lecture Notes in Math. No. 377, Springer, Berlin, 1974
61. A. Floer and E. Zehnder, *Fixed point results for symplectic maps related to the Arnold conjecture*, in Dynamical Systems and Bifurcations, Lect. Notes in Math. vol. 1125, Springer, Berlin, 1985, 47–63
62. A. Fonda, and J. Mawhin, *Multiple periodic solutions of conservative systems with periodic nonlinearity*, in Differential Equations and Applications, vol. 1, Ohio University Press, 1989, 298–304
63. A. Fonda, and J. Mawhin, *Critical point theory and multiple periodic solutions of conservative systems with periodic nonlinearity*, in The problem of Plateau : a tribute to J. Douglas and T. Rado, T. M. Rassias ed., World Scientific, Singapore, 1992, 111–128
64. A. Fonda, and M. Willem, *Subharmonic oscillations of forced pendulum-type equations*, J. Differential Equations **81** (1989), 215–220
65. A. Fonda, and F. Zanolin, *Periodic oscillations of forced pendulums with very small length*, Proc. Royal Soc. Edinburgh **127A** (1997), 67–76
66. G. Fournier, R. Iannacci and J. Mawhin, *Periodic solutions of pendulum-like third order differential equations*, in Nonlinear functional analysis and fixed point theory, Maratea 1985, Singh ed., Reidel, Dordrecht, 1986, 235–239
67. G. Fournier, D. Lupo, M. Ramos, and M. Willem, *Limit relative category and critical point theory*, Dynamics Reported **3** (1994), 1–24
68. G. Fournier, and J. Mawhin, *On periodic solutions of forced pendulum-like equations*, J. Differential Equations **60** (1985), 381–395
69. G. Fournier, A. Szulkin, and M. Willem, *Semilinear elliptic equations in R^N with almost periodic or unbounded forcing term*, SIAM J. Math. Anal. **27** (1996), 1653–1660
70. G. Fournier and M. Willem, *Multiple solutions of the forced double pendulum equation*, in Ann. Inst. H. Poincaré, Analyse non linéaire, 6, supplém. (1989), 259–281
71. J. Franks, *Generalizations of the Poincaré-Birkhoff theorem*, Ann. of Math. **128** (1988), 139–151
72. S. Fučik, *Solvability of Nonlinear Equations and Boundary value Problems*, Reidel, 1980
73. M. Furi and M. P. Pera, *On the existence of forced oscillations for the spherical pendulum*, Boll. Un. Mat. It. **4-B** (1990), 381–390
74. M. Furi and M. P. Pera, *A continuation principle for the forced spherical pendulum*, in Fixed Point Theory and Applications, Théra and Baillon ed., Longman, London, 1991, 141–154
75. M. Furi and M. P. Pera, *The forced spherical pendulum does have forced oscillations*, Lect. Notes in Math. vol. 1475, Springer, Berlin, 1991, 176–183
76. M. Furi and M. P. Pera, *On the notion of winding number for closed curves and applications to forced oscillations on even dimensional spheres*, Bol. Un. Math. Ital (7) **7-A** (1993), 397–407
77. M. Furi and M. Spadini, *Multiplicity of forced oscillations for the spherical pendulum*, preprint
78. R. E. Gaines and J. Mawhin, *Coincidence Degree and Nonlinear Differential Equations*, Lecture Notes in Math. No. 568, Springer, Berlin, 1977

79. N. Ghoussoub, *Duality and Perturbation Methods in Critical Point Theory*, Cambridge Univ. Press, Cambridge, 1993
80. I. Goldrish, Y. Imry, G. Wassenman, and E. Ben-Jacob, *Studies of the intermittent-type chaso in Ac- and Dc-driven Josephson junction*, Phys. Rev. B **29** (1984), 1218–1231
81. E. G. Gwinn, and R. M. Westervelt, *Intermittent chaos and low-frequency noise in the driven damped pendulum*, Phys. Rev. Lett. **54** (1985), 1613–1616
82. E. G. Gwinn, and R. M. Westervelt, *Fractal basin boundaries and intermittency in the driven damped pendulum*, Phys. Rev. A **33** (1986), 4143–4155
83. P. Habets and P. Torres, *Some multiplicity results for periodic solutions of a Rayleigh differential equation*, preprint
84. D. Hai, *Note on a differential equation describing the periodic motion of a satellite in its elliptical orbit*, Nonlinear Anal. **12** (1988), 1337–1338
85. A. Halanay, G. A. Leonov, and V. Rasvan, *From pendulum equation to an extended analysis of synchronous machines*, Rend. Sem. Mat. Univ. Polit. Torino **45** (1987), 91–106
86. G. Hamel, *Ueber erzwungene Schingungen bei endlischen Amplituden*, Math. Ann. **86** (1922), 1–13
87. D. R. He, W. J. Yeh, and Y. H. Kao, *Transition from quasiperiodicity to chaos in a Josephson-junction analog*, Phys. Rev. B **30** (1984), 197
88. K. Hocket, and P. Holmes, *Josephson's junction, annulus maps, Birkhoff attractors, horseshoes and rotation sets*, Ergodic Theory and Dynamical Systems **6** (1986), 205–239
89. B. A. Huberman, J. D. Crutchfield, and N. H. Packard, *Noise phenomena in Josephson junctions*, Appl. Phys. Letters **37** (1980), 750–772
90. R. Kannan and R. Ortega, *Periodic solutions of pendulum-type equations*, J. Differential Equations **59** (1985), 123–144
91. R. Kannan and R. Ortega, *An asymptotic result in forced oscillations of pendulum-type equations*, Applicable Analysis **22** (1986), 45–54
92. A. Katok and B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems*, Cambridge Univ. Press, Cambridge, 1995
93. G. Katriel, *Many periodic solutions for pendulum-type equations*, preprint
94. R. L. Kautz, *Chaotic states of Rf-biased Josephson junction*, J. Appl. Phys. **52** (1981), 6241
95. R. L. Kautz, *Chaos in a computer-animated pendulum*, Amer. J. Phys. **61** (1993), 407–415
96. R. L. Kautz, *Chaos in Josephson circuits*, IEEE Trans. Magn. **19** (1983), 465–474
97. R. L. Kautz, and R. Monaco, *Survey of chaos in the rf-biased Josephson junction*, J. Appl. Phys. **57** (1985), 875–889
98. O. Kavian, *Introduction à la théorie des points critiques*, Springer, Paris, 1993
99. I. D. Kill, *On periodic solutions of a certain nonlinear equation*, J. Appl. Math. Mech. **27** (1963), 1699–1704
100. M. Levi, *Beating modes in the Josephson junction*, in Chaos in Nonlinear Dynamical systems, Chandra ed., SIAM, 1984, 56–73
101. M. Levi, *Non-chaotic behavior in the Josephson junction*, Phys. Rev. A **27** (1988), 927–931
102. M. Levi, *Invariant foliations in forced oscillations*, in Geometry and analysis in nonlinear dynamics, Broer-Takens ed., Longman, London, 1992, 34–40
103. M. Levi, *KAM theory for particles in periodic potential*, Ergodic Theory and Dynamical Systems **10** (1990), 777–785

104. M. Levi, F. Hoppensteadt, and W. Miranker, *Dynamics of the Josephson junction*, Quart. Appl. Math. **36** (1978–79), 167–198
105. J. C. Lillo, and G. Seifert, *On stability questions for pendulum type equations*, Z. Angew. Math. Phys. **7** (1956), 238–247
106. Y. Long, *Periodic points of Hamiltonian diffeomorphisms on tori and a conjecture of C. Conley*, Nankai Inst. Math. Preprint Series No. 1996–M–004, 1996
107. D. Lupo, and S. Solimini, *A note on a resonance problem*, Proc. Royal Soc. Edinburgh **102** A (1986), 1–7
108. P. Martinez-Amores, J. Mawhin, R. Ortega, and M. Willem, *Generic results for the existence of nondegenerate periodic solutions of some differential systems with periodic nonlinearities*, J. Differential Equations **91** (1991), 138–148
109. J. N. Mather, *Existence of quasi-periodic orbits for twist homeomorphisms of the annulus*, Topology **21** (1982), 457–467
110. J. Mawhin, *Periodic oscillations of forced pendulum-like equations*, Lecture Notes in Math. No. 964, Springer, Berlin, 1982, 458–476
111. J. Mawhin, *Oscillations forcées du pendule*, in Nonlinear Partial Differential Equations and their Applications, Collège de France Sem. vol. 6, Pitman, Boston, 1985, 275–287
112. J. Mawhin, *Points fixes, points critiques et problèmes aux limites*, Presses Univ. Montréal, 1985
113. J. Mawhin, *Forced oscillations of pendulum-like equations*, in Xth Intern. Conf. Nonlinear Oscillations, Varna 1984, Brankov ed., Bulgarian Acad. Sci., Sofia, 1985, 192–194
114. J. Mawhin, *Recent results on periodic solutions of the forced pendulum equation*, Rend. Ist. Mat. Univ. Trieste **19** (1987), 119–129
115. J. Mawhin, *Problèmes de Dirichlet non linéaires*, Presses Univ. Montréal, 1987
116. J. Mawhin, *On a differential equation for the periodic motions of a satellite around its center of mass*, in Asymptotic methods of mathematical physics, Nauka Dumka, Kiev, 1988, 150–157
117. J. Mawhin, *The forced pendulum: a paradigm for nonlinear analysis and dynamical systems*, Expos. Math. **6** (1988), 271–287
118. J. Mawhin, *Les oscillations forcées du pendule: un paradigme en dynamique et en analyse non linéaire*, Bull. Cl. Sci. Acad. R. Belgique (5) **75** (1989), 58–68
119. J. Mawhin, *Forced second order conservative systems with periodic nonlinearity*, Ann. Inst. H. Poincaré, special issue dedicated to J. J. Moreau, **6** (1989) suppl., 415–434
120. J. Mawhin, *Generic properties of nonlinear boundary value problems*, in Proc. Intern. Confer. Diff. Equ. Math. Phys., Birmingham, 1990, Academic Press, 1992, 217–234
121. J. Mawhin, *Nonlinear oscillations: one hundred years after Liapunov and Poincaré*, Z. Angew. Math. Mech. **73** (1993), T 54–T 62
122. J. Mawhin, *Bounded solutions of nonlinear ordinary differential equations*, in Nonlinear Analysis and Boundary Value Problems for Ordinary Differential Equations, CISM Courses and Lectures No. 371, Udine 1995, Zanolin ed., Springer, Wien, 1996, 121–147
123. J. Mawhin, *Remarques sur les solutions bornées ou presque périodiques de l'équation du pendule forcé* (to appear)
124. J. Mawhin, and M. Willem, *Multiple solutions of the periodic boundary value problem for some forced pendulum-type equations*, J. Differential Equations **52** (1984), 264–287

125. J. Mawhin and M. Willem, *Variational methods and boundary value problems for vector second order differential equations and applications to the pendulum equation*, in *Nonlinear Anal. and Optimisation,* Bologna, 1982, Vinti ed., Springer, Berlin, 1984, 181–192
126. J. Mawhin and M. Willem, *Critical point theory and Hamiltonian systems*, Springer, New York, 1989
127. Q. Min, S. W. Xian, Z. Jinyan, *Global behavior in the dynamical equation of J-J type*, J. Differential Equations **71** (1988), 315–333
128. E. Mirenghi and A. Salvatore, *Remarks on forced Lagrangian systems with periodic potential*, Le Matematiche **46** (1991), 593–608
129. J. Moser, *Stable and Random Motions in Dynamical Systems*, Annals of Math. Studies 77, Princeton Univ. Press, Princeton, 1973
130. J. Moser, *Minimal solutions of variational problems on the torus*, Ann. Inst. H. Poincaré, Anal. non linéaire **3** (1986), 229–272
131. J. Moser, *Recent developments in the theory of Hamiltonian systems*, SIAM Review **28** (1986), 459–485
132. J. Moser, *Breakdown of stability*, Lecture Notes in Phys. vol. 247, Springer, Berlin, 1986, 492–518
133. J. Moser, *Minimal foliations on a torus*, in Topics in Calculus of Variations, CIME Montecatini 1987, Lecture Notes in Math. vol. 1365, Springer, 1989, 62–99
134. J. Moser, *A stability theory for minimal foliations on a torus*, Ergodic Theory and Dynamical Systems **8*** (1988), 251–288
135. J. Moser, *Quasi-periodic solutions of nonlinear elliptic partial differential equations*, Bol. Soc. Brasil. Mat. **20** (1989), 29–45
136. F. Nakajima, *Some conservative pendulum equations with forcing term*, preprint
137. D. Offin, *Subharmonic oscillations for forced pendulum type equations*, Differential and Integral Equations **3** (1990), 965–972
138. Z. Opial, *Sur les intégrales bornées de l'équation* $u'' = f(t, u, u')$, Ann. Polon. Math. **4** (1958), 314–324
139. R. Ortega, *Existencia de soluciones periodicas de la ecuacion del pendulo forzado*, in VII Congr. Ecuac. Differ. y Applic. (VII CEDYA), Granada, 1985, 317–322
140. R. Ortega, *A counterexample for the damped pendulum equation*, Bull. Cl. Sci. Acad. R. Belgique (5) **73** (1987), 405–409
141. R. Ortega, *Stability and index of periodic solutions of an equation of Duffing type*, Boll. U.M.I. (7) **3-B** (1989), 533–546
142. R. Ortega, *Topological degree and stability of periodic solutions for certain differential equations*, J. London Math. Soc. (2) **42** (1990) 505–516
143. R. Ortega, *Some applications of the topological degree to stability theory*, in Topological Methods in DEs and Inclusions, Kluwer, 1995, 377–409
144. R. Ortega, *A boundedness result of Landesman-Lazer type*, *Differential and Integral Equations* **8** (1995), 729–734
145. R. Ortega, *The number of stable periodic solutions of time-dependent Hamiltonian systems with one degree of freedom*, Ergodic Theory and Dynamical Systems (to appear)
146. R. Ortega, *A forced pendulum equation with many periodic solutions*, preprint
147. W. V. Petryshyn, and Z. S. Yu, *On the solvability of an equation describing the periodic motions of a satellite in its elliptic orbit*, Nonlinear Anal. **9** (1985), 969–975
148. H. Poincaré, *Sur un théorème de géométrie*, Rend. Circ. Mat. Univ. Palermo **33** (1912), 375–407

149. P. Rabinowitz, *On a class of functionals invariant under a Z_n action*, Trans. Amer. Math. Soc. **310** (1988), 3O3–311
150. R. Z. Sagdeev, P. A. Usikov, and G. M. Zaslavsky, *Nonlinear Physics: from the Pendulum to Turbulence and Chaos*, Harwood Acad. Publ., Chur, 1988
151. L. Sanchez, *Métodos da teoria de pontos criticos*, Univ. Lisboa, Lisboa, 1993
152. J. A. Sanders, *The (driven) Josephson equation: an exercise in asymptotics*, in Asymptotic analysis II, Verhulst ed., Lecture Notes in Math. vol. 985, Springer, 1983
153. G. Sansone, *Existence et stabilité asymptotique uniforme d'une solution périodique de l'équation $u'' + f(u,a)h(u') = g(u) + p(t)$*, in Les vibrations forcées dans les systèmes nonlinéaires, Coll. Int. CNRS, No 148, Marseille, 1964, 1965, 97–106
154. G. Sansone and R. Conti, *Nonlinear Differential Equations*, Pergamon Press, Oxford, 1964
155. N. Sari, *Oscillations lentement forcées du pendule simple*, Z. Angew. Math. Phys. **41** (1990), 480–500
156. R. Schaaf, and K. Schmitt, *A class of nonlinear Sturm-Liouville problems with infinitely many solutions*, Trans. Amer. Math. Soc. **306** (1988), 853–859
157. R. Schaaf, and K. Schmitt, *Periodic perturbations of linear problems at resonance on convex domains*, Rocky Mountain J. Math. **20** (1990), 1119–1131
158. R. Schaaf, and K. Schmitt, *On the number of solutions of semilinear elliptic problems at resonance: some numerical experiments*, Lecture in Appl. Math. vol. 26, Amer. Math. Soc., Providence, 1990, 541–559
159. R. Schaaf, and K. Schmitt, *Asymptotic behavior of positive solution branches of elliptic problems with linear part at resonance*, Z. Angew. Math. Phys. **43** (1992), 645–676
160. J. Scheurle, *Chaotic solutions of systems with almost periodic forcing*, Z. Angew. Math. Phys. **37** (1986), 12–26
161. B. V. Schmitt, and N. Sari, *Solutions périodiques paires et harmoniques-impaires de l'équation du pendule forcé*, J. Méc. Théor. Appl. **3** (1984), 979–993
162. B. V. Schmitt, and N. Sari, *Sur la structure de l'équation du pendule forcé*, J. Méc. Théor. Appl. **4** (1985), 615–628
163. G. Seifert, *On conditions for stability of solutions of pendulum type equations*, Z. Angew. Math. Phys. **6** (1955), 239–243
164. G. Seifert, *On stability in the large for periodic solutions of differential systems*, Ann. Math. **67** (1958), 83–89
165. G. Seifert, *Almost periodic solutions for systems of differential equations near points of nonlinear first approximation*, Proc. Amer. Math. Soc. **11** (1960), 429–435
166. E. Serra, and M. Tarallo, *A reduction method for periodic solutions of second order subquadratic equations*, Advances in Differential Equations (to appear)
167. E. Serra, M. Tarallo, and S. Terracini, *On the structure of the solution set of forced pendulum-type equations*, J. Differential Equations **131** (1996), 189–208
168. W. X. Shen, *Horseshoe motions and subharmonics of equation of J-J type with small parameters*, Acta math. Sinica **4** (1988), 345–354
169. S. Solimini, *On the solvability of some elliptic partial differential equations with the linear part at resonance*, J. Math. Anal. Appl. **117** (1986), 138–152
170. D. B. Sullivan, and J. E. Zimmerman, *Mechanical analogs of time dependent Josephson phenomena*, Amer. J. Physics **39** (1971), 1504–1517
171. A. Szulkin, *A relative category and applications to critical point theory for strongly indefinite functionals*, Nonlinear Anal. **15** (1990), 725–739

172. G. Tarantello, *On the number of solutions for the forced pendulum equations*, J. Differential Equations **80** (1989), 79–93
173. G. Tarantello, *Remarks on forced equations of the double pendulum type*, Trans. Amer. Math. Soc. **326** (1991), 441–452
174. G. Tarantello, *Multiple forced oscillations for the N-pendulum equation*, Comm. Math. Phys. **132** (1990), 499–517
175. F. Tricomi, *Sur une équation différentielle de l'électrotechnique*, C. R. Acad. Sci. Paris **193** (1931), 635–636
176. F. Tricomi, *Integrazione di un equazione differenziale presentatasi in elettrotecnica*, Ann. R. Scuola Norm. Sup. Pisa **2** (1933), 1–20
177. Q. Wang, Z. Q. Wang, and J. Y. Shi, *Subharmonic oscillations with prescribed minimal period for a class of Hamiltonian systems*, Nonlinear Anal. **28** (1996), 1273–1282
178. J. R. Ward, Jr., *A boundary value problem with a periodic nonlinearity*, Nonlinear Anal. **10** (1986), 207–213
179. M. Willem, *Oscillations forcées de l'équation du pendule*, Pub. IRMA Lille **3** (1981), V–1–V–3
180. M. Willem, *Oscillations forcées de systèmes hamiltoniens*, Publ. Math. Univ. Besançon, 1981
181. M. Willem, *Aspects of Morse theory*, Rend. Ist. Mat. Univ. Trieste **19** (1987), 155–164
182. M. Willem, *Perturbations of non degenerate periodic orbits of Hamiltonian systems*, in Periodic solutions of Hamiltonian systems and related topics, Rabinowitz et al ed., Reidel, Dordrecht, 1987, 261–266
183. M. Willem, *Periodic solutions of differential equations with periodic nonlinearities*, in Variational Methods in Hamiltonian systems and elliptic equations, Girardi et al ed., Longman, Harlow, 1992, 169–179
184. M. Willem, *Minimax Theorems*, Birkhäuser, Boston, 1996
185. J. You, *Invariant tori and Lagrange stability of pendulum-type equations*, J. Differential Equations **85** (1990), 54–65
186. J. You, *Periodic solutions to pendulum type equations*, Nanjing Daxie Xuebao Shuxue Bannian Kan **7** (1990), 218–222
187. J. You, *Boundedness of solutions and existence of quasiperiodic solutions for a non-conservative pendulum-type equation*, Chinese Science Bulletin **21** (1991), 1606–1609
188. J. You, *Boundedness for solutions of non-conservative pendulum-type equations*, preprint
189. E. Zehnder, *Fixed points of symplectic maps and a classical variational principle for forced oscillations*, in Perspectives in Math., Birkhäuser, Basel, 1984, 573–587
190. E. Zehnder, *The Arnold conjecture for fixed points of symplectic mappings and periodic solutions of Hamiltonian systems*, in Proc. Intern. Congress of Mathematicians, Berkeley, 1986, Amer. Math. Soc., Providence, 1987, 1237-1246

EQUADIFF 9 CD ROM, Brno 1997
Masaryk University

PROCEEDINGS
pp. 147–158

Multiple Solutions of Nonlinear Boundary Value Problems and Topological Degree

Irena Rachůnková

Department of Mathematical Analysis,
Faculty of Science, Palacký University
Tomkova 40, 77900 Olomouc, Czech Republic
Email: RACHUNKO@matnw.upol.cz

Abstract. This paper deals with the second order nonlinear boundary value problems. We consider the two-point, multipoint or nonlinear boundary conditions on a compact interval and suppose the existence of strict upper and lower solutions of the problem with the both types of ordering i.e. the lower (upper) solution is less than the upper (lower) one. We prove the relation between the topological degree and strict upper and lower solutions in the both cases and using this we get the existence and multiplicity results for the boundary value problems under consideration.

AMS Subject Classification. 34B15, 34B10

Keywords. Nonlinear second order ODE, two-point, multipoint and nonlinear boundary conditions, strict upper and lower solutions, topological degree, existence of more solutions

1 Introduction

When we study the boundary value problems for the second order differential equation

$$x'' = f(t, x, x'), \tag{1.1}$$

with certain linear or nonlinear boundary conditions on the compact interval $J = [a, b] \subset \mathbf{R}$ we often use the properties of lower and upper solutions for (1.1). Let us remind the definition.

Let f be continuous on $J \times \mathbf{R}^2$ (or let f satisfy the Carathéodory conditions on $J \times \mathbf{R}^2$). The functions $\sigma_1, \sigma_2 \in C^2(J)$ (or $AC^1(J)$) are called lower and upper solutions for (1.1), if they satisfy

$$\begin{aligned} \sigma_1''(t) &\geq f(t, \sigma_1(t), \sigma_1'(t)), \\ \sigma_2''(t) &\leq f(t, \sigma_2(t), \sigma_2'(t)), \end{aligned} \tag{1.2}$$

for all $t \in J$ (for a.e. $t \in J$). If the inequalities in (1.2) are strict, then σ_1, σ_2 are called strict lower and upper solutions.

We distinguish two basic cases:

1. The functions σ_1, σ_2 are well ordered, i.e.

$$\sigma_1(t) \le \sigma_2(t) \text{ for all } t \in J. \tag{1.3}$$

2. The functions σ_1, σ_2 are not well ordered, i.e. the condition (1.3) falls.

The most existence results concern the first case, but there are the existence results for the second case, as well. We can refer to the papers [7], [3] or [4].

Here, we want to present the existence and multiplicity results for (1.1) (with various boundary conditions) in the first case and also in the second case where σ_1, σ_2 have the opposite order, i.e.

$$\sigma_2(t) \le \sigma_1(t) \text{ for all } t \in J. \tag{1.4}$$

Our results are based on the relation between the topological degree of the operator corresponding to the boundary value problem and strict lower and upper solutions fulfilling (1.3) or (1.4) (in the strict sense).

For getting the existence and multiplicity results we need a priori estimates of solutions of the original boundary value problem or of solutions of proper auxiliary boundary value problems. Working with σ_1, σ_2, we want to estimate the solutions just by σ_1, σ_2. For the estimation at the endpoints a, b of J we use certain connection between σ_1, σ_2 and the boundary conditions. It is well known that for the classical two-point boundary conditions such connection has the form:

– for the periodic conditions

$$x(a) = x(b), \ x'(a) = x'(b), \tag{1.5}$$

we suppose

$$\sigma_i(a) = \sigma_i(b), \ \big(\sigma_i'(b) - \sigma_i'(a)\big)(-1)^i \ge 0, \ i = 1, 2; \tag{1.6}$$

– for the Neumann conditions

$$x'(a) = 0, \ x'(b) = 0, \tag{1.7}$$

we assume

$$\sigma_i'(a)(-1)^i \le 0, \ \sigma_i'(b)(-1)^i \ge 0, \ i = 1, 2. \tag{1.8}$$

Similarly,

– for the four-point conditions

$$x(a) = x(c), \ x(d) = x(b), \ a < c \le d < b, \tag{1.9}$$

σ_1, σ_2 have to satisfy

$$\begin{aligned} \big(\sigma_i(c) - \sigma_i(a)\big)(-1)^i \le 0, \\ \big(\sigma_i(b) - \sigma_i(d)\big)(-1)^i \ge 0, \ i = 1, 2, \end{aligned} \tag{1.10}$$

– for the nonlinear conditions

$$g_1\big(x(a),x'(a)\big)=0,\ \ g_2\big(x(b),x'(b)\big)=0, \tag{1.11}$$

where $g_1,g_2\in C(\mathbf{R}^2)$ are increasing in the second argument and g_1 is nonincreasing and g_2 nondecreasing in the first argument, we can impose on σ_1,σ_2

$$\begin{aligned} &g_1\big(\sigma_i(a),\sigma_i'(a)\big)(-1)^i\le 0,\\ &g_2\big(\sigma_i(b),\sigma_i'(b)\big)(-1)^i\ge 0,\ \ i=1,2. \end{aligned} \tag{1.12}$$

Let us note that for more general nonlinear two-point boundary conditions the compatibility of the boundary conditions with σ_1,σ_2 was introduced in [14]. For the special cases of the conditions (1.5), (1.7) and (1.11) this notion leads just to the assumptions (1.6), (1.8) and (1.12).

In this paper we will study the boundary value problems (1.1), (k), and we will assume the existence of lower and upper solutions σ_1,σ_2 of (1.1) with the property (k+.1), k∈{1.5,1.7,1.9,1.11}. The problem (1.1), (k), k∈{1.5,1.7,1.9,1.11}, can be written in the form of the operator equation

$$(L+N)\,x=0, \tag{1.13}$$

where $L:\operatorname{dom}L\to Y$ is a linear operator and it is a Fredholm map of index 0, and $N:C^1(J)\to Y$ is, in general, nonlinear and it is L-compact on any open bounded set $\Omega\subset C^1(J)$. The form of L and N and the choice of the spaces $\operatorname{dom}L$ and Y depend on the type of boundary value problems. Let us suppose that f is continuous on $J\times\mathbf{R}^2$. Then we put for k∈{1.5,1.7,1.9} $\operatorname{dom}L=\{x\in C^2(J):x$ satisfies (k)$\}$, $Y=C(J)$, $L:x\longmapsto x''$, $N:x\longmapsto -f(\cdot,x(\cdot),x'(\cdot))$; for the boundary condition (1.11) we put $\operatorname{dom}L=C^2(J)$, $Y=C(J)\times\mathbf{R}^2$, $L:x\longmapsto(x'',0,0)$, $N:x\longmapsto(-f(\cdot,x(\cdot),x'(\cdot)),g_1(x(a),x'(a)),g_2(x(b),x'(b)))$. For more details see [2], [8], [9].

If the equation (1.13) has no solution on the boundary of Ω then there exists the degree of the map $L+N$ in Ω with respect to L

$$d_L(L+N,\Omega).$$

In [6], the relation between the degree and strict lower and upper solutions satisfying (1.3) (in the strict sense) is shown. In the following section we will formulate this relation for the above boundary value problems.

2 Topological degree for f bounded

First, let us suppose that $f\in C(J\times\mathbf{R}^2)$ is bounded:

$$\exists M\in(0,\infty):|f(t,x,y)|<M \text{ for } \forall(t,x,y)\in J\times\mathbf{R}^2. \tag{2.1}$$

For f unbounded we will use the method of a priori estimates and replace the condition (2.1) by the conditions of the growth or sign types in the next sections.

Theorem 1. *Suppose* k∈{1.5,1.7,1.9,1.11}. *Let* (2.1) *be fulfilled,* (1.13) *be the operator equation corresponding to the problem*(1.1), (k) *and let* σ_1, σ_2 *be strict lower and upper solutions of* (1.1), (k) *with*

$$\sigma_1(t) < \sigma_2(t) \text{ for all } t \in J.$$

Then

$$d_L(L+N, \Omega_1) = 1 \pmod 2, \tag{2.2}$$

with

$$\begin{aligned}\Omega_1 = &\{x \in C^1(J) : \sigma_1(t) < x(t) < \sigma_2(t), |x'(t)| < c \text{ for all } t \in J\},\\ &\text{where } c \geq (2M+r+1)(b-a) \text{ for k} \in \{1.5, 1.7, 1.9\}\\ &\text{and } c \geq (2M+r+1)(b-a) + 2(r+1)/(b-a) \text{ for k=1.11},\\ &r = \| \sigma_1 \|_{\max} + \| \sigma_2 \|_{\max}.\end{aligned}$$

Theorem 1 concerns the case of well ordered σ_1, σ_2. the case where σ_1, σ_2 are ordered by the opposite way is described in Theorem 2.

Theorem 2. *Suppose* k∈{1.5,1.7,1.9,1.11}. *Let* (2.1) *be fulfilled,* (1.13) *be the operator equation corresponding to the problem* (1.1), (k) *and let* σ_1, σ_2 *be strict lower and upper solutions of* (1.1), (k) *satisfying*

$$\sigma_2(t) < \sigma_1(t) \text{ for all } t \in J.$$

Then

$$d_L(L+N, \Omega_2) = 1 \pmod 2, \tag{2.3}$$

where

$$\begin{aligned}\Omega_2 = \{x \in C^1(J) : \|x\|_{\max} < A,\ \|x'\|_{\max} < B,\\ \exists t_x \in J : \sigma_2(t_x) < x(t_x) < \sigma_1(t_x)\},\end{aligned}$$

with $B \geq 2(b-a)M$, $A \geq \|\sigma_1\|_{\max} + \|\sigma_2\|_{\max} + 2(b-a)^2 M$ *for* k ∈{1.5,1.7,1.9}, $B \geq 2(b-a)M + \|\sigma_2'\|_{\max}$, $A \geq \|\sigma_1\|_{\max} + \|\sigma_2\|_{\max} + (b-a)B$ *for* k=1.11.

Corollary 3. *Suppose* k∈{1.5,1.7,1.9,1.11}. *If* σ_1, σ_2 *in Theorem* 1 (2) *are not strict, then either the problem* (1.1), (k) *has a solution on* $\partial\Omega_1$ *(*$\partial\Omega_2$*) or the condition* (2.2) ((2.3)) *is valid.*

3 Existence and multiplicity for *f* bounded

As the direct consequence of Corollary 3, using a limiting process, we obtain the following existence results for the problems (1.1), (k), k∈{1.5,1.7,1.9,1.11}.

Theorem 4. *Suppose* k∈{1.5,1.7,1.9,1.11}. *Let* (2.1) *be fulfilled and let* σ_1, σ_2 *be lower and upper solutions of* (1.1), (k) *with*

$$\sigma_1(t) \le \sigma_2(t) \text{ for all } t \in J.$$

Then the problem (1.1), (k) *has at least one solution in* $\overline{\Omega}_1$, *where* Ω_1 *is the set from Theorem* 1.

Remark 5. The assumption about the monotonicity of g_1, g_2 can be omitted in Theorem 1 and 4. The existence results of Theorem 4 are known and they are presented here for the completeness, only.

Theorem 6. *Suppose* k∈{1.5,1.7,1.9,1.11}. *Let* (2.1) *be fulfilled and let* σ_1, σ_2 *be lower and upper solutions of* (1.1), (k) *with*

$$\sigma_2(t) \le \sigma_1(t) \text{ for all } t \in J.$$

Then the problem (1.1), (k) *has at least one solution in* $\overline{\Omega}_2$, *where* Ω_2 *is the set from Theorem* 2.

Remark 7. For k∈{1.5,1.7} the similar existence results are proven in [3], [4], [7].

Theorems 1 and 2 are a tool for proving multiplicity results for (1.1), (k), both for the linear two-point (k∈{1.5,1.7}) or multipoint boundary conditions (k=1.9) and for the nonlinear boundary condition (k=1.11).

Theorem 8. *Suppose* k∈{1.5,1.7,1.9,1.11}. *Let* (2.1) *be fulfilled and let* σ_1, σ_2, σ_3 *be strict lower, upper and lower solutions of* (1.1), (k) *with*

$$\sigma_1(t) < \sigma_2(t) < \sigma_3(t) \text{ for all } t \in J. \tag{3.1}$$

Then (1.1), (k) *has at least two different solutions* u, v *satisfying*

$$\sigma_1(t) < u(t) < \sigma_2(t), \sigma_1(t) < v(t) \text{ for all } t \in J,$$
$$\sigma_2(t_v) < v(t_v) < \sigma_3(t_v) \text{ for a } t_v \in J.$$

The dual situation is described in Theorem 9.

Theorem 9. *Let all assumptions of Theorem* 8 *be fulfilled with the exception that now* $\sigma_1, \sigma_2, \sigma_3$ *are strict lower, upper and upper solutions with*

$$\sigma_3(t) < \sigma_1(t) < \sigma_2(t) \text{ for all } t \in J \tag{3.2}$$

Then (1.1), (k) *has at least two different solutions* u, v *satisfying*

$$\sigma_1(t) < u(t) < \sigma_2(t), v(t) < \sigma_2(t) \text{ for all } t \in J,$$
$$\sigma_3(t_v) < v(t_v) < \sigma_1(t_v) \text{ for a } t_v \in J.$$

For constant lower and upper solutions we get the multiplicity result of the Ambrosetti-Prodi type.

Theorem 10. *Suppose* k∈{1.5,1.7,1.9}. *Let* (2.1) *be fulfilled and let* $n \in \mathbf{N}$, $n \geq 2$, $s_1, r_1, \ldots, r_{n+1} \in \mathbf{R}$ *be such that*

$$r_1 < r_2 < \cdots < r_{n+1} \tag{3.3}$$

and

$$\big(f(t, r_i, 0) - s_1\big)(-1)^i < 0 \text{ for all } t \in J,\ i \in \{1, \ldots, n\}. \tag{3.4}$$

Then there exist $s_2, s_3 \in (-M, s_1)$, $s_3 \leq s_2$, *such that the problem*

$$x'' + f(t, x, x') = s, \ (k) \tag{3.5}$$

has:

(i) *at least* n *different solutions greater than* r_1 *for* $s \in (s_2, s_1]$;
(ii) *at least* $\frac{n+1}{2}$ $(\frac{n}{2})$ *solutions greater than* r_1 *for* $s = s_2$ *and* n *odd (even);*
(iii) *provided* $s_3 < s_2$ *at least one solution greater then* r_1 *for* $s \in [s_3, s_2)$;
(iv) *no solution for* $s < s_3$.

4 Topological degree for f unbounded

In this section we suppose that k∈ {1.5, 1.7, 1.11}, that (1.13) is the operator equation corresponding to the problem (1.1), (k) and that σ_1, σ_2 are strict lower and upper solutions of (1.1), (k).

Using the method of a priori estimates we can replace the condition (2.1) in Theorem 1 by the Nagumo-Knobloch-Schmitt condition with bounding functions φ_1, φ_2 :

$$\begin{gathered}
\exists \varphi_1, \varphi_2 \in C^1(K) : \varphi_1(t, \sigma_i(t)) \leq \sigma_i'(t), \varphi_2(t, \sigma_i(t)) \geq \sigma_i'(t), \\
f(t, x, \varphi_1(t, x) < \tfrac{\partial \varphi_1(t,x)}{\partial t} + \tfrac{\partial \varphi_1(t,x)}{\partial x}\, \varphi_1(t, x), \\
f(t, x, \varphi_2(t, x) > \tfrac{\partial \varphi_2(t,x)}{\partial t} + \tfrac{\partial \varphi_2(t,x)}{\partial x}\, \varphi_2(t, x), \\
\text{for } i \in \{1, 2\} \text{ and for all } (t, x) \in K = J \times [\sigma_1(t), \sigma_2(t)].
\end{gathered} \tag{4.1}$$

Theorem 11. *Let* (4.1) *be fulfilled and let*

$$\sigma_1(t) < \sigma_2(t) \text{ for all } t \in J.$$

Further suppose that for k=1.5

$$(\varphi_i(b, x) - \varphi_i(a, x))(-1)^i \geq 0,$$

for k=1.7

$$(\varphi_i(b, x) - \sigma_i'(b))(-1)^i > 0,$$

and for k=1.11

$$g_2(x, \varphi_i(b, x))(-1)^i > 0,$$

with $i = 1, 2$, $x \in [\sigma_1(t), \sigma_2(t)]$.
Then

$$d_L(L + N, \Omega_3) = 1 \pmod 2,$$

where

$$\Omega_3 = \{x \in C^1(J) : \sigma_1(t) < x(t) < \sigma_2(t),\ \varphi_1(t, x) < x'(t) < \varphi_2(t, x) \text{ on } K\}.$$

For the constant functions $\sigma_1, \sigma_2, \varphi_1, \varphi_2$ Theorem 11 implies

Corollary 12. *Suppose that there exist real numbers* $r_1 < r_2$, $c_1 < 0 < c_2$, *such that*

$$f(t, r_1, 0) < 0,\ f(t, r_2, 0) > 0, \tag{4.2}$$
$$f(t, x, c_1) < 0,\ f(t, x, c_2) > 0, \tag{4.3}$$

for all $(t, x) \in J \times [r_1, r_2]$.
If k=1.11 *we suppose moreover that for* $x \in [r_1, r_2]$

$$\begin{aligned} &g_1(r_1, 0) \geq 0,\ g_1(r_2, 0) \leq 0, \\ &g_2(r_1, 0) \leq 0,\ g_2(r_2, 0) \geq 0, \end{aligned} \tag{4.4}$$
$$g_2(x, c_i)(-1)^i > 0,\ i = 1, 2. \tag{4.5}$$

Then

$$d_L(L + N, \Omega_4) = 1 \pmod 2,$$

where

$$\Omega_4 = \{x \in C^1(J) : r_1 < x(t) < r_2,\ c_1 < x'(t) < c_2, \forall\, t \in J.\}$$

Now, let us consider the special case of bounding functions depending on t only:

$$\begin{aligned} &\exists \beta_1, \beta_2 \in C^1(J)\colon\ \beta_1(t) \leq \sigma_i'(t),\ \beta_2(t) \geq \sigma_i'(t), \\ &f(t, x, \beta_1(t)) < \beta_1'(t),\ f(t, x, \beta_2(t)) > \beta_2'(t), \end{aligned} \tag{4.6}$$

for all $(t, x) \in J \times [s_2, s_1]$, where $s_2 = \min\{\sigma_2(t) : t \in J\} - \int_a^b \gamma(t)dt$, $s_1 = \max\{\sigma_1(t) : t \in J\} + \int_a^b \gamma(t)dt$, $\gamma(t) = \max\{|\beta_1(t)|, |\beta_2(t)|\}$.

Theorem 13. *Let* (4.6) *be fulfilled and let*

$$\sigma_2(t) < \sigma_1(t) \text{ for all } t \in J.$$

Further suppose that for k=1.5

$$(\beta_i(b) - \beta_i(a))(-1)^i \geq 0, \tag{4.7}$$

for k=1.7

$$(\beta_i(b) - \sigma_i'(b))(-1)^i > 0, \tag{4.8}$$

and for k=1.11

$$g_2(x, \beta_i(b))(-1)^i > 0, \tag{4.9}$$

with $i \in \{1, 2\}$, $x \in [s_2, s_1]$.
Then

$$d_L(L + N, \Omega_5) = 1 \pmod 2,$$

where

$$\Omega_5 = \{x \in C^1(J) : s_2 < x(t) < s_1,\ \beta_1(t) < x'(t) < \beta_2(t) \text{ for all } t \in J, \\ \exists t_x \in J : \sigma_2(t_x) < x(t_x) < \sigma_1(t_x)\}.$$

Corollary 14. *Suppose that there exist real numbers* $r_1 > r_2$, $c_1 < 0 < c_2$, *such that* (4.2) *and* (4.3) *are satisfied for all* $(t, x) \in J \times [r_2 + c_1(b-a), r_1 + c_2(b-a)]$. *If* k=1.11, *we suppose that* (4.4), (4.5) *are satisfied for* $x \in [r_2 + c_1(b-a), r_1 + c_2(b-a)]$.
Then

$$d_L(L + N, \Omega_6) = 1 \pmod 2,$$

where

$$\Omega_6 = \{x \in C^1(J) : r_2 + c_1(b-a) < x(t) < r_1 + c_2(b-a), \\ c_1 < x'(t) < c_2, \forall\, t \in J.\}$$

Example 15. Suppose $f_1, f_2, f_3 \in C(J), k, m \in \mathbf{N}$. The function

$$f(t, x, y) = f_1(t)x^{2k+1} + f_2(t)y^{2m+1} + f_3(t)$$

satisfies the conditions of Corollary 12, if $f_1, f_2 > 0$ on J, and it satisfies the conditions of Corollary 14, if $f_1 < 0$, $f_2 > 0$ on J and either $m > k$ or $m = k$, $f_2(t) > \|f_1\|_{\max}(b-a)^{2k+1}$ for all $t \in J$.

Other type of conditions which can be used instead of (2.1) in Theorem 1 and Theorem 2 are one-sided growth conditions which were used by Kiguradze [5] in some existence theorems.

1. The one-sided Bernstein-Nagumo condition:

$$\begin{gathered}\exists\,\omega\in C(\mathbf{R}_+),\ \omega \text{ positive},\ \int_0^\infty \frac{ds}{\omega(s)}=\infty \text{ and}\\ f(t,x,y)\le \omega(|y|)\cdot(1+|y|)\\ \forall(t,x)\in J\times[\sigma_1(t),\sigma_2(t)]\times\mathbf{R}.\end{gathered}\tag{4.10}$$

2. The one-sided linear growth condition:

$$\begin{gathered}\exists a_1,a_2\in(0,\infty),\rho\in C(J\times\mathbf{R}),\ \text{non-negative and non-decreasing}\\ \text{in the second argument such that}\\ f(t,x,y)\le a_1|x|+a_2|y|+\rho(t,|x|+|y|)\\ \forall(t,x,y)\in J\times\mathbf{R}^2,\end{gathered}\tag{4.11}$$

where

$$a_1(b-a)^2+a_2(b-a)<1$$

and

$$\lim_{z\to\infty}\frac{1}{z}\int_a^b\rho(t,z)dt=0.$$

Note 16. Let us remember that if f satisfies (4.11) it satisfies (4.10) as well.

For the proof of the following theorems we need lemmas on a priori estimates for solutions of the problems (1.1), (k), k$\in\{1.5, 1.7, 1.11\}$.

Lemma 17. *Suppose*

$$\sigma_1(t)<\sigma_2(t)\quad \textit{for all } t\in J.$$

Let (4.10) *be satisfied. If* k=1.11, *suppose moreover*

$$\lim_{y\to\infty} g_1(r_2,y)>0,\ \lim_{y\to-\infty} g_2(r_2,y)<0,\tag{4.12}$$

$r_1=\min\{\sigma_1(t):t\in J\}$, $r_2=\max\{\sigma_2(t):t\in J\}$.
Then there exists $\mu^*\in(0,\infty)$ *such that for any solution* u *of the problem* (1.1), (k), *the implication*

$$\sigma_1(t)<u(t)<\sigma_2(t)\ \text{on}\ J\Longrightarrow\|u'\|_{\max}<\mu^*$$

is valid.

Lemma 18. *Let* $r_1, r_2 \in \mathbf{R}, r_1 < r_2$ *and let* (4.11) *be satisfied. If* k=1.11, *suppose moreover*

$$\lim_{y\to\infty} g_1(x,y) > 0, \ \lim_{y\to-\infty} g_2(x,y) < 0, \tag{4.13}$$

uniformly for $x \in \mathbf{R}_+$.
Then there exists $\nu^* \in (0,\infty)$ *such that for any solution* u *of the problem* (1.1), (k), *the implication*

$$\exists t_u \in J : r_1 < u(t_u) < r_2 \Longrightarrow \|u'\|_{\max} < \nu^*$$

is valid.

Theorem 19. *Let* (4.10) *be fulfilled and let*

$$\sigma_1(t) < \sigma_2(t) \text{ for all } t \in J.$$

If k=1.11, *suppose moreover* (4.12).
Then there exists $r^* \in (0,\infty)$ *such that*

$$d_L(L+N, \Omega_6) = 1 \pmod 2,$$

where

$$\Omega_6 = \{x \in C^1(J) : \sigma_1(t) < x(t) < \sigma_2(t) \ \forall t \in J, \ \|x'\|_{\max} < r^*\}.$$

Theorem 20. *Let* (4.11) *be fulfilled and let*

$$\sigma_2(t) < \sigma_1(t) \text{ for all } t \in J.$$

If k=1.11, *suppose moreover* (4.13).
Then there exists $r^* \in (0,\infty)$ *such that*

$$d_L(L+N, \Omega_7) = 1 \pmod 2,$$

where

$$\Omega_7 = \{x \in C^1(J) : \|x\|_{\max} + \|x'\|_{\max} < r^*, \ \exists t_x \in J : \sigma_2(t_x) < x(t_x) < \sigma_1(t_x)\}.$$

5 Multiplicity results for f unbounded

We can extend the results of the Section 3 onto differential equations with an unbounded right-hand side $f \in C(J \times \mathbf{R}^2)$. We will present here such extension of some multiplicity results.

Let us suppose that σ_1, σ_2 and σ_3 are strict lower, upper and lower solutions of (1.1), (k), k$\in$\{1.5,1.7,1.11\}. Using Theorem 11 and Theorem 13 we get the following multiplicity result:

Theorem 21. *Suppose that* (3.1), (4.6) *and, according to* k, *the condition* (4.7) *or* (4.8) *or* (4.9) *are fulfilled for all* $(t,x) \in J\times[\sigma_1(t), s_3]$, *where* $s_3 = \max\{\sigma_3(t) : t \in J\} + \int_a^b \gamma(t)\,dt$.
Then the assertion of Theorem 8 *is valid.*

Similarly, by means of Theorem 19 and Theorem 20 and the fact that (4.11) and (4.13) are the special cases of (4.10) and (4.12), we get:

Theorem 22. *Let us suppose that* (3.1) *and* (4.11) *are fulfilled and, for* k=1.11, *suppose moreover* (4.13). *Then the assertion of Theorem* 8 *is valid.*

Now, let us consider the dual situation, where σ_3 is an upper solution of (1.1), (k).

Theorem 23. *Suppose that* (3.2), (4.6) *and, according to* k, *the condition* (4.7) *or* (4.8) *or* (4.9) *are fulfilled for all* $(t,x) \in J\times[b_3, \sigma_2(t)]$, *where* $b_3 = \min\{\sigma_3(t) : t \in J\} - \int_a^b \gamma(t)\,dt$.
Then the assertion of Theorem 9 *is valid.*

Theorem 24. *Let us suppose that* (3.2) *and* (4.11) *are fulfilled and, for* k=1.11, *suppose moreover* (4.13). *Then the assertion of Theorem* 9 *is valid.*

For constant lower and upper solutions we can generalize the theorems from [11], concerning the multiplicity results of the Ambrosetti-Prodi type for the periodic problem.

Theorem 25. *Suppose* k∈{1.5,1.7}. *Let* $n \in \mathbf{N}$, $n \geq 2$, $c_1, c_2, s_1, r_1, \dots, r_{n+1} \in \mathbf{R}$, $c_1 < 0 < c_2$, *satisfy* (3.3), (3.4), (4.3) *for all* $(t,x) \in J \times [r_1, r^*]$, *where*

$$r^* = \begin{cases} r_{n+1} & \text{for } n \text{ odd} \\ r_{n+1} + \max\{|c_1|, c_2\}(b-a) & \text{for } n \text{ even.} \end{cases} \tag{5.1}$$

Then there exist $s_2, s_3 \in (-\infty, s_1)$, $s_3 \leq s_2$, *such that the problem* (3.5) *has:*

(i) *at least* n *different solutions* u_i, $i = 1, .., n$, *satisfying*

$$r_1 < u_i(t) < r^* \text{ for all } t \in J, i \in \{1, \dots, n\}; \tag{5.2}$$

(ii) *at least* $\frac{n+1}{2}$ $(\frac{n}{2})$ *solutions satisfying* (5.2) *for* $s = s_2$ *and* n *odd (even);*

(iii) *provided* $s_3 < s_2$ *at least one solution satisfying* (5.2) *for* $s \in [s_3, s_2)$;

(iv) *no solution satisfying* (5.2) *for* $s < s_3$.

Theorem 26. *Suppose* k∈{1.5,1.7}. *Let* $n \in \mathbf{N}$, $n > 2$, *be odd and let further* $s_1, r_1, \dots, r_{n+1} \in \mathbf{R}$ *satisfy* (3.3) *and* (3.4). *Further, let* (4.10) *be fulfilled.*
Then there exists $r^* \geq r_{n+1}$ *such that* (i)–(iv) *of Theorem* 25 *are valid.*

Theorem 27. *Suppose* k∈{1.5,1.7}. *Let* $n \in \mathbf{N}$, $n \geq 2$, *be even and let further* $s_1, r_1, \dots, r_{n+1} \in \mathbf{R}$ *satisfy* (3.3) *and* (3.4). *Further let* (4.11) *be fulfilled.*
Then there exists $r^* \geq r_{n+1}$ *such that* (i)–(iv) *of Theorem* 25 *are valid.*

Note 28. Close results concerning the existence of two or three solutions of the periodic problem can be found also in [1] and [13].
For f satisfying the Carathéodory conditions on $J \times \mathbf{R}^2$ the results of Corollary 3, Theorem 4 and Theorem 6 can be proven as well. The multiplicity results of the Theorems 8–10 and the theorems for f unbounded of the Sections 4 and 5 have to be a little modified because in the Carathéodory case solutions can interact strict lower and upper solutions.

References

1. C. Fabry, J. Mawhin and M. N. Nkashama: *A multiplicity result for periodic solutions of forced nonlinear boundary value problem*, Bull. London. Math. Soc. **18** (1986), 173–186.
2. R. E. Gaines and J. L. Mawhin: *Coincidence Degree and Nonlinear Differential Equations*, Lecture Notes in Math. 568, Springer-Verlag, Berlin 1977.
3. J. P. Gossez and P. Omari: *Periodic solutions of a second order ordinary differential equation: a necessary and sufficient condition for nonresonance*, J. Differential Equations **94** (1991), 67–82.
4. P. Habets and P. Omari: *Existence and localization of solutions of second order elliptic problems using lower and upper solutions in the reversed order*, preprint U.C.L., June 1994, in print on Topological Methods in Nonlinear Analysis.
5. I. Kiguradze: *Some Singular Boundary Value Problems for Ordinary Differential Equations*, ITU, Tbilisi 1975 (in Russian).
6. J. Mawhin: *Points fixes, points critiques et problèmes aux limites*, Sémin. Math. Sup., No. 92, Presses Univ. Montréal, Montréal 1985.
7. P. Omari: *Non-ordered lower and upper solutions and solvability of the periodic problem for the Liénard and the Rayleigh equations*, Rend. Inst. Mat. Univ. Trieste **20** (1988), 54–64.
8. I. Rachůnková: *Sign conditions in nonlinear boundary value problems*, Acta Univ. Palack. Olom., Fac. Rer. Nat. 114, **33** (1994), 117–134.
9. I. Rachůnková: *On a transmission problem*, Acta Univ. Palack. Olom., Fac. Rer. Nat. 105, **31** (1992), 45–59.
10. I. Rachůnková: *Upper and lower solution and topological degree*. Preprint 23/1997.
11. I. Rachůnková: *On the existence of two solutions of the periodic problem for the ordinary second-order differential equation*, Nonlinear Analysis TMA **22** (1994), 1315–1322.
12. I. Rachůnková: *Upper and lower solution and multiplicity results*. Preprint 25/1997.
13. B. Rudolf: *A multiplicity result for a periodic boundary value problem*, Nonlinear Analysis TMA **28** (1997),137–144.
14. H. B. Thompson: *Second order ordinary differential equations with fully nonlinear two-point boundary conditions*, Pacif. Journal Math. **172** (1996), 255–297.

EQUADIFF 9 CD ROM, Brno 1997
Masaryk University

PROCEEDINGS
pp. 159–168

Generalized Linking Theorem and Nonlinear Equations in Unbounded Domains

Andrzej Szulkin

Department of Mathematics, Stockholm University,
106 91 Stockholm, Sweden
Email: andrzejs@matematik.su.se

Abstract. We consider the following three problems: existence of nontrivial solutions for a semilinear Schrödinger equation in $\mathbb{R}^N$, existence of homoclinics for a first order Hamiltonian system and existence of time–periodic motions in an infinite chain of particles. The common feature of these problems is that the associated Euler–Lagrange functional has the so–called linking geometry and the Palais–Smale condition is not satisfied.

AMS Subject Classification. 34C15, 34C37, 35J65, 58E05

Keywords. Generalized linking theorem, Schrödinger equation, Hamiltonian system, chain of particles

1 Introduction

The purpose of this paper is to survey some recent results in the theory of nonlinear differential equations in unbounded domains. Solutions of these equations will be found as critical points of an associated Euler-Lagrange functional Φ in a suitable Hilbert space. We consider three different problems whose common feature is that Φ has the so-called linking geometry and the Palais-Smale condition is not satisfied. To be more specific, in Section 2 we will be concerned with the problem of existence of a nontrivial solution of the Schrödinger equation $-\Delta u + V(x)u = f(x,u)$ in $\mathbb{R}^N$ in a situation when 0 is in a gap of the spectrum of the operator $-\Delta + V$ and the nonlinearity f is superlinear at $u = 0$ and $|u| = \infty$. In Section 3 the existence of homoclinic solutions of a Hamiltonian system in $\mathbb{R}^{2N}$ is considered, and in Section 4 we turn our attention to the problem of existence of time-periodic solutions for an infinite chain of particles with nearest neighbour interaction (the so-called Fermi-Pasta-Ulam model). The arguments presented here are very sketchy. Complete proofs may be found in the original work to which the reader is referred.

Our starting point is the following generalized linking theorem:

Theorem 1. *Let E be a separable real Hilbert space and suppose that $\Phi \in C^1(E, \mathbb{R})$ satisfies the following hypotheses:*
(i) $\Phi(u) = \frac{1}{2}\langle Lu, u\rangle - \psi(u)$, where L is a bounded selfadjoint linear operator, ψ is bounded below, weakly sequentially lower semicontinuous and $\nabla\psi$ is weakly sequentially continuous.
(ii) $E = Y \oplus Z$, where Y, Z are L-invariant and the quadratic form $\langle Lu, u\rangle$ is negative definite on Y and positive definite on Z.
(iii) There are constants $b, \rho > 0$ such that $\Phi|_{\partial B_\rho \cap Z} \geq b$, where $B_\rho := \{u \in E : \|u\| < \rho\}$.
(iv) There is $z_0 \in Z$, $\|z_0\| = 1$, and $R > \rho$ such that $\Phi|_{\partial M} \leq 0$, where $M := \{u = y + \lambda z_0 : y \in Y,\ \|u\| \leq R,\ \lambda \geq 0\}$.
Then there exists a sequence (u_n) such that $\nabla\Phi(u_n) \to 0$ and $\Phi(u_n) \to c$ for some $c \in [b, \sup_M \Phi]$.

The above result extends linking theorems of Rabinowitz [21,22] and Benci-Rabinowitz [9,22] (in the first of them Y is assumed to be finite-dimensional, in the second $\nabla\psi$ is compact). Theorem 1 may be found in [19], see also [34]. We would like to emphasize that in the problems considered in the next sections both Y and Z are infinite-dimensional and $\nabla\psi$ is not compact.

2 Schrödinger equation

Consider the semilinear Schrödinger equation

$$\begin{cases} -\Delta u + V(x)u = f(x,u), & x \in \mathbb{R}^N \\ u(x) \to 0 \quad \text{as} \quad |x| \to \infty, \end{cases} \tag{2.1}$$

where V and f are continuous and 1-periodic with respect to x_j, $1 \leq j \leq N$. It is known [23, Theorem XIII.100] that under such conditions the operator $-\Delta + V$ in $L^2(\mathbb{R}^N)$ has purely continuous spectrum which is bounded below (but not above) and consists of closed disjoint intervals. Intervals (a, b) such that $\sigma(-\Delta+V) \cap [a, b] = \{a, b\}$ will be called *spectral gaps* of $-\Delta+V$. Let $F(x, u) := \int_0^u f(x, \xi)\, d\xi$ and suppose that f and V satisfy the following hypotheses:

(A1) V is 1-periodic in x_j, $1 \leq j \leq N$, continuous, and 0 lies in a gap of the spectrum of $-\Delta + V$.

(A2) f is 1-periodic in x_j, $1 \leq j \leq N$, and continuous.

(A3) $f(x, u)/u \to 0$ uniformly in x as $u \to 0$.

(A4) There are $c > 0$ and $p \in (2, 2^*)$ such that $|f(x, u)| \leq c(1 + |u|^{p-1})$, where $2^* := 2N/(N-2)$ if $N \geq 3$ and $2^* := +\infty$ if $N = 1$ or 2.

(A5) There is $\gamma > 2$ such that $0 < \gamma F(x, u) \leq uf(x, u)$ whenever $u \neq 0$.

Since $f(x, 0) = 0$ according to (A3), it is clear that (2.1) has the trivial solution $u = 0$.

Theorem 2. *If the hypotheses* (A1)–(A5) *are satisfied, then* (2.1) *has at least one nontrivial solution.*

For the proof, we consider the functional

$$\begin{aligned}\Phi(u) &:= \frac{1}{2}\int_{\mathbb{R}^N}(|\nabla u|^2 + V(x)u^2)\,dx - \int_{\mathbb{R}^N} F(x,u)\,dx \\ &= \frac{1}{2}\langle Lu, u\rangle - \psi(u)\end{aligned}$$

on the (real) Sobolev space $E := H^1(\mathbb{R}^N)$. Since

$$|f(x,u)| \le c_0(|u| + |u|^{p-1}) \tag{2.2}$$

according to (A3)–(A4), $\Phi \in C^1(E,\mathbb{R})$ [34, Lemma 3.10] and it is easy to see that $\nabla\Phi(u) = 0$ if and only if u is a (weak) solution of the equation in (2.1). Moreover, it can be shown that $u \in E$ and $\nabla\Phi(u) = 0$ imply $u(x) \to 0$ as $|x| \to \infty$.

To verify (i) of Theorem 1 we observe that $\psi \ge 0$ and ψ is weakly sequentially lower semicontinuous according to Fatou's lemma. Moreover,

$$\langle \nabla\psi(u), v\rangle = \int_{\mathbb{R}^N} f(x,u)v\,dx,$$

and weak sequential continuity of $\nabla\psi$ follows from (2.2) since if $u_n \rightharpoonup u$, then $u_n \to u$ in $L^2_{loc}(\mathbb{R}^N)$ and $L^p_{loc}(\mathbb{R}^N)$.

Since 0 is in a gap of the spectrum of $-\Delta + V$, E decomposes as a direct sum of two infinite-dimensional L-invariant subspaces Y, Z such that $\langle Lu, u\rangle$ is negative definite on Y and positive definite on Z (cf. [28, Section 9]). Hence (ii) of Theorem 1. The quadratic form $\langle Lu, u\rangle$ is positive definite on Z and, according to (A3), $\psi(u) = o(\|u\|^2)$ as $u \to 0$; therefore $\Phi(u) \ge b > 0$ for $u \in \partial B_\rho \cap Z$ provided ρ is small enough. This gives (iii). Since $\langle Lu, u\rangle$ is negative definite on Y and $\psi \ge 0$, $\Phi|_Y \le 0$. Using the fact that $p > 2$ one can show that $\Phi \le 0$ on the set $\{u \in M : \|u\| = R\}$ whenever R is large enough. Hence also (iv) is satisfied.

Now it follows from Theorem 1 that there exists a sequence (u_n) such that $\Phi(u_n) \to c > 0$ and $\nabla\Phi(u_n) \to 0$. Furthermore, it can be shown that (u_n) is bounded, so $u_n \rightharpoonup \bar{u}$ after passing to a subsequence. Since $\nabla\Phi$ is weakly sequentially continuous, $\nabla\Phi(\bar{u}) = 0$. If $\bar{u} \ne 0$, the proof is complete. So assume $\bar{u} = 0$. According to a lemma due to P. L. Lions (see [12, Lemma 2.18], [20, Lemma I.1] or [34, Lemma 1.21]), if (u_n) is bounded and there exists $r > 0$ such that

$$\lim_{n\to\infty} \sup_{a\in\mathbb{R}^N} \int_{|x-a|<r} u_n^2\,dx = 0, \tag{2.3}$$

then $u_n \to 0$ in $L^s(\mathbb{R}^N)$ for all $s \in (2, 2^*)$. Hence either $u_n \to 0$ in $L^p(\mathbb{R}^N)$ or there exists a sequence $(a_n) \subset \mathbb{Z}^N$ and $r, \delta > 0$ such that

$$\int_{|x-a_n|<r} u_n^2\,dx \ge \delta$$

for almost all n. In the first case one shows that $u_n \to 0$ in E which is impossible since $\Phi(u_n) \to c > 0$. In the second one $v_n(x) := u_n(x + a_n) \rightharpoonup \bar{v} \neq 0$ after taking a subsequence. Since Φ is invariant with respect to the action of $\mathbb{Z}^N$ given by

$$(a * u)(x) = u(x + a), \qquad u \in E, \quad a \in \mathbb{Z}^N, \tag{2.4}$$

we have $\Phi(v_n) = \Phi(u_n)$ and $\nabla\Phi(v_n) \to 0$, so $\bar{v}$ is the nontrivial solution we were looking for.

Theorem 2 and its proof are taken from [19], see also [34]. Earlier versions of this result, under the assumption that the function F is strictly convex, have been obtained by Alama and Li [1], and Buffoni, Jeanjean and Stuart [10]. Although the techniques in [1] and in [10] are very different, in both papers the problem is eventually reduced to that of finding a critical point of a functional having the mountain pass geometry. The hypothesis that F is convex has been removed, first by Troestler and Willem [33], and then by Kryszewski and Szulkin [19]. An extension of Theorem 2 has been recently found by Bartsch and Ding [8]. They considered the situation where 0 is a left endpoint of a gap in the spectrum of $-\Delta + V$, i.e. $[0, \beta] \cap \sigma(-\Delta + V) = \{0\}$ for some $\beta > 0$.

We would also like to mention the work of Heinz, Küpper and Stuart, see [16,28] and the references there, and that of Troestler [32], on bifurcation into spectral gaps for (2.1) with $V(x)$ replaced by $V(x) - \lambda$.

It follows immediately from the periodicity assumptions on V and f that if u is a solution of (2.1), then so is $a * u$ (cf. (2.4)) for any $a \in \mathbb{Z}^N$. Two solutions u_1 and u_2 are said to be *geometrically distinct* if $a * u_1 \neq u_2$ for any $a \in \mathbb{Z}^N$. The problem of finding the number of geometrically distinct solutions of (2.1) has been studied by several authors. If $\sigma(-\Delta+V) \subset (0, \infty)$ (i.e. if the quadratic form $\langle Lu, u\rangle$ is positive definite), it has been shown by Coti Zelati and Rabinowitz [12] that there are infinitely many such solutions. The same result remains true if 0 is in a spectral gap of $-\Delta + V$ and $f(x, u) = W(x)|u|^{p-2}u$, where $W > 0$ and $2 < p < 2^*$ [2]. For nonconvex F it has been shown in [8,19] that (2.1) has infinitely many geometrically distinct solutions under the additional assumption that f is odd in u. It seems to be an open problem to decide whether oddness of f is really needed here.

3 Hamiltonian systems

Let

$$J = \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix}$$

be the standard symplectic $2N \times 2N$- matrix. In this section we will be concerned with the question of existence of homoclinic solutions for the Hamiltonian system

$$\dot{z} = JH_z(z, t), \qquad z \in \mathbb{R}^{2N}. \tag{3.1}$$

Recall that a solution z of (3.1) is said to be *homoclinic* (to 0) if $z \not\equiv 0$ and $z(t) \to 0$ as $|t| \to \infty$. Suppose that $H(z,t) = \frac{1}{2}Az \cdot z + F(z,t)$ satisfies the following assumptions:

(B1) A is a constant symmetric $2N \times 2N$-matrix and $\sigma(JA) \cap i\mathbf{R} = \emptyset$.

(B2) F and F_z are 1-periodic in t and continuous.

(B3) $F_z(z,t)/|z| \to 0$ uniformly in t as $z \to 0$.

(B4) There exists $\gamma > 2$ such that $0 < \gamma F(z,t) \le z \cdot F_z(z,t)$ for all $z \neq 0$.

(B5) There exist $c, r > 0$ such that $|F_z(z,t)|^2 \le cz \cdot F_z(z,t)$ for all $|z| \le r$.

(B6) There exist $c, R > 0$ and $q \in (1,2)$ such that $|F_z(z,t)|^q \le cz \cdot F_z(z,t)$ for all $|z| \ge R$.

It follows from (B6) that

$$|F_z(z,t)| \le \tilde{c}(1 + |z|^{p-1}) \tag{3.2}$$

for some $\tilde{c} > 0$ and $p = q/(q-1)$.

Theorem 3. *If the hypotheses* (B1)–(B6) *are satisfied, then* (3.1) *has at least one homoclinic solution.*

Let $E := H^{1/2}(\mathbb{R}, \mathbb{R}^{2N})$ be the Sobolev space of functions $z \in L^2(\mathbb{R}, \mathbb{R}^{2N})$ such that their Fourier transform $\hat{z}$ satisfies

$$\int_{\mathbb{R}} (1 + |\xi|^2)^{1/2} |\hat{z}(\xi)|^2 \, d\xi < \infty.$$

Then E is a Hilbert space and

$$\langle z, v \rangle := \int_{\mathbb{R}} (1 + |\xi|^2)^{1/2} \hat{z}(\xi) \cdot \overline{\hat{v}(\xi)} \, d\xi$$

is an inner product in E. Consider the functional

$$\begin{aligned} \Phi(z) &:= \frac{1}{2} \int_{\mathbb{R}} (-J\dot{z} - Az) \cdot z \, dt - \int_{\mathbb{R}} F(z,t) \, dt \\ &= \frac{1}{2} \langle Lz, z \rangle - \psi(z). \end{aligned}$$

According to (3.2) and (B3), $|F_z(z,t)| \le c_0(|z| + |z|^{p-1})$. Hence using the argument of [34, Lemma 3.10] and the fact that E is continuously embedded in $L^s(\mathbb{R}, \mathbb{R}^{2N})$ for each $s \ge 2$ (see e.g. [28, Lemma 10.4]) it is easy to show that $\Phi \in C^1(E, \mathbb{R})$ and $\nabla\Phi(z) = 0$ if and only if z is a solution of (3.1). Moreover, $F_z(z(.),.) \in L^2(\mathbb{R}, \mathbb{R}^{2N})$ for such z. It follows therefore from (3.1) that $z \in H^1(\mathbb{R}, \mathbb{R}^{2N})$, so $z(t) \to 0$ as $|t| \to \infty$. Hence critical points $z \neq 0$ of Φ are homoclinic solutions of (3.1).

According to (B1), $-i\xi J - A$ is an invertible matrix and $(-i\xi J - A)^{-1}$ is uniformly bounded with respect to $\xi \in \mathbb{R}$. Hence it follows from Plancherel's

formula that L is bounded, selfadjoint and has a bounded inverse. So E decomposes as in (ii) of Theorem 1. See [28, Section 10] for more details. In particular, it is shown in [28] that the spectrum of $-J\frac{d}{dt} - A$ is unbounded, from above and below, in $H^1(\mathbb{R}, \mathbb{R}^{2N})$. Therefore $\langle Lz, z\rangle$ is positive and negative definite on subspaces of infinite dimension.

Other hypotheses of Theorem 1 are verified in the same way as in the preceding section. Hence we obtain a sequence (z_n) such that $\Phi(z_n) \to c > 0$ and $\nabla\Phi(z_n) \to 0$. Moreover, (z_n) can be shown to be bounded. The argument is the same as for the Schrödinger equation and may be found in [7]. The proof of boundedness makes essential use of (B4)–(B6). Finally, since Φ is invariant with respect to the action of $\mathbb{Z}$ given by $(a * z)(t) = z(t + a)$ $(z \in E,\ a \in \mathbb{Z})$, cf. (2.4), an application of P. L. Lions' lemma gives a solution $\bar{z} \neq 0$. Here a remark is in order: in [12, Lemma 2.18] and [34, Lemma 1.21] the space is $H^1(\mathbb{R}^N)$; however, a simple adaptation of the argument in [12,34] shows that if (z_n) is bounded in $H^{1/2}(\mathbb{R}, \mathbb{R}^{2N})$ and (2.3) is satisfied, then $z_n \to 0$ in $L^s(\mathbb{R}, \mathbb{R}^{2N})$ for all $s \in (2, +\infty)$.

Theorem 3 for Hamiltonian systems with strictly convex F is due to Coti Zelati, Ekeland and Séré [11]. They reformulated the problem in terms of a dual functional which has the mountain pass geometry. The convexity assumption has been removed by Hofer and Wysocki [17] and Tanaka [29]. The proof in [29] is obtained by constructing a sequence of subharmonic solutions of (3.1) and passing to the limit. A truncation argument is also used there in order to weaken some of the hypotheses (in particular, in [29] it is assumed that $q = 1$ in (B6), so F need not satisfy any growth restriction like (3.2)). An extension of Theorem 3 in a similar spirit as in [8] has been obtained by Ding and Willem [14]. They allowed A to be t-dependent, 1-periodic and such that $[0, \beta] \cap \sigma(-J\frac{d}{dt} - A) = \{0\}$ for some $\beta > 0$.

It has been shown by Séré [25,26] that if F is strictly convex, then (3.1) has infinitely many geometrically distinct homoclinic solutions. Recently Ding and Girardi [13] have obtained a result on the existence of infinitely many homoclinics for F which is even in z but not necessarily convex. In [7] it will be shown that the same result remains valid for F invariant with respect to an action of a more general symmetry group. Also for Hamiltonian systems it seems to be unknown whether such invariance condition can be removed if F is nonconvex.

4 Infinite chain of particles

Consider a chain of particles arranged linearly in a doubly infinite sequence. Assume that each particle has unit mass and that it interacts only with its nearest neighbours. Denote the displacement of the i-th particle from its original position by q_i and let ϕ denote the potential of interaction. Then the equations of motion for this chain are

$$\ddot{q}_i = \phi'(q_{i-1} - q_i) - \phi'(q_i - q_{i+1}), \qquad i \in \mathbb{Z}. \tag{4.1}$$

If $\phi(x) = \frac{1}{2}\beta x^2$, $\beta > 0$, the system is linear and it is possible to explicitly find normal mode solutions of (4.1), see [31].

Nonlinear systems of this kind (for a finite number of particles) were considered for the first time by Fermi, Pasta and Ulam in [15]. They wanted to verify numerically the conjecture that while there is no exchange of energy between different modes when the system is linear, already a perturbation by a small nonlinear term causes the energy to be gradually shared by the modes. Contrary to what they expected, they found that only little energy was shared and the system returned periodically to the initial state. In a subsequent research Toda has found that if the force of interaction is exponential, then the system (4.1) is integrable and there exist both periodic solutions of finite energy and soliton solutions. See [31] and the references there for more information.

Suppose now that $\phi(x) = \frac{1}{2}\beta x^2 + V(x)$ and β, V satisfy the following conditions:

(C1) $\beta > 0$.

(C2) V is continuously differentiable.

(C3) $V'(x)/x \to 0$ as $x \to 0$.

(C4) There is $\gamma > 2$ such that $0 < \gamma V(x) \le V'(x)x$ whenever $x \neq 0$.

Note that since $\phi'(x)$ has the same sign as x, the potential ϕ is purely attractive.

Theorem 4. *If the hypotheses* (C1)–(C4) *are satisfied, then* (4.1) *has a nontrivial T-periodic solution of finite energy for each $T > 0$.*

Let $q := \{q_i\}_{i\in\mathbb{Z}}$, $S^1 := [0,T]/\{0,T\}$,

$$\langle q,p\rangle := \sum_i \int_0^T (\dot q_i(t)\dot p_i(t) + (q_i(t) - q_{i+1}(t))(p_i(t) - p_{i+1}(t)))\, dt,$$

$\|q\|^2 = \langle q,q\rangle$ and

$$E := \left\{ q \in H^1(S^1,\mathbb{R})^{\mathbb{Z}} : \int_0^T q_0(t)\, dt = 0,\ \|q\| < \infty \right\}.$$

Then E is a Hilbert space and

$$\begin{aligned}\Phi(q) &:= \sum_i \frac{1}{2}\int_0^T \left(\dot q_i^2 - \beta(q_i - q_{i+1})^2\right) dt - \sum_i \int_0^T V(q_i - q_{i+1})\, dt \\ &= \frac{1}{2}\langle Lq, q\rangle - \psi(q)\end{aligned}$$

is defined on E. Moreover, $\Phi \in C^1(E,\mathbb{R})$ and critical points of Φ are T-periodic solutions of (4.1) [6]. Note that if $q = \{q_i\}$ is a solution of (4.1), so is $\tilde q = \{q_i + \sigma\}$

for any constant $\sigma \in \mathbb{R}$. Therefore the condition that $\int_0^T q_0(t)\,dt = 0$ which appears in the definition of E is a way of normalizing (4.1) by dividing out the constants.

Let $T \in (0, \pi/\sqrt{\beta})$ be fixed. Then it can be shown that Φ satisfies all hypotheses of Theorem 1. The proof is similar to that in Section 2 but more technical. Here we only show how the decomposition $E = Y \oplus Z$ is obtained and refer to [6] for the other parts.

We first make a side remark that $\langle Lq, q\rangle$ is negative definite if $\beta < 0$, $Lq = 0$ for all constant sequences q if $\beta = 0$ and it has been shown in [6] that L has no bounded inverse if $\beta \geq \pi^2/T^2$.

Let $Y := \{q \in E : q_i = const.\ \text{for all } i\}$ and $Z := Y^\perp$. Suppose $T \in (0, \pi/\sqrt{\beta})$; then $0 < \beta < \pi^2/T^2$ and a simple computation using Wirtinger's inequality shows that $\langle Lq, q\rangle$ is positive definite on Z. Clearly, $\langle Lq, q\rangle$ is negative definite on Y. Hence by Theorem 1 there exists a sequence $(q^{(n)})$ such that $\Phi(q^{(n)}) \to c > 0$ and $\nabla\Phi(q^{(n)}) \to 0$. Moreover, it can be shown $(q^{(n)})$ is bounded, so we may assume it is weakly convergent. If $q^{(n)} \rightharpoonup \bar{q} \neq 0$, we are done. Otherwise one shows there is a sequence (i_n) of integers such that if $\tilde{q}_i^{(n)} := q_{i+i_n}^{(n)} + \sigma^{(n)}$, then $\tilde{q}^{(n)} \rightharpoonup \tilde{q} \neq 0$ ($\sigma^{(n)}$ is chosen in order to have $\int_0^T \tilde{q}_0^{(n)}(t)\,dt = 0$). Since Φ is invariant with respect to the action of $\mathbb{Z}$ given by $(k * q_i)(t) = q_{i+k}(t) + \sigma_k$, a familiar argument shows that $\tilde{q}$ is a T-periodic solution of (4.1). Moreover, the energy $\frac{1}{2}\langle L\tilde{q}, \tilde{q}\rangle + \psi(\tilde{q})$ is finite.

Suppose $T \geq \pi/\sqrt{\beta}$; then we can find an integer k such that $T/k < \pi/\sqrt{\beta}$. So (4.1) has a T/k-periodic solution which of course is T-periodic as well.

The special role played by the number $T_0 := \pi/\sqrt{\beta}$ raises the question of the behaviour of solutions as $T \nearrow T_0$. A partial answer may be found in [6] where it has been shown that if there exist $c > 0$ and $p \in (2, 4)$ such that $V(x) \geq c|x|^p$, then nontrivial solutions of (4.1) bifurcate at T_0. More precisely, there exist nontrivial solutions of arbitrarily small energy and L^∞-norm, with a period arbitrarily close to T_0.

The study of chains of particles by variational methods has been initiated by Ruf and Srikanth in [24]. They considered finite chains with different kinds of (nonlinear) potential. In a series of papers Arioli, Gazzola and Terracini considered the infinite chain (4.1) with $\beta < 0$ [3,5] (potential repulsive for small and attractive for large displacements) and $\beta = 0$ [4]. Theorem 4 here is a special case of a more general result contained in [6]. Finally, let us also mention two papers, by Smets and Willem [27], and by Tarallo and Terracini [30], on solitary waves for systems of equations like (4.1).

References

1. S. Alama and Y. Y. Li, *Existence of solutions for semilinear elliptic equations with indefinite linear part*, J. Diff. Eq. **96** (1992), 89–115.
2. S. Alama and Y. Y. Li, *On "Multibump" bound states for certain semilinear elliptic equations*, Indiana J. Math. **41** (1992), 983–1026.

3. G. Arioli and F. Gazzola, *Periodic motions of an infinite lattice of particles with nearest neighbor interaction*, Nonl. Anal. TMA **26** (1996), 1103–1114.
4. G. Arioli and F. Gazzola, *Existence and approximation of periodic motions of an infinite lattice of particles*, Z. Angew. Math. Phys. **46** (1995), 898–912.
5. G. Arioli, F. Gazzola and S. Terracini, *Multibump periodic motions of an infinite lattice of particles*, Math. Z. **223** (1996), 627–642.
6. G. Arioli and A. Szulkin, *Periodic motions of an infinite lattice of particles: the strongly indefinite case*, Preprint, 1997.
7. G Arioli and A. Szulkin, *Homoclinic solutions of Hamiltonian systems with symmetry*, in preparation.
8. T. Bartsch and Y. H. Ding, *On a nonlinear Schrödinger equation with periodic potential*, Preprint, 1997.
9. V. Benci and P. H. Rabinowitz, *Critical point theorems for indefinite functionals*, Invent. Math. **52** (1979), 241–273.
10. B. Buffoni, L. Jeanjean and C. A. Stuart, *Existence of nontrivial solutions to a strongly indefinite semilinear equation*, Proc. Amer. Math. Soc. **119** (1993), 179–186.
11. V. Coti Zelati, I. Ekeland and E. Séré, *A variational approach to homoclinic orbits in Hamiltonian systems*, Math. Ann. **288** (1990), 133–160.
12. V. Coti Zelati and P. H. Rabinowitz, *Homoclinic type solutions for a semilinear elliptic PDE on* $\mathbf{R}^n$, Comm. Pure Appl. Math. **45** (1992), 1217–1269.
13. Y. H. Ding and M. Girardi, *Infinitely many homoclinic orbits of a Hamiltonian system with symmetry*, Preprint, 1997.
14. Y. H. Ding and M. Willem, *Homoclinic orbits of a Hamiltonian system*, Preprint, 1997.
15. E. Fermi, J. Pasta and S. Ulam, *Studies of Nonlinear Problems*, Los Alamos Rpt. LA — 1940 (1955); also in Collected Works of E. Fermi, Vol. II, p. 978, University of Chicago Press, 1965.
16. H. P. Heinz, T. Küpper and C. A. Stuart, *Existence and bifurcation of solutions for nonlinear perturbations of the periodic Schrödinger equation*, J. Diff. Eq. **100** (1992), 341–354.
17. H. Hofer and K. Wysocki, *First order elliptic systems and the existence of homoclinic orbits in Hamiltonian systems*, Math. Ann. **288** (1990), 483–503.
18. L. Jeanjean, *Solutions in spectral gaps for a nonlinear equation of Schrödinger type*, J. Diff. Eq. **112** (1994), 53–80.
19. W. Kryszewski and A. Szulkin, *Generalized linking theorem with an application to semilinear Schrödinger equation*, Adv. Diff. Eq., to appear.
20. P. L. Lions, *The concentration compactness principle in the calculus of variations. The locally compact case. Part II*, Ann. Inst. H. Poincaré, Analyse non linéaire **1** (1984), 223–283.
21. P. H. Rabinowitz, *Some critical point theorems and applications to semilinear elliptic partial differential equations*, Ann. Sc. Norm. Sup. Pisa Cl. Sci (4) **5** (1978), 215–223.
22. P. H. Rabinowitz, *Minimax Methods in Critical Point Theory with Applications to Differential Equations*, CBMS 65, Amer. Math. Soc., Providence, R.I., 1986.
23. M. Reed and B. Simon, *Methods of Modern Mathematical Physics, Vol. IV*, Academic Press, New York, 1978.
24. B. Ruf and P. N. Srikanth, *On periodic motions of lattices of Toda type* , Arch. Rat. Mech. Anal. **126** (1994), 369–385.
25. E. Séré, *Existence of infinitely many homoclinic orbits in Hamiltonian systems*, Math. Z. **209** (1992), 27–42.

26. E. Séré, *Looking for the Bernoulli shift*, Ann. Inst. H. Poincaré, Analyse non linéaire **10** (1993), 561–590.
27. D. Smets and M. Willem, *Solitary waves with prescribed speed on infinite lattices*, J. Funct. Anal. **149** (1997), 266–275.
28. C. A. Stuart, *Bifurcation into spectral gaps*, Bull. Belgian Math. Soc., Supplement, 1995.
29. K. Tanaka, *Homoclinic orbits in a first order superquadratic Hamiltonian system: convergence of subharmonic orbits*, J. Diff. Eq. **94** (1991), 315–339.
30. M. Tarallo and S. Terracini, *On the existence of periodic and solitary travelling waves in some non linear lattices*, Dynam. Systems Appl. **4** (1995), 429–458.
31. M. Toda, *Theory of nonlinear lattices*, Springer-Verlag, Berlin, 1981.
32. C. Troestler, *Bifurcation into spectral gaps for a noncompact semilinear Schrödinger equation with nonconvex potential*, Preprint, 1996.
33. C. Troestler and M. Willem, *Nontrivial solution of a semilinear Schrödinger equation*, Comm. P.D.E. **21** (1996), 1431–1449.
34. M. Willem, *Minimax Theorems*, Birkhäuser, Boston, 1996.

EQUADIFF 9 CD ROM, Brno 1997
Masaryk University

PROCEEDINGS
pp. 169–181

Branching of Periodic Orbits in Hamiltonian and Reversible Systems

André Vanderbauwhede

University of Gent
Department of Pure Mathematics and Computer Algebra
Krijgslaan 281, B-9000 Gent, Belgium
Email: avdb@cage.rug.ac.be

Abstract. In this paper we survey a number of results on periodic orbits in Hamiltonian and reversible systems: the appearance of such orbits in one-parameter families, the bifurcation of such families from equilibria, the period blow-up near homoclinics, the branching of subharmonics and the phenomenon of subharmonic cascades.

AMS Subject Classification. 58F, 34C

Keywords. Periodic orbits, Hamiltonian and reversible systems, subharmonic bifurcation

1 Introduction

One of the characteristic properties of Hamiltonian and reversible systems is that (symmetric) periodic orbits of such systems typically appear in one-parameter families, in contrast to periodic orbits of general systems which are typically limit cycles, i.e. they are isolated. Starting from this observation one can raise a number of questions, such as (1) how do branches of periodic orbits originate or terminate? (2) is there any "branching", i.e. can one branch of periodic orbits bifurcate from another such branch? and (3) how does this branching process change when parameters in the system are changed? In this paper we survey a number of results on these issues which we obtained in recent years in collaboration with Bernold Fiedler, Jan-Cees van der Meer, Jürgen Knobloch and Maria-Cristina Ciocci.

We will consider two different types of systems, namely Hamiltonian systems from one side, and reversible systems from the other side. Although in practice many Hamiltonian systems are also reversible, the two classes do not coincide, and we will treat them here strictly separated. Some of the results which we quote

for Hamiltonian systems remain valid for the much larger class of conservative systems, i.e. for systems which have a first integral. Also, some of the results are for fixed systems, while others require one or more external parameters.

The Hamiltonian systems which we will consider have the form

$$\dot{x} = X_H(x,\lambda) := J\nabla_x H(x,\lambda), \tag{1.1}$$

where $x \in \mathbb{R}^{2n}$, $\lambda \in \mathbb{R}^m$, $H : \mathbb{R}^{2n} \times \mathbb{R}^m \to \mathbb{R}$ is a smooth function (the Hamiltonian), and $J \in \mathcal{L}(\mathbb{R}^{2n})$ is the standard symplectic matrix defined by $J(y,z) := (z,-y)$ for all $y,z \in \mathbb{R}^n$. It is immediate to see that $H(\cdot,\lambda)$ is a first integral for $(1.1)_\lambda$. We also consider reversible systems of the form

$$\dot{x} = f(x,\lambda), \tag{1.2}$$

again with $x \in \mathbb{R}^{2n}$ and $\lambda \in \mathbb{R}^m$, and with $f : \mathbb{R}^{2n} \times \mathbb{R}^m \to \mathbb{R}^{2n}$ a smooth parameter-dependent vectorfield such that

$$f(Rx,\lambda) = -Rf(x,\lambda) \tag{1.3}$$

for some linear operator $R \in \mathcal{L}(\mathbb{R}^{2n})$ satisfying $R^2 = I$ (i.e. R is a linear involution on $\mathbb{R}^{2n}$) and $\dim \mathrm{Fix}(R) = n$. If $\tilde{x}(t)$ is a solution of (1.2) then so is $\tilde{y}(t) := R\tilde{x}(-t)$; a (maximal) solution of (1.2) with orbit γ is called *symmetric* if $R\gamma = \gamma$.

We first show why periodic orbits of (1.1) or (1.2) appear typically in one-parameter families (at fixed values of the parameter λ). Let γ_0 be a periodic orbit of a Hamiltonian vectorfield X_H, let Σ be a transversal section to γ_0 at a point $x_0 \in \gamma_0$, and let $P : \Sigma \to \Sigma$ be the corresponding Poincaré mapping. For each $h \in \mathbb{R}$ near $h_0 := H(x_0)$ we set $\mathcal{E}_h := \{x \in \mathbb{R}^{2n} \mid H(x) = h\}$ and $\Sigma_h := \Sigma \cap \mathcal{E}_h$. Since H is a first integral for X_H it follows that P leaves each Σ_h invariant, which allows us to define $P_h : \Sigma_h \to \Sigma_h$ as the restriction of P to Σ_h. Clearly x_0 is a fixed point of P_{h_0}, and if 1 is not an eigenvalue of DP_{h_0} (which is typically the case) this fixed point persists for all nearby values of h. Hence we obtain a 1-parameter family of periodic orbits parametrized by the "energy" h. In the reversible case we use the property that a nontrivial orbit γ is symmetric and periodic if and only if γ intersects $\mathrm{Fix}\,(R)$ in exactly two points; the period then equals twice the time needed to travel along γ between these two points. Now suppose that γ_0 is a symmetric periodic orbit for a reversible vectorfield $f(x)$, with minimal period $T_0 > 0$, and let x_0 and y_0 be the two intersection points of γ_0 and $\mathrm{Fix}\,(R)$. Then x_0 and y_0 also belong to the intersection of $\mathrm{Fix}\,(R)$ with $\phi_{T_0/2}(\mathrm{Fix}\,(R))$ (where $\phi_t(x)$ denotes the flow of f), and generically this intersection will be transversal. If this is the case then the two intersection points will persist for nearby values of T, i.e. for each T near T_0 the intersection of $\mathrm{Fix}\,(R)$ with $\phi_{T/2}(\mathrm{Fix}\,(R))$ will contain two points x_T and y_T which generate a symmetric T-periodic orbit of f. We conclude that typically symmetric periodic orbits of reversible vectorfields appear in one-parameter families parametrized by the period T.

In the main part of this paper we will discuss how branches of (symmetric) periodic orbits can originate at equilibria (Section 2) or terminate at homoclinics (Section 3); we will also show how the bifurcation of subharmonic solutions leads to "branching" (Section 4). Finally we will very briefly discuss the phenomenon of subharmonic cascades (Section 5).

2 Branches originating at equilibria

The simplest conditions under which a branch of periodic orbits can originate from an equilibrium are given by the classical Liapunov Center Theorem. In the Hamiltonian case this theorem reads as follows.

Theorem 1. *Consider a Hamiltonian vectorfield X_H and let $x_0 \in \mathbb{R}^{2n}$ be such that:*

(i) $X_H(x_0) = 0$*;*

(ii) $A_0 := DX_H(x_0)$ *has a pair of simple purely imaginary eigenvalues* $\pm i\omega_0$ *(with* $\omega_0 > 0$*);*

(iii) (nonresonance condition) A_0 *has no other eigenvalues of the form* $\pm ik\omega_0$*, with* $k \in \mathbb{Z}$*,* $k \neq \pm 1$*.*

Then the vectorfield X_H has a smooth 2-dimensional locally invariant manifold containing x_0 and foliated by periodic orbits surrounding x_0. As one moves along this 1-parameter family of periodic orbits towards x_0 the minimal period tends to $T_0 := 2\pi/\omega_0$. □

In the reversible case a similar result holds:

Theorem 2. *Let f be a reversible vectorfield, and let $x_0 \in \mathrm{Fix}\,(R)$ be a symmetric equilibrium of f such that the linearization $A_0 := Df(x_0)$ has a pair of simple purely imaginary eigenvalues $\pm i\,\omega_0$ $(\omega_0 > 0)$ and no other eigenvalues of the form $\pm ik\omega_0$ $(k \in \mathbb{Z},\ k \neq \pm 1)$. Then the vectorfield f has a smooth R-invariant 2-dimensional locally invariant manifold containing x_0 and foliated by a 1-parameter family of symmetric periodic orbits. As one moves along this family of periodic orbits towards the equilibrium the minimal period tends to $T_0 := 2\pi/\omega_0$.* □

The situations described by the theorems 1 and 2 are robust under perturbations: if in a parametrized family of Hamiltonian (respectively reversible) vectorfields the conditions of Theorem 1 (respectively Theorem 2) are satisfied for a certain value λ_0 of the parameter then they remain satisfied for all nearby values of the parameter. The reason for this is that if $\mu_0 \in \mathbb{C}$ is an eigenvalue of A_0 then so is $-\mu_0$; as a consequence the simple purely imaginary eigenvalues whose existence was assumed in the foregoing theorems cannot move off the imaginary axis when

the system is perturbed. However, in parametrized families of Hamiltonian or reversible systems it is possible to find in a generic way equilibria for which the linearization has a pair of purely imaginary eigenvalues for which the conditions of Theorems 1 and 2 are not satisfied, because either these eigenvalues are not simple, or because the nonresonance condition is not satisfied, or both. A well known example is that of a so-called *Krein instability* (also called a $1:1$-*resonance* or a *Hamiltonian Hopf bifurcation*) in a one-parameter family of Hamiltonian systems: in their simplest form the hypotheses are that there is an equilibrium (say at $x = 0$) at which the linearization $A_\lambda := D_x X_H(0, \lambda)$ has for small $\lambda < 0$ two pairs of simple purely imaginary eigenvalues close to each other which merge for $\lambda = 0$ into a single pair of non-semisimple purely imaginary eigenvalues and split off the imaginary axis for $\lambda > 0$. An application of Theorem 1 shows that for fixed small $\lambda < 0$ the system has two one-parameter families of periodic orbits emanating from the equilibrium $x = 0$; the question arises what happens to these periodic orbits as λ passes through zero and becomes positive.

The answer to these question depends on some third order coefficient in the normal form of the vectorfield $X_H(\cdot, 0)$, i.e. on some fourth order coefficient in the normal form of the Hamiltonian $H(\cdot, 0)$. Generically (when considering one-parameter problems as described above) this coefficient is non-zero; depending on its sign we have either an *elliptic* or a *hyperbolic* bifurcation. In the elliptic case the two families of periodic orbits which emanate from the equilibrium for $\lambda < 0$ are connected and form one single branch which at both sides tends to the equilibrium; we call this a *local branch*. As λ increases towards zero this local branch shrinks and is absorbed by the equilibrium for $\lambda = 0$. For $\lambda \geq 0$ there are no nontrivial periodic orbits nearby the equilibrium. In the hyperbolic case we have the following scenario. For $\lambda < 0$ the two families of periodic orbits emanating from the equilibrium are not connected to each other (at least not locally); we say that we have two *global branches*. For $\lambda = 0$ these two global branches become at the equilibrium tangent to each other; for $\lambda > 0$ they detach from the equilibrium and merge into one single branch of periodic orbits which no longer contains the equilibrium. A complete analysis of this Hamiltonian Hopf bifurcation can be found in [20].

The same bifurcation scenario as described above also appears at generic $1:1$-resonances in one-parameter families of conservative or reversible systems (see respectively [6] and [7]). The result can be extended to equivariant conservative or equivariant reversible systems (see [14] and [8]). It is also possible to consider situations where $k > 2$ pairs of purely imaginary eigenvalues come together and split off the imaginary axis under a change of parameters; such situations appear generically in $k-1$-parameter families of conservative or reversible systems. An analysis of the bifurcation of periodic orbits at such k-fold resonances can be found in [6] and [7].

A further question which arises in the context of such resonances is about the stability of the periodic orbits appearing in these bifurcation scenario's. It is important to notice that if $\mu \in \mathbb{C}$ is a characteristic multiplier of a periodic orbit

in a Hamiltonian or reversible system, then so is $1/\mu$; consequently a periodic orbit is called *stable* if all its multipliers are on the unit circle, and *unstable* if there are some multipliers off the unit circle. Taking into account only the critical multipliers it can be shown for the Hamiltonian Hopf bifurcation described above (see [20]) that in the hyperbolic case all periodic orbits appearing in the bifurcation scenario are stable; in the elliptic case the local branch which exists for $\lambda < 0$ is divided into three parts: the periodic solutions along the middle part are unstable, those along the two outer parts (adjacent to the equilibrium) are stable. The same result also holds at a $1:1$-resonance in reversible systems (see [4] and [9]); here the transition points between stable and unstable solutions along the local branch in the elliptic case are sometimes called *Eckhaus points*. At these Eckhaus points there can be secondary bifurcations, in particular of orbits homoclinic to periodic orbits (again, see [4]). Finally, the stability of periodic orbits near a 3-fold resonance in reversible systems will be discussed in some forthcoming paper [9].

There are several tools available for studying the bifurcation of periodic orbits at resonances in Hamiltonian or reversible systems; the most popular ones are the Liapunov-Schmidt reduction and normal form theory. We conclude this section by describing a general type of reduction result which can (and has) been used for analyzing the type of resonances considered here. More details and proofs can be found in [17] and [5]. These proofs are based on a combined use of normal form theory and the Liapunov-Schmidt reduction; however, the reduction result can be used directly, without going into the details of either of these methods (see [6] and [7] for some examples).

Consider a system

$$\dot{x} = f(x, \lambda), \tag{2.1}$$

where the vectorfield $f : \mathbb{R}^{2n} \times \mathbb{R}^m \to \mathbb{R}^{2n}$ is either Hamiltonian or reversible, and satisfies $f(0, \lambda) = 0$ for all λ. We are then interested in solving the following problem:

(**P**) Find, for all (λ, T) near a given $(\lambda_0, T_0) \in \mathbb{R}^m \times]0, \infty[$, all sufficiently small T-periodic solutions of $(2.1)_\lambda$.

Let $A_0 := D_x f(0, \lambda_0)$ be the linearization of $f(\cdot, \lambda_0)$ at the equilibrium in the origin, and assume that A_0 is nonsingular, such that there is no bifurcation of equilibria at $\lambda = \lambda_0$. Let $A_0 = S_0 + N_0$ be the Jordan decomposition of A_0 into its semisimple and nilpotent parts (i.e. S_0 is semisimple, N_0 is nilpotent, and $S_0 N_0 = N_0 S_0$). Next we introduce the so-called *reduced phase space* for our problem; this is a subspace of $\mathbb{R}^{2n}$ defined by

$$U := \ker \left(e^{S_0 T_0} - I \right). \tag{2.2}$$

There exists a natural S^1-action on U, generated by $S := S_0|_U$ and explicitly given by

$$\varphi \in S^1 \cong \mathbb{R}/T_0\mathbb{Z} \longmapsto e^{S\varphi} \in \mathcal{L}(U). \tag{2.3}$$

Also, the space U is even-dimensional and invariant under J or R depending on whether f is Hamiltonian or reversible; therefore it makes sense to talk about a Hamiltonian (respectively reversible) vectorfield on U.

We have then the following reduction result.

Theorem 3. *Under the foregoing conditions there exists for each* (λ, T) *near* (λ_0, T_0) *a one-to-one correspondence between the small* T*-periodic solutions of* $(2.1)_\lambda$ *and the small* T*-periodic solutions of a* reduced equation

$$\dot{u} = f_r(u, \lambda), \tag{2.4}$$

where the reduced vectorfield $f_r : U \times \mathbb{R}^m \to U$ *has the following properties:*

(1) $f_r(0, \lambda) = 0$ *for all* λ*, and* $D_u f_r(0, \lambda_0) = S + N$*, where* $N := N_0|_U$*;*

(2) f_r *is Hamiltonian or reversible, depending on whether* f *is Hamiltonian or reversible;*

(3) f_r *is* S^1*-equivariant, i.e. we have*

$$f_r(e^{S\varphi}u, \lambda) = e^{S\varphi} f_r(u, \lambda), \qquad \forall \varphi \in S^1; \tag{2.5}$$

Moreover, all small T*-periodic solutions of* $(2.4)_\lambda$ *have the form*

$$\tilde{u}(t) = e^{(1+\sigma)St}u \tag{2.6}$$

with $u \in U$ *small and* $T = T_0/(1 + \sigma)$. □

An analogous result holds for conservative systems, under appropriate non-degeneracy conditions for the first integral. The last conclusion of Theorem 3 combined with the S^1-equivariance of f_r shows that in order to obtain the bifurcation picture for our problem (**P**) we have to study the *determining equation*

$$(1 + \sigma)Su = f_r(u, \lambda) \tag{2.7}$$

for (u, λ, σ) near $(0, \lambda_0, 0)$. It is shown in [17] and [5] how the reduced vectorfield f_r can be calculated or approximated by bringing the original vectorfield f into normal form. It should be emphasized that although (2.7) is a finite-dimensional equation it is in general not yet the *bifurcation equation* for our problem (**P**) since its linearization at $(u, \lambda, \sigma) = (0, \lambda_0, 0)$ is not identically zero but gives the equation $Nu = 0$; however, when the nilpotent operator N is known it is fairly simple to deduce the bifurcation equations from (2.7).

3 Branches terminating at homoclinics

When moving along a branch of periodic orbits in a Hamiltonian, conservative or reversible system it is possible that the period tends to infinity while the

orbit itself remains bounded; the limiting orbit may then for example be a homoclinic orbit. Examples of such *homoclinic period blow-up* are well known for one-degree-of-freedom Hamiltonian systems, i.e. when $n = 1$ in (1.1). Consider for example the phase portrait for the Hamiltonian system with Hamiltonian $H(y, z) = 1/2z^2 + y^3 - y^2$; this system has two equilibria, a center and a saddle; the periodic orbits which originate at the center terminate in a period blow-up at an orbit homoclinic to the saddle. In [15] it is shown that this type of behavior is typical near (symmetric) homoclinic orbits in conservative or reversible systems, whatever their dimension. In this section we briefly describe the main result of [15].

We consider a system

$$\dot{x} = f(x). \tag{3.1}$$

In the conservative case we assume that $x \in \mathbb{R}^n$, that $f : \mathbb{R}^n \to \mathbb{R}^n$ is smooth, and that there exists a smooth function $H : \mathbb{R}^n \to \mathbb{R}$ such that $DH(x) \cdot f(x) = 0$ for all $x \in \mathbb{R}^n$; moreover it is assumed that:

(**C**) (i) there exists an orbit γ_0 of (3.1) which is homoclinic to a hyperbolic equilibrium $x_0 \in \mathbb{R}^n$;

(ii) the homoclinic orbit γ_0 is non-degenerate, i.e.

$$\dim\,(T_y W^s(x_0) \cap T_y W^u(x_0)) = 1, \qquad \forall y \in \gamma_0, \tag{3.2}$$

where $W^s(x_0)$ and $W^u(x_0)$ denote the stable (respectively unstable) manifold of x_0;

(iii) $DH(y_0) \neq 0$ for some $y_0 \in \gamma_0$.

These hypotheses are robust under perturbations and imply a period blow-up at γ_0; more precisely:

Theorem 4. *Under the assumptions* (**C**) *we have that* $\gamma_0 \cup \{x_0\}$ *forms the limit of a one-parameter family of periodic orbits along which the minimal period* T *tends to infinity as one approaches the homoclinic orbit.* □

When the system (3.1) is reversible (see Section 1) one has to assume that the homoclinic orbit is symmetric, and then necessarily also the limiting equilibrium is symmetric. Such symmetric homoclinic orbits have a unique intersection point with $\mathrm{Fix}\,(R)$. Also, if $x_0 \in \mathrm{Fix}\,(R)$ is a symmetric and hyperbolic equilibrium, then both the stable manifold $W^s(x_0)$ and the unstable manifold $W^u(x_0)$ have dimension n, since $W^u(x_0) = R(W^s(x_0))$. This allows us to formulate our hypotheses for the reversible case as follows:

(**R**) (i) the system (3.1) is reversible and has a symmetric orbit γ_0 (i.e. $R(\gamma_0) = \gamma_0$) which is homoclinic to a symmetric and hyperbolic equilibrium $x_0 \in \mathrm{Fix}\,(R)$;

(ii) γ_0 is an *elementary* homoclinic orbit, which means that $W^s(x_0)$ and $\mathrm{Fix}\,(R)$ intersect transversely at the unique intersection point of γ_0 and $\mathrm{Fix}\,(R)$.

Again these hypotheses are robust under perturbations, and they imply a homoclinic period blow-up along a family of *symmetric* periodic orbits.

Theorem 5. *Under the assumptions* (**R**) *we have that* $\gamma_0 \cup \{x_0\}$ *forms the limit of a one-parameter family of symmetric periodic orbits along which the minimal period* T *tends to infinity as one approaches the homoclinic orbit.* □

The proofs of these theorems as given in [15] is based on a simplified form of Lin's method (see [10]); this method has recently become quite popular for the study of bifurcations near homoclinics (see e.g. the recent work of B. Sandstede).

4 Subharmonic branching

In the foregoing sections we have seen how branches of periodic orbits in Hamiltonian or reversible systems can originate at equilibria or terminate at homoclinics. In this section we discuss some elementary "branching phenomena" which can occur along branches of periodic orbits; we also describe a reduction result for mappings (analogous to Theorem 3) which can be used to study such branchings. For the sake of simplicity we will restrict here to Hamiltonian systems, although most of the results have analogues for reversible systems (see e.g. [16] for a study of subharmonic branching in reversible systems).

To start consider a Hamiltonian system

$$\dot{x} = X_H(x) = J\nabla_x H(x), \tag{4.1}$$

with $x \in \mathbb{R}^{2n}$ and $H : \mathbb{R}^{2n} \to \mathbb{R}$ smooth. Let γ_0 be a given (nontrivial) periodic orbit of (4.1), $x_0 \in \gamma_0$ and $h_0 := H(x_0)$. As described in the Introduction we can then construct a one-parameter family of (restricted) Poincaré maps

$$P_h : \Sigma_h \longrightarrow \Sigma_h, \tag{4.2}$$

well defined for $h \in \mathbb{R}$ close to h_0 (see Section 1 for the notations). Fixed points of P_h correspond to periodic orbits of (4.1) close to γ_0, periodic orbits of P_h correspond to so-called *subharmonic solutions* of (4.1), i.e. periodic solutions whose orbit remains in a neighborhood of γ_0 but whose minimal period is close to an integer multiple of of γ_0. So the study of periodic orbits near γ_0 leads to an analysis of the bifurcation of fixed points and periodic points from the fixed point x_0 of P_{h_0}. The following properties of Σ_h and P_h are crucial for this analysis.

Lemma 6. *We have for each* h *near* h_0 *that* Σ_h *is a* $2(n-1)$*-dimensional symplectic submanifold of* $\mathbb{R}^{2n}$*, and* P_h *is a symplectic diffeomorphism.* □

For a proof see e.g. [13] or [19]. Using the classical Darboux theorem (see [1] or [19]) this Lemma implies that the family of mappings P_h ($h \in \mathbb{R}$) can be identified with a one-parameter family of symplectic diffeomorphisms on a fixed *symplectic vectorspace* (V, J) with $\dim V = 2(n-1)$; this means that $J \in \mathcal{L}(V)$ is anti-symmetric with respect to some scalar product on V and satisfies $J^2 = -I_V$, while the diffeomorphisms $P_h : V \to V$ are such that

$$DP_h(x)^T J DP_h(x) = J, \qquad \forall x \in V. \tag{4.3}$$

In this identification the base point x_0 at which we constructed the Poincaré map corresponds to the origin of V; hence we have $P_{h_0}(0) = 0$. The eigenvalues of $DP_{h_0}(0)$ are the nontrivial characteristic multipliers of the periodic orbit; observe that because of the symplectic structure 1 will always be a multiplier with at least multiplicity 2. Generically 1 will be a multiplier with multiplicity equal to 2, and in that case 1 will not be an eigenvalue of $DP_{h_0}(0)$ and the fixed point of P_{h_0} will persist for nearby values of h. Therefore we can (possibly after an appropriate translation) assume that

$$P_h(0) = 0, \qquad \forall h \in \mathbb{R}. \tag{4.4}$$

The fixed point set $\{(0, h) \mid h \in \mathbb{R}^m\}$ corresponds to the branch of periodic solutions of (4.1) which we discussed in Section 1.

Now let us consider the eigenvalues of $DP_h(0) \in \mathcal{L}(V)$. Setting $x = 0$ in (4.3) it is easy to show that if $\mu \in \mathbb{C}$ is an eigenvalue of $DP_h(0)$ then so are $1/\mu$, $\bar{\mu}$ and $1/\bar{\mu}$. It follows that if $DP_{h_0}(0)$ has a pair of *simple* eigenvalues $\{\mu, \bar{\mu}\}$ on the unit circle (i.e. $|\mu| = 1$ and $\mu \neq \pm 1$), then the continuation of these eigenvalues stays on the unit circle for all h near h_0. Hence we expect to see many values of h for which $DP_h(0)$ has a pair of simple eigenvalues which are *roots of unity*, i.e. of the form $\exp(\pm 2\pi i p/q)$, with $0 < p < q$ and $\gcd(p, q) = 1$. This means that the linearization $DP_h(0)$ has q-periodic points, and hence there is a possibility that in the family of diffeomorphisms P_h we see a bifurcation of q-periodic points from the fixed point at 0. This in turn would mean that we have bifurcation of subharmonic solutions near γ_0 for our original Hamiltonian system (4.1).

We now describe a general reduction result which is very useful in studying the bifurcation of periodic points from fixed points in families of symplectic mappings and which forms an analogy for mappings of what we found in Theorem 3 for vectorfields. The proof can be found in [2], and a similar result for reversible mappings will be given in [3].

Let (V, J) be a symplectic vectorspace and $\Phi : V \times \mathbb{R}^m \to V$ a parametrized family of symplectic diffeomorphisms, i.e. we have

$$D\Phi_\lambda(x)^T J D\Phi_\lambda(x) = J, \qquad \forall x \in V, \forall \lambda \in \mathbb{R}^m, \tag{4.5}$$

with $\Phi_\lambda := \Phi(\cdot, \lambda)$ for $\lambda \in \mathbb{R}^m$. We also assume that $\Phi_\lambda(0) = 0$ for all λ, and (taking $\lambda = 0$ as a critical parameter value) we set $A_0 := D\Phi_0(0)$. Given an integer $q \geq 1$ we then consider the following problem:

($\mathbf{P}_q$) Find, for all small λ, all small q-periodic points of Φ_λ.

To solve ($\mathbf{P}_q$) we have to solve the equation

$$\Phi_\lambda^q(x) = x \tag{4.6}$$

for all (x, λ) near $(0, 0)$. We notice that this equation has an *implicit $\mathbb{Z}_q$-symmetry*: if, for a given λ, $x \in V$ is a solution of (4.6), then so are $\Phi_\lambda(x)$, $\Phi_\lambda^2(x)$, $\ldots$, $\Phi_\lambda^{q-1}(x)$ and $\Phi_\lambda^q(x) = x$. The result which follows will make this implicit symmetry explicit, so that in applications it can be used to simplify the equations. Let $A_0 = S_0 + N_0$ be the Jordan decomposition of A_0 into its semisimple part S_0 and its nilpotent part N_0, and define the *reduced phase space* U by

$$U := \ker\,(S_0^q - I) \tag{4.7}$$

One can then show that S_0 is a symplectic linear operator on V, that U is a symplectic subspace of V, and that $S := S_0|_U \in \mathcal{L}(U)$ generates a natural symplectic $\mathbb{Z}_q$-action on U.

Theorem 7. *For each sufficiently small λ there exists a one-to-one correspondence between the small q-periodic points of Φ_λ and the small q-periodic points of a* reduced mapping $\Phi_{r,\lambda}$, *where*

$$\Phi_r : U \times \mathbb{R}^m \longrightarrow U$$

has the following properties:

(i) $\Phi_r(0, \lambda) = 0$ *for all* λ, *and* $D_u\Phi_r(0, 0) = S + N$, *where* $N := N_0|_U$;

(ii) $\Phi_{r,\lambda}$ *is a symplectic diffeomorphism on* U, *for each* λ;

(iii) Φ_r *is $\mathbb{Z}_q$-equivariant, i.e. we have*

$$\Phi_r(Su, \lambda) = S\Phi_r(u, \lambda). \tag{4.8}$$

Moreover, all small q-periodic orbits of $\Phi_{r,\lambda}$ are also $\mathbb{Z}_q$-orbits, i.e. they can be found by solving the $\mathbb{Z}_q$-equivariant determining equation

$$\Phi_r(u, \lambda) = Su. \tag{4.9}$$

Finally, the reduced mapping Φ_r can be approximated up to any finite order by bringing the original mapping Φ into normal form. □

We conclude this section with a brief indication on how the reduction result of Theorem 7 can be used to prove a classical result of Meyer [11] on the bifurcation of periodic points in symplectic mappings. Again the details of our approach can be found in [2]. We take $m = 1$ in the foregoing and fix some $q \geq 3$. We also assume the following:

(a) A_0 has a pair of simple eigenvalues $\exp(\pm 2\pi ip/q)$, with $0 < p < q$ and $\gcd(p, q) = 1$;

(b) A_0 has no other eigenvalues μ such that $\mu^q = 1$.

Then dim $U = 2$ and we can identify U with the complex plane, such that the reduced mapping Φ_r given by Theorem 7 now becomes a mapping from $\mathbb{C} \times \mathbb{R}$ into $\mathbb{C}$. The $\mathbb{Z}_q$-equivariance (4.7) then takes the form

$$\Phi_r(e^{2\pi ip/q} z, \lambda) = e^{2\pi ip/q} \Phi_r(z, \lambda), \qquad \forall (z, \lambda) \in \mathbb{C} \times \mathbb{R}, \tag{4.10}$$

while the determining equation (4.9) becomes

$$\Phi_r(z, \lambda) = e^{2\pi ip/q} z. \tag{4.11}$$

It was shown in [18] that (4.10) implies that Φ_r must have the form

$$\Phi_r(z, \lambda) = \phi_1(z, \lambda)\, z + \phi_2(z, \lambda)\, \bar{z}^{q-1}, \tag{4.12}$$

with the functions $\phi_i : \mathbb{C} \times \mathbb{R} \to \mathbb{C}$ $(i = 1, 2)$ such that

$$\phi_i(e^{2\pi ip/q} z, \lambda) = \phi_i(z, \lambda) = \phi_i(\bar{z}, \lambda), \qquad \forall (z, \lambda) \in \mathbb{C} \times \mathbb{R},\ i = 1, 2. \tag{4.13}$$

Since Φ_r is symplectic it is also area-preserving, which in combination with (4.12) and (4.13) implies that

$$|\phi_1(z, \lambda)| = 1 + \mathrm{O}(|z^q|).$$

Using polar coordinates and assuming some generically satisfied conditions one can then solve the determining equation (4.11). The result is that (4.11) has $2q$ branches of nontrivial solutions, each parametrized by the amplitude ρ of z, and of the form

$$\{(\rho e^{i(\theta_i^*(\rho) + 2\pi jp/q)}, \lambda_i^*(\rho)) \mid 0 < \rho < \rho_0\}, \qquad (0 \le j \le q - 1,\ i = 1, 2)$$

with $\theta_2^*(0) = \theta_1^*(0) + \pi/q$, $\lambda_i^*(\rho) = \mathrm{O}(\rho^2)$ $(i = 1, 2)$ and $\lambda_2^*(\rho) = \lambda_1^*(\rho) + \mathrm{O}(\rho^{q-2})$. So there are two branches of periodic orbits which bifurcate at $\lambda = 0$ from the fixed point at the origin in the family of symplectic diffeomorphisms Φ_λ $(\lambda \in \mathbb{R})$.

When we apply the foregoing result to the family P_h $(h \in \mathbb{R})$ of Poincaré maps discussed in the beginning of this section we conclude the following: when at some point along a one-parameter branch of periodic orbits of (4.1) the nontrivial characteristic multipliers satisfy the conditions (a) and (b) (for some $q \ge 3$) then at that point two branches of subharmonic solutions will bifurcate from the first branch. The higher the value of q, the closer to each other these two branches will be. The bifurcating subharmonic solutions will (next to the double multiplier 1) have two multipliers close to 1; along one of the two branches these "critical multipliers" are on the real axis, along the other branch they are on the unit circle.

5 Subharmonic cascades

In this last section we briefly indicate an interesting but still largely open problem. Consider again, as in the foregoing section, a one-parameter branch of periodic orbits of the Hamiltonian system (4.1). Assume that along part of this branch there are some simple multipliers on the unit circle. Then there will generically be an infinite number of points along the branch where some multipliers are roots of unity and where we will have bifurcation of two branches of subharmonic solutions. At most of these bifurcation points the value of q will be high, and hence the bifurcating subharmonics will have large periods. Now concentrate on one such branching point; as indicated at the end of Section 4 the multipliers along one of the two bifurcating branches will be on the unit circle and close to 1. Hence, applying again the same results, we will find along this secondary branch an infinite number of points where two branches of subharmonics bifurcate; since the critical multipliers along the secondary branch are close to 1 the subharmonics bifurcating from this branch will have very high periods (corresponding to very large values of q). Iterating this argument we obtain a *cascade* of subharmonic branchings, all in the same fixed Hamiltonian system (4.1). A similar argument can be given for reversible systems. It leads to a very rich and complicated structure for the set of periodic orbits of Hamiltonian or reversible systems, and it would certainly be interesting to understand this structure in a more global way.

The methods described in the foregoing sections do not allow such global study since they are local (near each of the branching points) and they concentrate on solutions with a given (approximate) period. One will need a different approach to answer such questions as: (i) is there any self-similarity in such cascades? (ii) can one use renormalization techniques? (iii) are there any universal constants? In some very particular cases (mainly concentrating on period-doubling) some of these questions have been answered by a number of authors such as M. Feigenbaum and R. MacKay. In the reversible case there is some recent contribution by J. Roberts and J. Lamb ([12]). But to a large extend the problem remains open.

References

1. R. Abraham and J. Marsden. *Foundations of Mechanics*. Benjamin/Cummings Publ. Co., Reading, Massachusetts, 1978.
2. M. C. Ciocci and A. Vanderbauwhede. Bifurcation of periodic orbits for symplectic mappings. Journ. Diff. Eqns. and Appl. 3 (1998) 485–500.
3. M. C. Ciocci and A. Vanderbauwhede. Bifurcation of periodic points in reversible mappings. In preparation.
4. G. Iooss and M.-C. Pérouème. Perturbed homoclinic solutions in reversible $1:1$ resonance vector fields. Journ. Diff. Eqns. 102 (1993) 62–88.
5. J. Knobloch and A. Vanderbauwhede. A General Reduction Method for Periodic Solutions in Conservative and Reversible Systems. J. Dyn. Diff. Eqns. 8 (1996) 71–102.

6. J. Knobloch and A. Vanderbauwhede. Hopf bifurcations at k-fold resonances in conservative systems. In: H. W. Broer, S. A. van Gils, I. Hoveijn and F. Takens (Eds.), *Nonlinear Dynamical Systems and Chaos*, Birkhäuser, Progress in Nonlin. Diff. Eqns. and Their Appl., Vol. 19 (1996) 155–170.
7. J. Knobloch and A. Vanderbauwhede. Hopf bifurcation at k-fold resonances in reversible systems. Preprint T. U. Ilmenau, 1995.
8. J. Knobloch and A. Vanderbauwhede. Hopf bifurcation at k-fold resonances in equivariant reversible systems. In: P. Chossat (Ed.), *Dynamics, Bifurcation and Symmetry. New Trends and New Tools*, NATO ASI Series C, Vol. 437, Kluwer Academic, Dordrecht (1994) 167–179.
9. J. Kobloch and A. Vanderbauwhede. Stability of periodic orbits bifurcating at k-fold resonances in reversible systems. In preparation.
10. X.-B. Lin. Using Melnikov's method to solve Silnikov's problems. Proc. Royal Society of Edinburgh 116A (1990) 295–325.
11. K. R. Meyer. Generic bifurcation of periodic points. Trans. Amer. Math. Soc. 149 (1970) 95–107.
12. J. Roberts and J. Lamb. Self-similarity of period-doubling branching in 3-D reversible mappings. Physica D 82 (1995) 317–332.
13. F. Takens. Hamiltonian systems: generic properties of closed orbits and local perturbations. Math. Ann. 188 (1970) 304–312.
14. A. Vanderbauwhede Hopf bifurcation for equivariant conservative and time-reversible systems. Proc. Royal Society of Edinburgh 116A (1990) 103–128.
15. A. Vanderbauwhede and B. Fiedler. Homoclinic period blow-up in reversible and conservative systems. Zeitschrift für Angew. Math. Phys. (ZAMP) 43 (1992) 292–318.
16. A. Vanderbauwhede. Branching of periodic solutions in time-reversible systems. In: H. Broer and F. Takens (Eds.), *Geometry and Analysis in Non-Linear Dynamics*. Pitman Res. Notes in Math. 222 (1992) 97–113.
17. A. Vanderbauwhede and J.-C. van der Meer. A general reduction method for periodic solutions near equilibria in Hamiltonian systems. In: W. F. Langford and W. Nagata (Eds.), *Normal Forms and Homoclinic Chaos*, Fields Institute Communications, A.M.S. Providence (1995) 273–294.
18. A. Vanderbauwhede. Subharmonic bifurcation at multiple resonances. Preprint University of Gent, 1996. To appear in the Proceedings of the 2nd Marrakesh International Conference on Differential Equations.
19. A. Vanderbauwhede. A short tutorial on Hamiltonian systems and their reduction near a periodic orbit. Preprint University of Gent, 1997.
20. J.-C. van der Meer. *The Hamiltonian Hopf Bifurcation*. Lect. Notes in Math. 1160, Springer-Verlag, Berlin, 1986.

EQUADIFF 9 CD ROM, Brno 1997
Masaryk University

PROCEEDINGS
pp. 183–200

Some Partial Differential Volterra Equation Problems Arising in Viscoelasticity

Simon Shaw and J. R. Whiteman

BICOM, Brunel University, Uxbridge, Middlesex, UB8 3PH, England
Email: simon.shaw@brunel.ac.uk
john.whiteman@brunel.ac.uk
WWW: http://www.brunel.ac.uk/~icsrbicm

Abstract. The constitutive law relating stress to strain for viscoelastic materials can be written as a Volterra equation of the second kind. This results in the mathematical models of viscoelastic behaviour taking the form of partial differential equations with memory. In this article we illustrate how the memory terms arise in these equations and also summarize the various partial differential Volterra equations used when modelling problems of quasistatic and dynamic viscoelasticity, and non-Fickian diffusion in polymers. We also indicate some of the numerical analysis work that has been carried out for these problems.

AMS Subject Classification. 73F15, 45D05, 45K05, 65M15

Keywords. Volterra equation, viscoelasticity, finite element method, error estimates

1 Introduction

This paper is concerned with the modelling of problems involving viscoelastic materials which, even in their simplest form, exhibit behaviour characteristic of both classical Hookean solids and Newtonian fluids. The resulting effects are important when the material is deforming under an applied load. This load could, for example, be due to externally applied forces; internal deformation caused by a diffusing penetrant; or, constrained thermal expansion caused by temperature gradients. See for example [21,6,28]. Moreover, the material somehow keeps a record of its response history and, for this reason, viscoelastic materials are said to possess *memory*. This memory is manifest in the constitutive relationship between the stress and strain tensors, $\underline{\boldsymbol{\sigma}}$ and $\underline{\boldsymbol{\varepsilon}}$, and as a result mathematical models of viscoelastic behaviour take the form of partial differential Volterra (PDV) equation problems. The canonical forms of these equations are: the *elliptic Volterra* problem,

$$Au(t) = f(t) + \int_0^t B(t,s)u(s)\,ds; \tag{1.1}$$

the *parabolic Volterra* problem,

$$u'(t) + Au(t) = f(t) + \int_0^t B(t,s)u(s)\,ds; \tag{1.2}$$

and, the *hyperbolic Volterra* problem,

$$u''(t) + Au(t) = f(t) + \int_0^t B(t,s)u(s)\,ds. \tag{1.3}$$

These are supplied with initial and/or boundary data as appropriate, and the dependence on the space variable $\boldsymbol{x}$ is suppressed. In these problems we use A and $B(t,s)$ to represent partial differential operators (acting only in the space variables) where, for example, we could have

$$A := -\nabla^2 \qquad \text{and} \qquad B(t,s) := -\nabla \cdot \phi(t,s)\nabla,$$

although for (1.1) and (1.3) the appropriate form for A is the linear elasticity operator—with $B(t,s)$ "similar".

The purpose of this article is to illustrate how the memory terms arise in these equations and also to summarize the various PDV equations used when modelling problems of *quasistatic* and *dynamic* viscoelasticity, and *non-Fickian* diffusion in polymers. We also indicate some of the numerical analysis work that has been carried out for these problems (but we do not claim to be exhaustive, for a fuller account see [39]).

Throughout, the positive real number T will denote a final time and we use $\mathcal{J} := [0,T]$ and $\mathcal{I} := (0,T]$ to denote time intervals. Also, for $n = 1, 2$ or 3 we consider $\Omega \subset \mathbb{R}^n$ to be an open bounded domain with boundary $\partial\Omega$. Furthermore, we consider $\partial\Omega$ in the form

$$\partial\Omega := \overline{\Gamma_D \cup \Gamma_N} \qquad \text{with} \qquad \Gamma_D \cap \Gamma_N = \emptyset,$$

where the closed set $\Gamma_D \subseteq \partial\Omega$ is called the *Dirichlet boundary* and is of positive measure so that

$$\int_{\Gamma_D} d\Gamma > 0.$$

We call the (possibly empty) open set $\Gamma_N \subset \partial\Omega$ the *Neumann boundary*. The reason for this terminology is the obvious one where we refer to the type of boundary condition specified on these subsets. We indicate vector-valued quantities with boldface so that, for example, we use $\boldsymbol{x} := (x_i)_{i=1}^n$ to indicate a point in $\mathbb{R}^n$. Tensors are indicated by a further underlining: $\underline{\boldsymbol{\sigma}} = (\sigma_{ij})_{i,j=1}^n$.

2 Hereditary constitutive relationships

Suppose that the interior of a compressible viscoelastic body $\mathcal{G}$ occupies Ω and that its surface coincides with $\partial\Omega$. If at a time t this body is subjected to

a system of body forces $\boldsymbol{f} := (f_i(\boldsymbol{x},t))_{i=1}^n$, for $\boldsymbol{x} \in \Omega$, and surface tractions $\boldsymbol{g} := (g_i(\boldsymbol{x},t))_{i=1}^n$, for $\boldsymbol{x} \in \Gamma_N$, then the body $\mathcal{G}$ will deform from its equilibrium configuration. A material particle originally at the point $\boldsymbol{x}$ will move to the new time dependent location $\boldsymbol{x}+\boldsymbol{u}(\boldsymbol{x},t)$ where $\boldsymbol{u} := (u_i)_{i=1}^n$ denotes the displacement vector. In the linear theory these displacements define the symmetric strain tensor $\underline{\boldsymbol{\varepsilon}} := (\varepsilon_{ij})_{i,j=1}^n$ by the relationships:

$$\varepsilon_{ij}(\boldsymbol{u}) := \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right). \tag{2.1}$$

In addition to this strain field there will also be induced in $\mathcal{G}$ a stress field described by the symmetric stress tensor $\underline{\boldsymbol{\sigma}} := (\sigma_{ij})_{i,j=1}^n$. This stress field rationalizes the internal force field which is set up within $\mathcal{G}$ to resist the external forces $\boldsymbol{f}$ and $\boldsymbol{g}$.

The stress field can be related to $\boldsymbol{u}$, $\boldsymbol{f}$ and $\boldsymbol{g}$ by Newton's second law of motion (see later in equation (3.1)) and so it is of interest to derive a *constitutive relationship* linking $\underline{\boldsymbol{\sigma}}$ and $\boldsymbol{u}$, or in practice, linking the tensors $\underline{\boldsymbol{\sigma}}$ and $\underline{\boldsymbol{\varepsilon}}$.

In classical linear elasticity theory this relationship is provided by Hooke's law:

$$\sigma_{ij} = D_{ijkl}\varepsilon_{kl} \qquad \text{or} \qquad \underline{\boldsymbol{\sigma}} = \underline{\boldsymbol{D}}\,\underline{\boldsymbol{\varepsilon}},$$

where $\underline{\boldsymbol{D}}$ is a positive-definite fourth-order tensor of elastic coefficients satisfying the symmetries

$$D_{ijkl} = D_{jikl}, \qquad D_{ijkl} = D_{ijlk}, \qquad \text{and} \qquad D_{ijkl} = D_{klij}.$$

The first two of these are implied by the symmetry of $\underline{\boldsymbol{\sigma}}$ and $\underline{\boldsymbol{\varepsilon}}$ while the third follows from energy considerations. However, in viscoelasticity the third of these only applies when the material is isotropic, see [21, Equations (1.10) and (2.62)].

One way of deriving a constitutive relationship for viscoelastic materials is to assume that a Boltzmann superposition of stress increments can be applied where these stress increments are related by Hooke's law to corresponding strain increments. For example, suppose that $\mathcal{G}$ is quiescent for $t < 0$ so that $\underline{\boldsymbol{\varepsilon}}(t) \equiv \underline{\boldsymbol{0}}$ for $t < 0$, and that at $t = 0$ the body undergoes a strain $\underline{\boldsymbol{\varepsilon}}(0)$. Then for $t \geq 0$ the resulting stress is assumed to be given by

$$\underline{\boldsymbol{\sigma}}_0(t) = \underline{\boldsymbol{D}}(t)\underline{\boldsymbol{\varepsilon}}(0),$$

where a time dependence has been introduced into the Hooke's tensor $\underline{\boldsymbol{D}}$. Physically we expect $\underline{\boldsymbol{D}}$ to be a smooth monotone decreasing function of t since it is unrealistic to expect $\underline{\boldsymbol{\sigma}}$ to grow over time for the fixed strain $\underline{\boldsymbol{\varepsilon}}(0)$. (Where would the strain energy come from?) In fact experiments on polymers show that $\underline{\boldsymbol{D}}$ does in fact decrease and this phenomena is known as *stress relaxation*.

Now, let Δt be a small time interval and set $t_i := i\Delta t$. We approximate the strain evolution by the step function

$$\widetilde{\underline{\boldsymbol{\varepsilon}}}(t) := \underline{\boldsymbol{\varepsilon}}(t_i) \qquad \text{in } [t_i, t_{i+1}) \text{ for } i = 0, 1, 2, \ldots,$$

and then each strain increment,

$$\Delta\underline{\boldsymbol{\varepsilon}}(t_{i+1}) := \underline{\boldsymbol{\varepsilon}}(t_{i+1}) - \underline{\boldsymbol{\varepsilon}}(t_i),$$

induces a stress increment according to Hooke's law:

$$\Delta\underline{\boldsymbol{\sigma}}_j(t_i) := \underline{\boldsymbol{D}}(t_i - t_j)\Delta\underline{\boldsymbol{\varepsilon}}(t_j) \qquad \text{for } 1 \le j \le i.$$

Notice that each of these stress increments will also relax according to the time dependence of $\underline{\boldsymbol{D}}$. The total stress at time t_i is now given by superposition:

$$\begin{aligned}\underline{\boldsymbol{\sigma}}(t_i) &:= \underline{\boldsymbol{\sigma}}_0(t_i) + \sum_{j=1}^{i} \Delta\underline{\boldsymbol{\sigma}}_j(t_i),\\ &= \underline{\boldsymbol{D}}(t_i)\underline{\boldsymbol{\varepsilon}}(0) + \sum_{j=1}^{i} \underline{\boldsymbol{D}}(t_i - t_j)\Delta\underline{\boldsymbol{\varepsilon}}(t_j),\end{aligned}$$

and by taking an appropriate limit we get the hereditary constitutive law as

$$\underline{\boldsymbol{\sigma}}(\boldsymbol{x}, t) = \underline{\boldsymbol{D}}(t)\underline{\boldsymbol{\varepsilon}}(\boldsymbol{u}(\boldsymbol{x}, 0)) + \int_0^t \underline{\boldsymbol{D}}(t - s)\underline{\boldsymbol{\varepsilon}}(\boldsymbol{u}'(\boldsymbol{x}, s))\, ds. \tag{2.2}$$

Since we are assuming that $\underline{\boldsymbol{D}}(t)$ is smooth we can arrive at an alternate form by partial integration,

$$\underline{\boldsymbol{\sigma}}(\boldsymbol{x}, t) = \underline{\boldsymbol{D}}(0)\underline{\boldsymbol{\varepsilon}}(\boldsymbol{u}(\boldsymbol{x}, t)) - \int_0^t \underline{\boldsymbol{D}}_s(t - s)\underline{\boldsymbol{\varepsilon}}(\boldsymbol{u}(\boldsymbol{x}, s))\, ds, \tag{2.3}$$

where the subscript s indicates partial differentiation with respect to the *history variable* s. Either of these may be used as the constitutive relationship, and each demonstrates clearly the role of memory in viscoelastic modelling.

To get a feel for the form of the time dependence of the stress relaxation tensor $\underline{\boldsymbol{D}}$ we can also quote a perhaps more intuitive method for deriving these constitutive relationships.

We start with the physical observation that viscoelastic materials display the characteristics of both elastic solids and viscous fluids. The kinetics of these type of substances are modelled respectively by the spring and the dashpot.

Fig. 1. A Hookean (linear) spring: $\sigma = E\varepsilon$; E is the spring stiffness

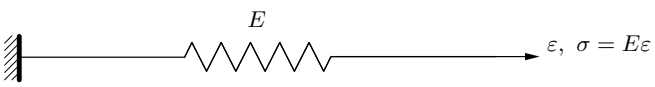
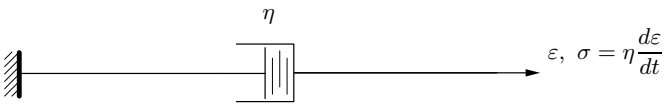


Fig. 2. A Newtonian (linear) dashpot: $\sigma = \eta \dfrac{d\varepsilon}{dt}$; η is the viscosity



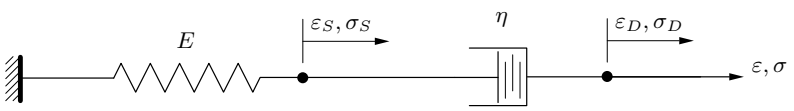
In these models the stress carried by the spring is proportional to the strain in the spring and is given by Hooke's law: $\sigma = E\varepsilon$. The stress carried in the dashpot is proportional to the strain rate and is given by Newton's law of viscosity: $\sigma = \eta\varepsilon'$.

One then models a viscoelastic material by considering a notional system of springs and dashpots with independent stiffness and viscosity parameters. There are essentially two ways to connect a spring to a dashpot: in *series* and in *parallel*. These are the building blocks and are named the "Maxwell" and "Voigt" models.

The Maxwell model

The Maxwell model is a series connection of a spring and dashpot.

Fig. 3. The Maxwell model



In this model ε_S and σ_S denote the strain and stress in the spring alone, and ε_D, σ_D denote those in the dashpot alone. The total stress is given by $\sigma = \sigma_S = \sigma_D$ and the total strain by $\varepsilon = \varepsilon_S + \varepsilon_D$. Differentiating and using Hooke's and Newton's laws yield

$$\frac{d\varepsilon}{dt} = \frac{1}{E}\frac{d\sigma_S}{dt} + \frac{\sigma_D}{\eta} \quad \Longrightarrow \quad \frac{d\sigma}{dt} + \frac{\sigma}{\tau} = E\frac{d\varepsilon}{dt}, \tag{2.4}$$

where $\tau := \eta/E$ is the so-called *relaxation time*. Using $\sigma(0) = E\varepsilon(0)$ this ODE is easily solved to give

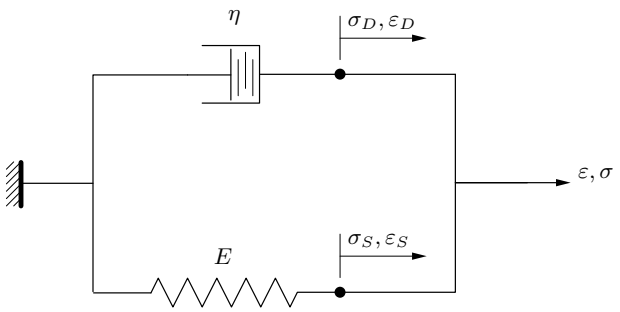
$$\sigma(t) = Ee^{-t/\tau}\varepsilon(0) + E\int_0^t e^{-(t-s)/\tau}\varepsilon'(s)\,ds,$$

and this is essentially (2.2) with the scalar analogue of $\boldsymbol{\underline{D}}$ given by $D(t) = Ee^{-t/\tau}$.

The Voigt model

Connecting the spring and dashpot in parallel yields the Voigt model. This time $\varepsilon_S = \varepsilon_D = \varepsilon$ and equilibrium demands that $\sigma = \sigma_S + \sigma_D$, hence

Fig. 4. The Voigt model



$$\eta\frac{d\varepsilon}{dt} + E\varepsilon = \sigma \quad \Longrightarrow \quad \frac{d\varepsilon}{dt} + \frac{\varepsilon}{\tau} = \frac{\sigma}{\eta}.$$

This gives the constitutive law in hereditary form as

$$\varepsilon(t) = e^{-t/\tau}\varepsilon(0) + \frac{1}{\eta}\int_0^t e^{-(t-s)/\tau}\sigma(s)\,ds.$$

The Maxwell solid

In his internal variable formulation A. Johnson, in for example [20], uses these basic building blocks in the Maxwell solid. Here E_0 and E_1 are spring stiffnesses and σ^*, ε^* are internal stress and strain variables. This time $\sigma^* = E_1\varepsilon^*$, $\varepsilon_D = \varepsilon - \varepsilon^*$ and $\sigma_S = E_0\varepsilon_S$. Also $\sigma^* = \sigma_D$ and this gives

$$E_1\varepsilon^* = \eta\frac{d}{dt}(\varepsilon - \varepsilon^*) \quad \Longrightarrow \quad \frac{d\varepsilon^*}{dt} + \frac{\varepsilon^*}{\tau} = \frac{d\varepsilon}{dt},$$

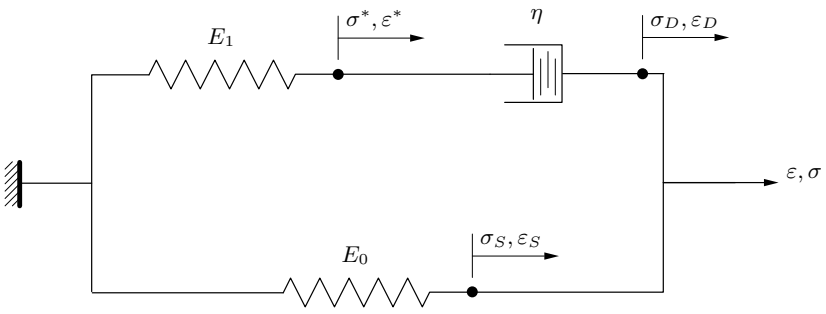
where now $\tau := \eta/E_1$. Solving this we get

$$\varepsilon^*(t) = e^{-t/\tau}\varepsilon(0) + \int_0^t e^{-(t-s)/\tau}\varepsilon'(s)\,ds. \tag{2.5}$$

Now, defining the *stress relaxation function*

$$D(t) := E_0 + E_1 e^{-t/\tau}$$

Fig. 5. The Maxwell solid



as the scalar analogue to the tensor $\underline{\boldsymbol{D}}(t)$ in (2.2) and (2.3), and using this in (2.5) along with the relation

$$\sigma = \sigma_S + \sigma^* = E_0\varepsilon + E_1\varepsilon^* \qquad (\text{since } \varepsilon_S = \varepsilon),$$

gives

$$\begin{aligned}\sigma(t) &= E_0\varepsilon(t) + E_1 e^{-t/\tau}\varepsilon(0) + \int_0^t E_1 e^{-(t-s)/\tau}\varepsilon'(s)\,ds,\\ &= D(0)\varepsilon(t) - \int_0^t D_s(t-s)\varepsilon(s)\,ds.\end{aligned}$$

This is the scalar analogue of equation (2.3) and suggests that we model $\underline{\boldsymbol{D}}$ with the *Dirichlet-Prony series*,

$$\underline{\boldsymbol{D}}(t) = \varphi(t)\underline{\boldsymbol{D}}(0) \tag{2.6}$$

where $\varphi(t)$ is a generic stress relaxation function given by

$$\varphi(t) = \varphi_0 + \sum_{i=1}^{N} \varphi_i e^{-\alpha_i t}. \tag{2.7}$$

Here the (possibly $\boldsymbol{x}$ dependent) coefficients $\{\varphi_i\}_{i=0}^N$ are non-negative and normalized so that $\varphi(0) = 1$, and the (possibly $\boldsymbol{x}$ dependent) $\{\alpha_i\}_{i=1}^N$ are non-negative. More generally one could of course write

$$D_{ijkl}(t) := (D_{ijkl})_0 + \sum_{m=1}^{N_{ijkl}} (D_{ijkl})_m \exp(-(\alpha_{ijkl})_m t).$$

The Dirichlet-Prony series is an extremely convenient form to take for large scale computational approximations to problems (1.1), (1.2) and (1.3) since if

$$\psi(t) := e^{-\alpha t},$$

then one can exploit the simple recurrence

$$\psi(t+k) = e^{-\alpha k}\psi(t)$$

to update the history term arising from a discretization of the Volterra integral. For general Volterra problems one must usually store the entire solution history as the computation advances through the time levels and moreover, at each time level this history needs to be summed to approximate the integral. For such methods the number of operations required at time level N is of the order $O(N^2)$. The Dirichlet-Prony series provides a very useful short cut around this "N^2 problem". (In certain special cases one can also overcome this difficulty using other means, see for example [19,16]).

We now return to the Maxwell solid and generalize the conceptual spring and dashpot model in order to motivate the choice of the Dirichlet-Prony series for the relaxation function as given in (2.7). To begin with we assume again a state of uniaxial stress and strain.

The generalized Maxwell solid, shown in Figure 6, consists of a Hookean spring connected in parallel to a sequence of N spring-dashpot components. In this model

$$\varepsilon_0 = \varepsilon, \qquad \sigma_0 = E_0\varepsilon, \qquad \text{and} \qquad \sigma_i^* = E_i\varepsilon_i^*.$$

Balancing the stresses carried by each of the spring-dashpot pairs we get for each $i \in \{1, \ldots, N\}$ that

$$\begin{aligned} &\frac{d\varepsilon_i^*}{dt} + \frac{\varepsilon_i^*}{\tau_i} = \frac{d\varepsilon}{dt}, \\ \Longrightarrow \quad &\varepsilon_i^*(t) = e^{-t/\tau_i}\varepsilon(0) + \int_0^t e^{-(t-s)/\tau_i}\varepsilon'(s)\,ds, \end{aligned}$$

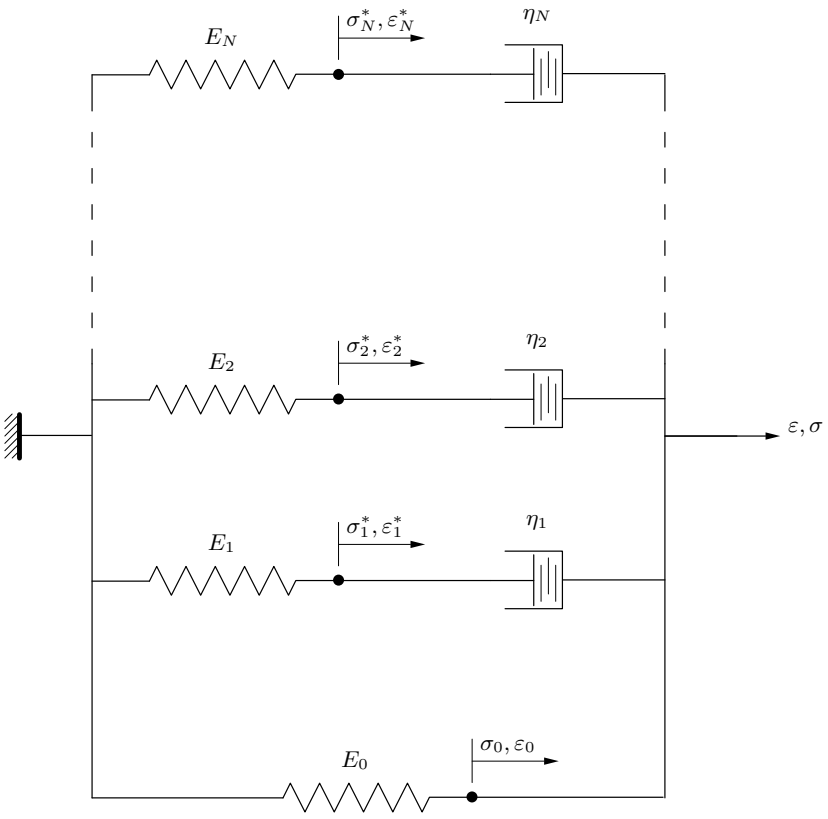
where now we have set $\tau_i := E_i/\eta_i$. The total stress carried by the assemblage is therefore given by:

$$\begin{aligned} \sigma(t) &= \sigma_0(t) + \sigma_1(t) + \cdots + \sigma_N(t), \\ &= E_0\varepsilon(t) + E_1\varepsilon_1^*(t) + \cdots + E_N\varepsilon_N^*(t), \\ &= E_0\varepsilon(0) + E_0(\varepsilon(t) - \varepsilon(0)) \\ &\qquad + \sum_{i=1}^{N}\left(E_i e^{-t/\tau_i}\varepsilon(0) + \int_0^t E_i e^{-(t-s)/\tau_i}\varepsilon'(s)\,ds\right), \\ &= E(t)\varepsilon(0) + \int_0^t E(t-s)\varepsilon'(s)\,ds, \end{aligned} \tag{2.8}$$

where

$$E(t) := E_0 + \sum_{i=1}^{N} E_i e^{-t/\tau_i}.$$

Fig. 6. The generalized Maxwell solid.



The constitutive relationship (2.8) is the scalar analogue of (2.2) with the analogue of $\underline{\boldsymbol{D}}(t-s)$ given by $E(t-s)$, which itself is an example of the Dirichlet-Prony series given in (2.7). Note that if we set $E_0 := 0$ then this generalized Maxwell solid actually models a fluid since $\lim E(t) = 0$.

So much for uniaxial states of stress and strain. In fact it can be shown that for each *relaxation mode* (i.e. each spring-dashpot pair) there is an ODE governing the evolution of each of the internal strain tensor components. Thus we have

$$\frac{d(\varepsilon_{ij})_n^*}{dt} + \frac{(\varepsilon_{ij})_n^*}{\tau_n} = \frac{d\varepsilon_{ij}}{dt},$$

and for the details we refer to [20]. The significance of these internal variable formulations for the viscoelastic constitutive behaviour lies in the fact that it

is possible to solve some kinds of viscoelasticity problems, when the relaxation functions are in the form of a Dirichlet-Prony series (2.7), using only a linear elasticity solver and an ODE solver. This obviates the need to create special software for quasistatic viscoelasticity problems. For more on this we refer again to [20] and also to [33]

The Dirichlet-Prony series is not however the only form used to model the stress relaxation functions, for example the authors of [1] use the *stretched* relaxation function

$$\varphi(t) = \varphi_0 \exp(-(\alpha t)^p) \qquad \text{for } p \in (0, 1]. \tag{2.9}$$

Obviously no simple recurrence exists for this form. Another popular choice for φ is the *power law* where

$$\varphi(t) = \varphi_0 t^{-p} \qquad \text{for } p \in (0, 1), \tag{2.10}$$

although from either of (2.2) or (2.3) this implies that either $\underline{\boldsymbol{\varepsilon}}(0)$ is zero irrespective of the magnitude of the load, or $\underline{\boldsymbol{\sigma}}(0)$ is infinite. Neither of these are physically realistic and so we would prefer to modify this law to

$$\varphi(t) = \varphi_0 (t + \varphi_1)^{-p} \qquad \text{for } p \in (0, 1), \tag{2.11}$$

where $\varphi_1 > 0$ in order to remove the non-physical behaviour. Nonetheless, it is instructive to see how one might "derive" the power law, and for this we borrow heavily from Chern's thesis [3] which exploits the fractional calculus.

The formulation is based on the observed fact that viscoelastic materials behave in a way intermediate to that of solids and fluids. Interpreting this literally yields a constitutive law that contains fractional derivatives. Unfortunately we are unable here to give this interpretation the depth it deserves and instead try only to illustrate the main point. Recall that the stress in a solid is proportional to the strain while the stress in a fluid is proportional to the strain rate. Accepting the intermediate nature of viscoelastic materials the idea is to define the viscoelastic constitutive law as:

$$\underline{\boldsymbol{\sigma}}(t) = \underline{\boldsymbol{D}}^{(0)} \underline{\boldsymbol{\varepsilon}}(t) + \underline{\boldsymbol{D}}^{(1)} \partial_t^\alpha \underline{\boldsymbol{\varepsilon}}(t), \tag{2.12}$$

for constant fourth order tensors $\underline{\boldsymbol{D}}^{(0)}$ and $\underline{\boldsymbol{D}}^{(1)}$, and where $\alpha \in [0, 1)$. The fractional derivative operator may be defined as:

$$\partial_t^\alpha \underline{\boldsymbol{\varepsilon}}(t) := \frac{\partial}{\partial t} \left(\frac{1}{\Gamma(1-\alpha)} \int_0^t (t-s)^{-\alpha} \underline{\boldsymbol{\varepsilon}}(s)\, ds \right), \qquad \text{for } \alpha \in [0, 1). \tag{2.13}$$

(Note that α can be irrational, even though the word "fractional" is always used.) By firstly integrating by parts in (2.13) and then taking the differentiation through, Chern arrives at a constitutive law which is suitable for use within the standard finite element framework. Two solution schemes are considered: a solution in the Laplace transform domain and a direct time domain solution. However, in this case there is no efficient history storage and so the operation

counts and computer memory requirement grow without bound as the time step is diminished.

The "justification" for the power law is as follows. Carrying out this integration-differentiation process gives

$$\partial_t^\alpha \varepsilon(t) = \frac{t^{-\alpha}}{\Gamma(1-\alpha)}\varepsilon(0) + \frac{1}{\Gamma(1-\alpha)}\int_0^t (t-s)^{-\alpha}\varepsilon'(s)\,ds, \tag{2.14}$$

and using this in the scalar analogue of equation (2.12) we now arrive at the constitutive law:

$$\sigma(t) = E_0\varepsilon(t) + \frac{E_1 t^{-\alpha}}{\Gamma(1-\alpha)}\varepsilon(0) + \frac{E_1}{\Gamma(1-\alpha)}\int_0^t (t-s)^{-\alpha}\varepsilon'(s)\,ds. \tag{2.15}$$

This seems to combine (2.2) and (2.3) when $\varphi(t)$ is given by the power law, (2.10).

We now have several candidates for the constitutive law and these may be used to generate a variety of differential equation problems. In the following pages we do just this and demonstrate how concrete forms of the abstract problems (1.1), (1.2) and (1.3), as well as some non-standard variants, can be derived to model viscoelastic behaviour.

3 Viscodynamics

To obtain the governing equations for the dynamic response of a viscoelastic body one uses Newton's second law to relate the stress field $\underline{\boldsymbol{\sigma}}$ and the forces $\boldsymbol{f}$ and $\boldsymbol{g}$ to the acceleration, or *inertia*, of the body $\mathcal{G}$. This process is familiar from linear elasticity theory and gives, with boundary and initial data, the following. For $i = 1, \ldots, n$:

$$\left.\begin{aligned} \varrho u_i'' - \sigma_{ij,j} &= f_i && \text{in } \Omega\times\mathcal{I},\\ u_i &= 0 && \text{in } \Gamma_D\times\mathcal{I},\\ \sigma_{ij}\widehat{n}_j &= g_i && \text{in } \Gamma_N\times\mathcal{I},\\ u_i(\boldsymbol{x},0) &= u_{i0} && \text{in } \Omega,\\ u_i'(\boldsymbol{x},0) &= u_{i1} && \text{in } \Omega. \end{aligned}\right\} \tag{3.1}$$

Here: repeated indices imply summation; ϱ is the mass-density of $\mathcal{G}$; and, $\widehat{\boldsymbol{n}} := (\widehat{n}_i)_{i=1}^n$ is the unit outward directed normal to Γ_N.

Using (2.3) to substitute for the stress one arrives at the PDV problem: find $\boldsymbol{u}$ such that

$$\varrho u_i''(t) - \big(D_{ijkl}(0)\varepsilon_{kl}(\boldsymbol{u}(t))\big)_{,j} = f_i(t) - \int_0^t \left(\frac{\partial D_{ijkl}(t-s)}{\partial s}\varepsilon_{kl}(\boldsymbol{u}(s))\right)_{,j} ds,$$

in $\Omega\times\mathcal{I}$ with the indicated initial-boundary data. With an appropriate definition of A and $B(t,s)$ this is clearly a realization of the abstract problem (1.3). Note

that it is "safe" to use the Dirichlet-Prony series (2.7) or modified power law (2.11) in this problem, but we may not use the power law (2.10) directly because we cannot then interpret $\underline{\boldsymbol{D}}(0)$.

In terms of existence, uniqueness and stability of solutions this problem has been studied in [9,10,24]. Numerical schemes are given in [12,45,29,32].

One could also use the fractional calculus model to substitute for $\underline{\boldsymbol{\sigma}}$ in Newton's second law. This will yield a PDV equation of the form

$$\varrho u_i''(t) - \left(D^{(0)}_{ijkl}\varepsilon_{kl}(\boldsymbol{u}(t))\right)_{,j} = f_i(t) + \frac{D^{(1)}_{ijkl}}{\Gamma(1-\alpha)}\frac{\partial}{\partial t}\int_0^t (t-s)^{-\alpha}\varepsilon_{kl}(\boldsymbol{u}(s))\,ds.$$

On the other hand one could use (2.2) and then arrive at

$$\varrho u_i''(t) - \left(D_{ijkl}(t)\varepsilon_{kl}(\boldsymbol{u}(0))\right)_{,j} = f_i(t) + \int_0^t \left(D_{ijkl}(t-s)\varepsilon_{kl}(\boldsymbol{u}'(s))\right)_{,j}\,ds.$$

Note that $\boldsymbol{u}$ does not occur as a natural "unknown" in this problem and so it is possible to replace $\boldsymbol{u}$ with $\boldsymbol{u}'$ and arrive at the alternative problem: find $\boldsymbol{u}$ such that

$$\varrho u_i'(t) + \int_0^t \left(D_{ijkl}(t-s)\varepsilon_{kl}(\boldsymbol{u}(s))\right)_{,j}\,ds = f_i(t) - \left(D_{ijkl}(t)\varepsilon_{kl}(\boldsymbol{u}_0)\right)_{,j},$$

which makes sense if $\boldsymbol{u}_0$ is smooth enough. The initial datum for this problem is now $\boldsymbol{u}(0) = \boldsymbol{u}_1$. Properties of the solution of these type of problems are studied in [10,24] and numerical analysis is given in [25,23].

However, one must resist the temptation to interpret this as a parabolic problem for, in general, it is not. To see this use the power law (2.10) with (2.6) to obtain (with $\varrho = 1$ and $\underline{\boldsymbol{D}}$ not $\boldsymbol{x}$ dependent for simplicity):

$$u_i'(t) + D_{ijkl}\int_0^t (t-s)^{-p}(\varepsilon_{kl}(\boldsymbol{u}(s))_{,j}\,ds = \widetilde{f}_i(t), \tag{3.2}$$

where $\widetilde{f}$ now incorporates the additional term in $\boldsymbol{u}_0$. In the case $p = \frac{1}{2}$ we find that the operator I defined by,

$$Iw(t) := \frac{1}{\sqrt{\pi}}\int_0^t (t-s)^{-\frac{1}{2}}w(s)\,ds$$

has the property,

$$I^2w(t) \equiv I(Iw)(t) = \int_0^t w(s)\,ds,$$

and so may be regarded as the *square root* of the definite integral operator. Applying $\partial_t^{\frac{1}{2}}$ to both sides of (3.2) in the case $p = \frac{1}{2}$ we arrive at

$$\left(\frac{\partial}{\partial t}\right)^{\frac{3}{2}} u_i(t) + \sqrt{\pi}D_{ijkl}(\varepsilon_{kl}(\boldsymbol{u}(t))_{,j} = \partial_t^{\frac{1}{2}}\widetilde{f}_i(t).$$

This equation is *half way* between being parabolic and hyperbolic. Similar manipulations are also possible in the case $p \neq \frac{1}{2}$, with the final time derivative being of order between 1 and 2. Numerical methods for fractional order differential equations are studied in [31,22,11].

For more detail on these type of problems see [27], as well as the other papers in that collection.

4 Viscostatics

Recall that the classical linear elasticity equations are "derived" from Newton's second law (3.1) by dropping the inertia term $\varrho \boldsymbol{u}''$. This corresponds to modelling the problem at times long after the load has been applied when the transient response has died out, and results in a very well-known elliptic problem. A similar approach can be adopted for viscoelastic response although this time it is a true approximation since the resulting problem is not time independent due to the persistence of the Volterra term. It seems that this approximation can be useful when the inertia term is negligible, which may occur when the load is smoothly and slowly applied (and non-oscillatory), or when it is the long-time *creep* response that is of interest. Since the time dependence persists we refer to the resulting problems as modelling *quasistatic* viscoelastic response.

The governing equations for these type of problem are obtained from (3.1) by setting $\varrho \boldsymbol{u}''(t) := 0$ and discarding the initial data. Thus, for $i = 1, \ldots, n$ we have

$$\left.\begin{aligned} -\sigma_{ij,j} &= f_i && \text{in } \Omega \times \mathcal{J}, \\ u_i &= 0 && \text{in } \Gamma_D \times \mathcal{J}, \\ \sigma_{ij}\widehat{n}_j &= g_i && \text{in } \Gamma_N \times \mathcal{J}, \end{aligned}\right\} \tag{4.1}$$

which are turned into differential equation problems for $\boldsymbol{u}$ by substituting for the stress using either of (2.2) or (2.3). These give respectively the PDV problems: find $\boldsymbol{u}$ such that for each $i \in \{1, \ldots, n\}$,

$$-\int_0^t \big(D_{ijkl}(t-s)\varepsilon_{kl}(\boldsymbol{u}'(s))\big)_{,j}\, ds = f_i(t) + \big(D_{ijkl}(t)\varepsilon_{kl}(\boldsymbol{u}(0))\big)_{,j},$$

and

$$-\big(D_{ijkl}(0)\varepsilon_{kl}(\boldsymbol{u}(t))\big)_{,j} = f_i(t) - \int_0^t \left(\frac{\partial D_{ijkl}(t-s)}{\partial s}\varepsilon_{kl}(\boldsymbol{u}(s))\right)_{,j} ds.$$

The first of these is essentially a Volterra first-kind equation for $\boldsymbol{u}'$, while the second is a second-kind equation for $\boldsymbol{u}$. In both cases one obtains $\boldsymbol{u}(0)$ by solving a linear elasticity problem.

Numerical schemes and *a priori* error estimates were first provided for both of these problems in [35]. Later and for the second-kind problem only, the estimates were improved (in terms of the size of the error constant) in [34]. These results

depend on by-passing Gronwall's inequality and using more sensitive comparison results to obtain sharp data-stability estimates. These estimates have now been generalized in [40]. Also for the second-kind problem, *a posteriori* error estimates for a space-time finite element discretization of a model problem have been given in [38] and [42]. These results are based on the error estimates in [37] and are currently being generalized to the multidimensional problem described above in [41].

5 Non-Fickian diffusion

In classical diffusion theory the gradient of the concentration u of an active agent (the penetrant) diffusing through a carrier medium is related to the mass flux by Fick's law: $\boldsymbol{J} = -\lambda \nabla u$, where λ is the diffusivity of the carrier substance. Conservation of mass then demands that $u' = -\nabla \cdot \boldsymbol{J}$ which yields the familiar heat equation,

$$u'(t) = \nabla \cdot \lambda \nabla u.$$

If we define $M(t)$ as the total mass of penetrant absorbed by the carrier per unit area at time t then it is well known (from similarity solutions) that $M(t) \sim t^{\frac{1}{2}}$ for Fickian diffusion.

Diffusion in rubbery polymers, those well above their glass transition temperature (GTT), is according to Durning in [13] adequately described by Fick's law, but the situation is much more complicated for glassy polymers, those near but above their GTT. As the penetrant moves through the polymer it can force a phase change and so leave behind it the polymer carrier in its rubbery state. The stiffness and relaxation properties of the polymer change abruptly by orders of magnitude across this phase change (see for example [15]), and as a result a differential stress is set up across the penetrant boundary. Moreover, because the carrier is viscoelastic this stress is described by a hereditary constitutive law and this behaviour provides a mechanism for the observed non-Fickian effects. Workers in the field make the following very rough classification.

Case I diffusion: standard Fickian diffusion where $M(t) \sim t^{\frac{1}{2}}$, applies to polymers in the rubbery state high above the GTT.

Case II diffusion: non-Fickian diffusion, $M(t) \sim t^{\alpha}$ where $\frac{1}{2} < \alpha \leq 1$, applies to glassy polymers near to but above the GTT.

There is also a "Super Case II" category corresponding to $\alpha > 1$, see [5]. For Case II sharp fronts (rather like shocks) may appear as the penetrant diffuses through the carrier. This front moves initially at a constant speed and then slows down, [7], and this explains why $M(t)$ is almost linear, and thus $M'(t)$—the rate of absorbtion—is almost constant. By contrast $M'(t)$ for Case I is, in the words of Cox in [7], "delta-function-like", and this property of glassy polymers has an interesting application in the area of controlled drug delivery products. Cox gives a nice example.

An active agent (the drug) is embedded into a polymer through which it cannot diffuse. This may for example be a tablet which is to be swallowed. When the carrier is invaded by a solvent, such as digestive fluid, the drug can then diffuse out of the polymer through the solvent in a non-Fickian way. Since $M'(t)$ is almost constant, this allows a controlled, constant-rate delivery of the drug to the body for several hours.

The polymer doesn't have to be a tablet. In fact, according to Cohen and White in [5] (who also describe other applications of non-Fickian diffusion), such "smart" pharmaceutical products can be designed to be "swallowed, smelled, surgically implanted, rubbed on, taped on, strapped on", and can in effect be applied to any part of the body. There is an extensive literature on this science and in addition to those already cited we refer also to [17,6,14].

To get a flavour of the mathematical modelling that these people employ we borrow from [4] and consider the modelling of one-space dimensional diffusion through a glassy polymer. Our development yields a linear model, but it is unlikely that this will reproduce the sharp fronts characteristic of polymer diffusion. The references cited deal with realistic nonlinear models.

To account for the differential stress set up at the penetrant front Fick's law is modified to include a stress dependence in the following way:

$$J = -(\lambda u_x + \kappa \sigma_x).$$

Here u is the concentration, λ the usual (Fickian) diffusion constant, and κ is a proportionality constant. Conservation of mass again demands that $u' = -J_x$ and this gives

$$u' = \lambda u_{xx} + \kappa \sigma_{xx}.$$

The stress is viscoelastic and the usual approach is to adopt the Maxwell model, given earlier in (2.4), with the assumption that u depends linearly on strain rate ε' (in order to get true Case II behaviour—see [8]). Thus

$$\frac{\partial \sigma}{\partial t} + \frac{\sigma}{\tau} = \mu u,$$

where μ is a proportionality constant. In the nonlinear theory the dependence of τ on u is crucial, but here we shall assume that τ is constant. Integrating we get

$$\sigma(t) = \mu e^{-t/\tau} u(0) + \mu \int_0^t e^{-(t-s)/\tau} u(s)\, ds.$$

Eliminating the stress from the transport equation and using mass conservation gives the single differential-Volterra equation,

$$u'(t) = \lambda u_{xx} + \kappa \mu e^{-t/\tau} u_{xx}(0) + \kappa \mu \int_0^t e^{-(t-s)/\tau} u_{xx}(s)\, ds.$$

Assuming for simplicity that $u(0) = 0$ we can generalize this to a multidimensional model and obtain the PDV equation,

$$u'(t) = \nabla \cdot \lambda \nabla u + \nabla \cdot \left(\kappa \nabla \int_0^t \mu e^{-(t-s)/\tau} u(s)\, ds \right).$$

This is a concrete realization of the abstract problem (1.2).

Equations of this nature have been studied in [26] and [18], and some numerical analysis is given in [45,2,43,30,44]. Also, *a priori* and *a posteriori* error estimates for a finite element discretization of a scalar prototype ODE with memory, of the type that arises after spatial finite element semi-discretization of this problem, are provided in [36].

References

1. J. T. Bendler, B. Noble, and M. A. Hussain. Solution of an equation for creep in solid polymers. In *Proc. Computers in Engineering*, number 3 in 1, pages 365—368, 1988.
2. J. R. Cannon and Y. Lin. *A priori* L^2 error estimates for finite-element methods for nonlinear diffusion equations with memory. *SIAM J. Numer. Anal.*, 27:595—607, 1990.
3. J. T. Chern. *Finite element modeling of viscoelastic materials on the theory of fractional calculus.* PhD thesis, Penn. State Uni., USA, 1993.
4. D. S. Cohen and A. B. White Jr. Sharp fronts due to diffusion and stress at the glass transition in polymers. *J. Polymer Sci. B: Polymer Physics*, 27:1731—1747, 1989.
5. D. S. Cohen and A. B. White Jr. Sharp fronts due to diffusion and viscoelastic relaxation in polymers. *SIAM J. Appl. Math.*, 51:472—483, 1991.
6. D. S. Cohen, A. B. White Jr., and T. P. Witelski. Shock formation in a multidimensional viscoelastic diffusive system. *SIAM J. Appl. Math.*, 55:348—368, 1995.
7. R. W. Cox. Shocks in a model for stress-driven diffusion. *SIAM J. Appl. Math.*, 50:1284—1299, 1990.
8. R. W. Cox and D. S. Cohen. A mathematical model for stress driven diffusion in polymers. *J. Polymer Sci. B: Polymer Physics*, 27:589—602, 1989.
9. C. M. Dafermos. An abstract Volterra equation with applications to linear viscoelasticity. *J. Diff. Eqns.*, 7:554—569, 1970.
10. C. M. Dafermos and J. A. Nohel. Energy methods for nonlinear hyperbolic Volterra type equations. *Comm. Part. Diff. Eqns.*, 4:219—278, 1970.
11. Kai Diethelm. An algorithm for the numerical solution of differential equations of fractional order. *Electronic Transactions on Numerical Analysis*, 5:1—6, 1997.
12. J. Douglas and B. F. Jones. Numerical methods for integro-differential equations of parabolic and hyperbolic types. *Numer. Math.*, 4:96—102, 1962.
13. C. J. Durning. Differential sorption in viscoelastic fluids. *J. Polymer Sci. B: Polymer Physics*, 23:1831—1855, 1985.
14. David A. Edwards. Constant front speed in weakly diffusive non-Fickian systems. *SIAM J. Appl. Math.*, 55:1039—1058, 1995.
15. J. D. Ferry. *Viscoelastic properties of polymers.* John Wiley and Sons Inc., 1970.

16. E. Hairer, CH. Lubich, and M. Schlichte. Fast numerical solution of nonlinear Volterra convolution equations. *SIAM J. Sci. Stat. Comput.*, 6:532—541, 1985.
17. C. K. Hayes and D. S. Cohen. The evolution of steep fronts in non-Fickian polymer-penetrant systems. *J. Polymer Sci. B: Polymer Physics*, 30:145—161, 1992.
18. Melvin L. Heard. An abstract parabolic Volterra integrodifferential equation. *SIAM J. Math. Anal.*, 13:81—105, 1982.
19. V. Janovsky, Simon Shaw, M. K. Warby, and J. R. Whiteman. Numerical methods for treating problems of viscoelastic isotropic solid deformation. *J. Comput. Appl. Math.*, 63:91—107, 1995.
20. A. R. Johnson and A. Tessler. A viscoelastic high order beam finite element. In J. R. Whiteman, editor, *The Mathematics of Finite Elements and Applications.* MAFELAP *1996*, pages 333—345. Wiley, Chichester, 1997.
21. F. J. Lockett. *Nonlinear viscoelastic solids.* Academic Press, 1972.
22. J. C. López-Marcos. A difference scheme for a nonlinear partial integrodifferential equation. *SIAM J. Numer. Anal.*, 27:20—31, 1990.
23. CH. Lubich, I. H. Sloan, and V. Thomée. Nonsmooth data error estimates for approximations of an evolution equation with a positive-type memory term. *Math. Comp.*, 65:1—17, 1996.
24. R. C. MacCamy. A model for one-dimensional, nonlinear viscoelasticity. *Q. Appl. Math.*, 35:21—33, 1977.
25. W. Mclean and V. Thomée. Numerical solution of an evolution equation with a positive-type memory term. *J. Austral. Math. Soc.*, 35:23—70, 1993.
26. R. K. Miller. An integrodifferential equation for rigid heat conductors with memory. *J. Math. Anal. Appl.*, 66:313—332, 1978.
27. J. A. Nohel. A nonlinear hyperbolic Volterra equation. In *Volterra equations*, volume 737 of *Lecture Notes in Mathematics*, pages 220—235. Spinger-Verlag, 1979.
28. J. W. Nunziato. On heat conduction in materials with memory. *Quart. Appl. Maths.*, 29:187—204, 1971.
29. A. K. Pani, V. Thomée, and L. B. Wahlbin. Numerical methods for hyperbolic and parabolic integro-differential equations. *J. Integral Equations Appl.*, 4:533—584, 1992.
30. Amiya K. Pani and Todd E. Peterson. Finite element methods with numerical quadrature for parabolic integrodifferential equations. *SIAM J. Numer. Anal.*, 33:1084—1105, 1996.
31. J. M. Sanz-Serna. A numerical method for a partial integro-differential equation. *SIAM J. Numer. Anal.*, 25:319—327, 1988.
32. Simon Shaw, M. K. Warby, and J. R. Whiteman. An error bound via the Ritz-Volterra projection for a fully discrete approximation to a hyperbolic integrodifferential equation. Technical report, 94/3, BICOM, Brunel University, Uxbridge, U.K., 1994. (http://www.brunel.ac.uk/~icsrbicm).
33. Simon Shaw, M. K. Warby, and J. R. Whiteman. A comparison of hereditary integral and internal variable approaches to numerical linear solid viscoelasticity. In *Proceedings of the XIII Polish Conference on Computer Methods in Mechanics*, 1997. Poznan, May 1997
(BICOM Tech. Rep. 97/2, see http://www.brunel.ac.uk/~icsrbicm).
34. Simon Shaw, M. K. Warby, and J. R. Whiteman. Error estimates with sharp constants for a fading memory Volterra problem in linear solid viscoelasticity. *SIAM J. Numer. Anal.*, 34:1237—1254, 1997.
(See also, http://www.brunel.ac.uk/~icsrbicm).

35. Simon Shaw, M. K. Warby, J. R. Whiteman, C. Dawson, and M. F. Wheeler. Numerical techniques for the treatment of quasistatic viscoelastic stress problems in linear isotropic solids. *Comput. Methods Appl. Mech. Engrg.*, 118:211—237, 1994.
36. Simon Shaw and J. R. Whiteman. Backward Euler and Crank-Nicolson finite element variants with rational adaptivity and *a posteriori* error estimates for an integrodifferential equation. Submitted to Math. Comp. 1996. (see http://www.brunel.ac.uk/~icsrbicm)
37. Simon Shaw and J. R. Whiteman. Discontinuous Galerkin method with *a posteriori* $L_p(0, t_i)$ error estimate for second-kind Volterra problems. *Numer. Math.*, 74:361—383, 1996.
38. Simon Shaw and J. R. Whiteman. Towards adaptive finite element schemes for partial differential Volterra equation solvers. *Advances in Computational Mathematics*, 6:309—323, 1996.
39. Simon Shaw and J. R. Whiteman. Applications and numerical analysis of partial differential Volterra equations: a brief survey. *Comput. Methods Appl. Mech. Engrg.*, 150:397—409, 1997.
40. Simon Shaw and J. R. Whiteman. Optimal long-time $L_p(0, T)$ data stability estimates for the Volterra formulation of the linear quasistatic viscoelasticity problem. BICOM Tech. Rep. 97/6, submitted to SIAM J. Appl. Math. Also, see: http://www.brunel.ac.uk/~icsrbicm, 1997.
41. Simon Shaw and J. R. Whiteman. Space-time finite element method with *a posteriori* Galerkin energy-error estimate for linear quasistatic viscoelasticity problems. BICOM Technical Report 97/7, see http://www.brunel.ac.uk/~icsrbicm, 1997.
42. Simon Shaw and J. R. Whiteman. Towards robust adaptive finite element methods for partial differential Volterra equation problems arising in viscoelasticity theory. In J. R. Whiteman, editor, *The Mathematics of Finite Elements and Applications.* MAFELAP *1996*, pages 55—80. Wiley, Chichester, 1997.
43. I. H. Sloan and V. Thomée. Time discretization of an integro-differential equation of parabolic type. *SIAM J. Numer. Anal.*, 23:1052—1061, 1986.
44. V. Thomée and L. B. Wahlbin. Long-time numerical solution of a parabolic equation with memory. *Math. Comp.*, 62:477—496, 1994.
45. E. G. Yanik and G. Fairweather. Finite element methods for parabolic and hyperbolic partial integro-differential equations. *Nonlinear Analysis, Theory, Methods & Applications*, 12:785—809, 1988.

EQUADIFF 9 CD ROM, Brno 1997
Masaryk University

PROCEEDINGS
pp. 201–251

The Use of Semiregular Finite Elements

Alexander Ženíšek

Department of Mathematics, Technical University, Technická 2
616 69 Brno, Czech Republic
Email: zenisek@kinf.fme.vutbr.cz
zenisek@mat.fme.vutbr.cz

Abstract. This text is extended Equadiff 9 plenary lecture. Sections 1–4 contain a survey of published results which concern triangular and quadrilateral finite elements. Sections 1 and 2 are devoted to interpolation problems. These two sections contain also results of other authors. The analysis of both the effect of numerical integration and approximation of a boundary is restricted to triangular elements with linear polynomials and to quadrilateral elements with four-node isoparametric functions. The corresponding results in the case of smooth solutions are introduced in Section 3, where the rate of convergence $O(h)$ is proved; the case of nonsmooth solutions is studied in Section 5. This section is restricted to triangular elements. In Sections 3 and 5 the domain considered has a form of a narrow ring with a great diameter. In this case the elements cannot be arbitrarily narrow. In Section 4 a composite domain indicated in Fig. 5 is approximated by triangular elements and applications of the finite element method in magnetostatical problems are introduced. In this case the triangular elements can be arbitrarily narrow. Section 6 is an Appendix where a special form of a discrete Friedrichs' inequality, suitable for semiregular elements, is proved. Sections 5 and 6, which complete the survey introduced in Sections 1–4, have not yet been published and were written specially for Equadiff 9.
The notation of derivatives and Sobolev spaces is identical with the notation used in [9].
As to the notion of semiregular elements, semiregular triangles can have one angle arbitrarily small. Triangles with two arbitrarily small angles are irregular. A semiregular quadrilateral K can be arbitrarily narrow and it satisfies the condition

$$|\cos \vartheta_i| \le \sigma < 1 \quad (i = 1, \dots, 4),$$

where $\vartheta_1, \dots, \vartheta_4$ are the angles of K.

AMS Subject Classification. 65N30

Keywords. Finite element method, elliptic problems, semiregular elements, maximum angle condition, effect of numerical integration, approximation of the boundary, magnetostatical problems, discrete Friedrichs' inequality

1 Triangular and quadrilateral elements of the Lagrange type

First interpolation estimates which can be used in the finite element theory were derived by Synge in the year 1957 (see [14, pp. 209–213]). His a little improved result can be formulated in the following theorem:

Theorem 1.1. *Let u be a function continuous on a closed triangle $\overline{T}$ with bounded second partial derivatives in its interior T,*

$$\left|\frac{\partial^2 u}{\partial x_i \partial x_j}\right| \le M_2,$$

and let $p(x_1, x_2)$ be a linear polynomial satisfying

$$p(P_i) = u(P_i) \quad (i = 1, 2, 3)$$

with P_1, P_2, P_3 the vertices of $\overline{T}$. Then it holds on $\overline{T}$

$$\left|\frac{\partial u}{\partial x_i} - \frac{\partial p}{\partial x_i}\right| \le \frac{2M_2 h}{\cos(\gamma/2)} \quad (i = 1, 2) \tag{1.1}$$

$$|u - p| \le \frac{2M_2 h^2}{\cos(\gamma/2)} \tag{1.2}$$

Result (1.2) was obtained by means of (1.1). Another independent consideration (where we first estimate the difference $g = u - p$ on P_2P_3 and then on P_1P' with $P' \in P_2P_3$ an arbitrary point) gives us

$$|u - p| \le \frac{1}{2} M_2 h^2. \tag{1.3}$$

This result implies a question whether estimate (1.1) cannot be improved, as far as the geometry is concerned. An example showing that the answer is negative was presented in [15]. Here is its simplified version: Let us consider a set of triangles with vertices

$$P_1(-h/2, 0), P_2(h/2, 0), P_3(0, y_0),$$

where h is fixed and y_0 $(0 < y_0 < \sqrt{3}h/2)$ is variable, and a function $u(x_1, x_2) = x_1^2$. Its first degree interpolant has the form

$$p(x_1, x_2) = \frac{h^2}{4}\left(1 - \frac{x_2}{y_0}\right).$$

Hence

$$\left|\frac{\partial u}{\partial x_2} - \frac{\partial p}{\partial x_2}\right| = \left|\frac{\partial p}{\partial x_2}\right| = \frac{h^2}{4y_0} = \frac{h}{2}\cot\alpha = \frac{h}{2}\tan(\gamma/2), \tag{1.4}$$

where α and γ are the minimum and maximum angles of $\overline{T}$, respectively. If $y_0 \to 0$ then $\alpha \to 0$, $\gamma \to \pi$ and

$$\left| \frac{\partial u}{\partial x_2} - \frac{\partial p}{\partial x_2} \right| \to \infty.$$

Zlámal knew both estimate (1.1) and result (1.4) when he started to work on his paper "On the finite element method" (see [24]). Nevertheless, instead of the maximum angle condition

$$\gamma_T \leq \gamma_0 < \pi \quad \forall T \in \mathcal{T}_h, \quad \forall h \in (0, h_0) \tag{1.5}$$

where $\mathcal{T}_h$ denotes a triangulation of a given (polygonal) domain, he introduced the minimum angle condition

$$\vartheta_T \geq \vartheta_0 > 0 \quad \forall T \in \mathcal{T}_h, \quad \forall h \in (0, h_0) \tag{1.6}$$

where ϑ_T is the minimum angle of T. Reading Zlámal's papers one sees that the finite element theory is relatively easy under condition (1.6). Also other mathematicians started to use condition (1.6) and when it was used in Ciarlet's 1978-book [3] it has become a standard finite element condition.

However, there are situations where the minimum angle condition (1.6) is too restrictive because it forbids to use triangles with one small angle. Such triangles are permitted according to the maximum angle condition. Thus it is quite natural to try to generalize the standard finite element theory to the case of condition (1.5).

We start with the interpolation theorems and first we remind Jamet's result [5].

For a better understanding we introduce from [5] only a special situation which is for applications quite sufficient. Let $\mathcal{L}(X, Y)$ denote the set of all linear bounded operators from a normed space X into a normed space Y. Let

$$\Pi \in \mathcal{L}(W^{k,p}(T), W^{1,p}(T)),$$

where k is a positive integer and $p \in [1, \infty]$, be an operator satisfying the following hypotheses:

(H.1) We have

$$\Pi u = u \quad \forall u \in P_{k,2}\,,$$

where $P_{k,n}$ denotes the set of all polynomials in n variables of degree not greater than k.

(H.2) There exists a unit vector ξ such that

$$\frac{\partial u}{\partial \xi}(P) = 0 \quad \forall P \in T \quad \Rightarrow \quad \frac{\partial (\Pi u)}{\partial \xi}(P) = 0 \quad \forall P \in T.$$

(We restrict ourselves to this special type of (H.2) because we are interested only in estimates of type (1.7).)

Theorem 1.2. *Let $\overline{T}$ be a closed triangle with the interior T and vertices P_1, P_2, P_3 and let α_T, β_T and γ_T be the angles at P_1, P_2 and P_3, respectively. Let the vertices be denoted in such a way that $\alpha_T \le \beta_T \le \gamma_T$. Let s_1 and s_2 be the unit vectors parallel to the sides P_3P_2 and P_3P_1, respectively. Let $\Pi \in \mathcal{L}(W^{k,p}(T), W^{1,p}(T))$ be an operator satisfying hypotheses (H.1) and (H.2) for $\xi = s_1$ and $\xi = s_2$. Let $u \in W^{k+1,p}(T)$. Then we have for $m = 0$ and $m = 1$*

$$|u - \Pi u|_{m,p,T} \le C \frac{h_T^{k+1-m}}{(\cos(\gamma_T/2))^m} |u|_{k+1,p,T}\,, \tag{1.7}$$

where $h_T = \operatorname{dist}(P_1, P_2)$ and C is a constant not depending on u and T.

Proof. The assertion is a special case of [5, Theorem 2.2]. □

In [5] Theorem 1.2 is applied on compatible triangular finite elements of the Lagrange type for arbitrary k. (For $k = 1$, $p = \infty$ estimates (1.7) are identical with Synge's result.) This means that the operator Π is defined by the relations

$$(\Pi u)(P_i) = u(P_i) \quad (i = 1, \dots, N,\ N := (n+1)(n+2)/2),$$

where $P_1, \dots, P_N$ are the nodal points which are situated on $\overline{T}$ as the first N integers in the Pascal triangle (see Fig. 1 where the black circles denote prescribed function values).

However, in the case $k = 1$ estimates (1.7) hold only for $p \in (2, \infty]$. The important case $p = 2$ is treated in [2] for $k \ge 1$. A further generalization in the case $k = 1$ is given in [6]. The interpolation result proved in [6] can be formulated as follows.

Theorem 1.3. *Let $\overline{T}$ be the same triangle as in Theorem 1.2 and let $p \in (1, \infty)$. Let $u \in W^{2,p}(T)$ and let $I_h u$ be the linear function satisfying $(I_h u)(P_i) = u(P_i)$ $(i = 1, 2, 3)$. Then we have*

$$|u - I_h u|_{m,p,T} \le C \frac{h_T^{2-m}}{(\sin \gamma_T)^m} |u|_{2,p,T} \quad (m = 0, 1), \tag{1.8}$$

where C is a constant independent of u and T.

Theorem 1.3 will be useful in our further considerations.

Now we introduce interpolation results in the case of semiregular (i.e., narrow) convex four-node quadrilateral isoparametric finite elements. In [1] such elements are called anisotropic. However, in [1] the error of the interpolation is estimated on rectangular elements; quadrilaterals are not considered.

The symbol $\overline{K}_0$ will denote the closed square in the (ξ, η)-plane with vertices $\widehat{M}_1(1,0)$, $\widehat{M}_2(1,1)$, $\widehat{M}_3(0,1)$, $\widehat{M}_4(0,0)$. The functions $\widehat{\varphi}^{(i)} : (\xi, \eta) \to R^1$ with

$$\begin{aligned} &\widehat{\varphi}^{(1)}(\xi,\eta) = \xi(1-\eta),\ \widehat{\varphi}^{(2)}(\xi,\eta) = \xi\eta, \\ &\widehat{\varphi}^{(3)}(\xi,\eta) = (1-\xi)\eta,\ \widehat{\varphi}^{(4)}(\xi,\eta) = (1-\xi)(1-\eta) \end{aligned} \tag{1.9}$$

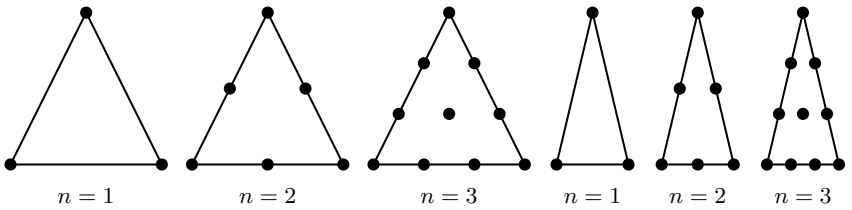


Fig. 1. Triangular finite elements of the Lagrange type.

are called bilinear basis functions; they have the property

$$\widehat{\varphi}^{(i)}(\widehat{M}_j) = \delta_{ij}.$$

Let $\overline{K}$ be a closed convex quadrilateral in the (x, y)-plane. Let two sides of $\overline{K}$ be much greater than the remaining two ones. Let us consider first the case that these two longer sides are parallel. (Such quadrilaterals are important, for example, in modelling a gap between rotor and stator in an electrical machine.) Let α_K be the smallest angle of $\overline{K}$ and let us denote by M_1 the vertex of $\overline{K}$ at the angle α_K. (If $\overline{K}$ has two or four angles which can be denoted by α_K then, of course, we have two or four choices.) One short side and one long side of $\overline{K}$ meet at M_1. The second end-point of the long one will be denoted by M_2 and the second end-point of the short one by M_4. The numbering of the vertices of $\overline{K}$ is thus either anticlockwise, or clockwise.

In applications the local numbering of the vertices of $\overline{K}$ obeys a different rule which is usually anticlockwise; let $N_1, \ldots, N_4$ denote the vertices of $\overline{K}$ according to this different rule, let (for simplicity) the numbering of $M_1, \ldots, M_4$ be also anticlockwise and let

$$M_1 = N_{j+1},\ M_2 = N_{j+2},\ M_3 = N_{j+3},\ M_4 = N_j,$$

where $N_{j+i} \equiv N_{j+i-4}$ if $j + i \geq 5$. As N_i corresponds by definition to $\widehat{M}_i$ the isoparametric transformation of $\overline{K}_0$ onto $\overline{K}$ has the form

$$\begin{aligned} x = x_K(\xi, \eta) &:= \sum_{i=1}^{4} x_i \widehat{\varphi}^{(j+i)}(\xi, \eta), \\ y = y_K(\xi, \eta) &:= \sum_{i=1}^{4} y_i \widehat{\varphi}^{(j+i)}(\xi, \eta), \end{aligned} \tag{1.10}$$

where x_i, y_i are the coordinates of M_i $(i = 1, \ldots, 4)$ and where the indices $j + i$ $(0 \leq j \leq 3$ fixed, $i = 1, \ldots, 4)$ are considered modulo 4. (In the case when the numbering of $M_1, \ldots, M_4$ is clockwise the corresponding isoparametric transformation of $\overline{K}_0$ onto $\overline{K}$ has again the form of (1.10).) As $\overline{K}$ is convex, transformation (1.10) maps $\overline{K}_0$ one-to-one onto $\overline{K}$.

Let

$$\xi = \xi_K(x,y),\ \ \eta = \eta_K(x,y) \tag{1.11}$$

denote the inverse transformation to transformation (1.10). We set

$$\varphi^{(i)}(x,y) := \widehat{\varphi}^{(i)}(\xi_K(x,y), y_K(x,y)) \quad (i = 1,\dots,4). \tag{1.12}$$

If $u \in C(\overline{K})$, then we define the isoparametric interpolation of u on $\overline{K}$ by

$$(Qu)(x,y) = \sum_{i=1}^{4} u(M_i)\varphi^{(j+i)}(x,y). \tag{1.13}$$

Theorem 1.4. *Let $\overline{K}$ be a narrow quadrilateral with parallel long sides which satisfy the assumption*

$$\operatorname{dist}(M_1,M_4) \le \frac{1}{12}\operatorname{dist}(M_1,M_2). \tag{1.14}$$

Let $u \in H^2(K)$. Then we have

$$\|u - Qu\|_{0,K} \le \left(C_1 + \frac{C_2\varepsilon_K}{h_K \sin\beta_K}\right) h_K^2 |u|_{2,K}, \tag{1.15}$$

$$|u - Qu|_{1,K} \le \left(C_3 + \frac{C_4}{\sin\alpha_K}\right) \frac{h_K}{\sin\beta_K} |u|_{2,K}, \tag{1.16}$$

where Qu is defined in (1.13), ε_K=dist $(M_1,M_4) < h_K$=dist (M_1,M_2), $\alpha_K \le \beta_K$, α_K and β_K being the angles at M_1 and M_2, respectively, and the constants C_1, C_2, C_3, C_4 satisfy

$$C_1 = 55.019093,\ C_2 = 21.658241,\ C_3 = 12.801823,\ C_4 = 19.47235264.$$

For the proof see [22].

Remark 1.5. Using the more standard approach with the bilinear isoparametric mapping of K_0 onto K we obtain (by means of the sharp form of the Bramble-Hilbert lemma) the estimate $\|u - Qu\|_{0,K} \le Ch_K^2|u|_{2,K}$ which does not depend on the geometry of K. However, this approach completely fails in estimating $|u - Qu|_{1,K}$ where we loose all powers of h_K.

Remark 1.6. It can be shown by an example that the dependence of the estimate of $|u - Qu|_{1,K}$ on $\sin^{-1}\alpha_K$ is essential (see [23]). The dependence on $\sin^{-1}\beta_K$ in both (1.15) and (1.16) is a cosmetic defect which is a consequence of the approach used in [22].

Remark 1.7. If we change assumption (1.14) to

$$\operatorname{dist}(M_1,M_4) \le \frac{1}{2n}\operatorname{dist}(M_1,M_2), \quad n \ge 6,$$

then the numerical constants in Theorem 1.4 will be smaller. (In more detail see [22].)

Theorem 1.4 can be generalized to the case that the long sides are not parallel. We again assume that $\overline{K}$ is a convex quadrilateral. Moreover, we assume that the long sides do not have any common vertex.

Our considerations are based on the following simple fact: Let $\overline{K}$ be an arbitrary convex quadrilateral. Then there exists a parallelogram $\overline{D}$ which has three vertices common with $\overline{K}$ and is such that $\overline{K} \subset \overline{D}$.

Let us denote these three vertices by M_1, M_2, M_3 in such a way that M_1M_2 and M_2M_3 are sides of K with the property

$$\operatorname{dist}(M_2, M_3) < \operatorname{dist}(M_1, M_2). \tag{1.17}$$

We shall denote

$$h_K := \operatorname{dist}(M_1, M_2), \quad a_K := \operatorname{dist}(M_2, M_3). \tag{1.18}$$

Of course it may happen that h_K is not the length of the greatest side of $\overline{K}$ and that the numbering of M_1, M_2, M_3, M_4 is not anticlockwise.

We shall assume that

$$a_K \le \frac{1}{2n} h_K, \quad \varepsilon_K \le \frac{1}{2n} h_K, \tag{1.19}$$

$$\frac{1}{2} \le \frac{\operatorname{dist}(M_4, p)}{\operatorname{dist}(M_3, p)} \le 1, \tag{1.20}$$

where $n \ge 6$ is a given integer, $\varepsilon_K := \operatorname{dist}(M_1, M_4)$ and p denotes the straight-line passing through M_1 and M_2.

In applications we usually have

$$\frac{\pi}{4} \le \alpha_K \le \frac{3\pi}{4}, \quad \frac{\pi}{4} \le \beta_K \le \frac{3\pi}{4}.$$

The interpolation theorem has in this more general case the following form (see [22]).

Theorem 1.8. *Let $\overline{K}$ be a quadrilateral satisfying assumptions (1.17)–(1.20) and let $u \in H^2(K)$. Then we have*

$$\|u - Qu\|_{0,K} \le \left(\widehat{C}_1(n) + \frac{\widehat{C}_2(n) \sqrt{\varepsilon_K a_K}}{h_K \sqrt{\sin \beta_K \sin \alpha_K}} \right) h_K^2 |u|_{2,K}, \tag{1.21}$$

$$|u - Qu|_{1,K} \le \left(\widehat{C}_3(n) + \frac{\widehat{C}_4(n) \sqrt{\varepsilon_K}}{\sqrt{a_K \sin \beta_K \sin \alpha_K}} \right) \frac{h_K}{\sin \beta_K} |u|_{2,K}, \tag{1.22}$$

where Q is an interpolation operator of type (1.13), $a_K = \operatorname{dist}(M_2, M_3)$ and $\varepsilon_K = \operatorname{dist}(M_1, M_4)$ satisfy (1.19), α_K and β_K are the angles at M_1 and M_2, respectively, and the positive constants $\widehat{C}_1(n)$, $\widehat{C}_2(n)$, $\widehat{C}_3(n)$ and $\widehat{C}_4(n)$ are decreasing when n is increasing, n being the integer which appears in (1.19).

2 Triangular elements of the Hermite type

Let us define $\Pi u \in P_{3,2}$, where $u \in C^1(\overline{T})$ and $\overline{T}$ is the same as in Theorem 1.2, by the relations

$$\begin{gathered}(D^\alpha \Pi u)(P_i) = D^\alpha u(P_i) \quad |\alpha| \le 1 \quad (i = 1,2,3), \\ \frac{\partial(\Pi u)}{\partial s_2}(Q_1) = \frac{\partial u}{\partial s_2}(Q_1),\end{gathered} \tag{2.1}$$

where Q_1 is the mid-point of the side P_2P_3.

Theorem 2.1. *The polynomial Πu is uniquely determined by relations (2.1). We have*

$$\Pi \in \mathcal{L}(W^{3,p}(T), W^{1,p}(T)), \ p \in [1,\infty]$$

and the operator Π satisfies hypotheses (H.1) and (H.2) for $\xi = s_1$ and $\xi = s_2$. Hence estimates (1.7) hold for $k = 3$, $p \in [1,\infty]$ and $m = 0, 1$:

$$|u - \Pi u|_{m,p,T} \le C \frac{h_T^{4-m}}{(\cos(\gamma_T/2))^m} |u|_{4,p,T}.$$

Proof. The unique determination will be proved in Remark 2.10. The property $\Pi \in \mathcal{L}(W^{3,p}(T), W^{1,p}(T))$ follows for $p > 1$ from the Sobolev imbedding theorem and for $p = 1$ from the fact that $W^{2,1}(T) \subset C(\overline{T})$. Hypothesis (H.1) is obvious and hypothesis (H.2) is proved in [19]. □

Remark 2.2. The tenth parameter $(\partial(\Pi u)/\partial s_2)(Q_1)$ has no influence on the global smoothness of a global finite element function defined in a given triangulation; thus it can be different in two adjacent triangles with a common shortest side.

Now we introduce a triangular finite element of the Hermite type which does not satisfy Jamet's hypothesis (H.2); nevertheless, it satisfies estimates not depending on the minimum angle of T.

Theorem 2.3. *Let $\overline{T}$ be the same triangle as in Theorem 1.2 and let $a = \operatorname{dist}(P_2, P_3)$, $b = \operatorname{dist}(P_1, P_3)$, $c \equiv h_T = \operatorname{dist}(P_1, P_2)$. Let $\varphi \in C^1(\overline{T})$ and let*

$$|D^\alpha \varphi(P)| \le M_4 \quad \forall |\alpha| = 4, \quad \forall P \in T, \tag{2.2}$$

$$D^\alpha \varphi(P_j) = 0 \quad \forall |\alpha| \le 1 \ (j = 1,2,3), \quad \frac{\partial \varphi}{\partial n_a}(Q_1) = 0 \tag{2.3}$$

where Q_1 is the mid-point of the side P_2P_3 and n_a the unit normal to P_2P_3. Then we have for all $P \in \overline{T}$

$$|\varphi(P)| \le \frac{1}{96}\left(1 + 4\left(\frac{a}{c}\right)^3\right) M_4 \, c^4, \tag{2.4}$$

$$\left|\frac{\partial \varphi}{\partial x_j}(P)\right| \le \frac{4}{15}\left(1 + 5\left(\frac{a}{c}\right)^2\right) \frac{1}{\sin \beta_T} M_4 \, c^3 \quad (j = 1,2). \tag{2.5}$$

Proof. Theorem 2.3 is proved in [19]. Nevertheless, we reproduce this proof because it is surprisingly short. We restrict our considerations to the case

$$|D^i\varphi(P)| \le M_4 \quad \forall |i| = 4\,, \quad \forall P \in \overline{T}\,. \tag{2.6}$$

In the case (2.2) we can use the trick with an inscribed triangle $\overline{T}' \subset T$ in the same way as in [24]. The proof is based on the following four lemmas.

Lemma 2.4. *Let s_1, s_2 be two noncollinear directions making an angle ω. Let $\frac{\partial\varphi}{\partial s_j}(P) = k_j$ $(j = 1, 2)$, P being a point of the (x_1, x_2)-plane. Then*

$$\left|\frac{\partial\varphi}{\partial x_j}(P)\right| \le \frac{|k_1| + |k_2|}{|\sin\omega|} \quad (j = 1, 2)\,.$$

Further, let s_1 and s_2 be two directions orthogonal to one another. If $|\frac{\partial\psi}{\partial s_i}(P)| \le k_i$ $(i = 1, 2)$ then we have for an arbitrary direction s

$$\left|\frac{\partial\psi}{\partial s}(P)\right| \le |k_1| + |k_2|.$$

Lemma 2.5. *Let $g(0) = \eta_1$, $g(l) = \eta_2$, $g'(0) = k_1$, $g'(l) = k_2$ and $|g^{(4)}(s)| \le K_4$ in $(0, l)$. Then for $s \in [0, l]$*

$$|g(s)| \le \max|\eta_j| + \frac{4\,l}{27}(|k_1| + |k_2|) + \frac{K_4}{16 \cdot 24}l^4, \tag{2.7}$$

$$|g'(s)| \le \frac{3}{2\,l}(|\eta_1| + |\eta_2|) + \max|k_j| + \frac{K_4}{24}l^3 \tag{2.8}$$

Further, if $g(0) = g(l) = g'(0) = g'(l) = 0$ then

$$|g''(s)| \le \frac{1}{2}K_4 l^2. \tag{2.9}$$

Lemma 2.6. *Let $g(0) = \eta_1$, $g(l/2) = \eta_2$, $g(l) = \eta_3$ and $|g^{(3)}(s)| \le K_3$ in $(0, l)$. Then for $s \in [0, l]$*

$$|g(s)| \le \frac{5}{4}\max|\eta_j| + \frac{\sqrt{3}}{6^3}\,K_3 l^3\,, \tag{2.10}$$

$$|g'(s)| \le \frac{8}{l}\max|\eta_j| + \frac{1}{4}K_3 l^2. \tag{2.11}$$

Lemma 2.7. *Let $g(0) = \eta_1$, $g(l) = \eta_2$, $g'(l) = k_1$ and $|g^{(3)}(s)| \le K_3$ in $(0, l)$. Then for $s \in [0, l]$*

$$|g(s)| \le \max|\eta_i| + \frac{l}{4}|k_1| + \frac{2}{81}K_3 l^3. \tag{2.12}$$

Lemmas 2.4–2.7 are taken from [24] with a modification in (2.7) and improvements in (2.8) and (2.12).

We have by Lemma 2.5 (with $g = \varphi|_{P_2P_3}$) and assumptions (2.3) and (2.6)

$$\left|\left(\varphi\,\big|_{P_2P_3}\right)\right| \le \frac{1}{16\cdot 24}\cdot 4\,M_4a^4 = \frac{1}{96}\,M_4a^4\,, \tag{2.13}$$

$$\left|\left(\frac{\partial\varphi}{\partial a}\,\Big|_{P_2P_3}\right)\right| \le \frac{1}{24}\cdot 4\,M_4a^3 = \frac{1}{6}\,M_4a^3\,, \tag{2.14}$$

where $\partial/\partial a$ denotes the derivative in the direction of P_2P_3. Similarly, Lemma 2.6 with $g = \partial\varphi/\partial n_a|_{P_2P_3}$ yields

$$\left|\left(\frac{\partial\varphi}{\partial n_a}\,\Big|_{P_2P_3}\right)\right| \le \frac{4\sqrt{3}}{6^3}\,M_4a^3\,. \tag{2.15}$$

Using estimates (2.14), (2.15) and Lemma 2.4 we find for an arbitrary direction s

$$\left|\left(\frac{\partial\varphi}{\partial s}\,\Big|_{P_2P_3}\right)\right| \le \frac{43}{6^3}\,M_4a^3\,. \tag{2.16}$$

Let $P \in \overline{T}$, $P \neq P_1$ and let B be the point of the segment P_2P_3 which lies on the straight line determined by P_1 and P. Setting $l = \operatorname{dist}(B, P_1)$ and considering the function $g = \varphi\,\big|_{P_1B}$ we obtain by means of Lemma 2.5 and (2.3), (2.6), (2.13), (2.16)

$$|\varphi(P)| \le \frac{1}{96}\,M_4a^4 + \frac{4\,l}{27}\,\frac{43}{6^3}\,M_4a^3 + \frac{1}{16\cdot 24}\cdot 4M_4l^4, \tag{2.17}$$

$$\left|\frac{\partial\varphi}{\partial s}(P)\right| \le \frac{3}{2\cdot 96}M_4\frac{a^4}{l} + \frac{43}{6^3}M_4a^3 + \frac{1}{6}M_4l^3. \tag{2.18}$$

Estimate (2.17) implies (2.4). Estimate (2.18) will be used in deriving (2.5).

Relation (2.9) from Lemma 2.5 with $g = \varphi|_{P_2P_3}$ and relation (2.11) from Lemma 2.6 with $g = \partial\varphi/\partial n_a|_{P_2P_3}$ together with assumption (2.3) yield

$$\left|\frac{\partial^2\varphi}{\partial a^2}(B)\right| \le 2M_4a^2, \quad \left|\frac{\partial^2\varphi}{\partial a\partial n_a}(B)\right| \le M_4a^2.$$

Hence, according to the second part of Lemma 2.4 where we set $\psi = \partial\varphi/\partial a$,

$$\left|\frac{\partial^2\varphi}{\partial a\partial s}(B)\right| \le 3M_4a^2. \tag{2.19}$$

Using Lemma 2.7 with $g = \partial\varphi/\partial a|_{P_1B}$ and taking into account relations (2.3), (2.14), (2.19) we find

$$\left|\frac{\partial\varphi}{\partial a}(P)\right| \le \frac{1}{6}M_4a^3 + \frac{3}{4}M_4a^2l + \frac{8}{81}M_4l^3. \tag{2.20}$$

Inequalities (2.18) and (2.20) together with Lemma 2.4 imply (2.5). □

Now we introduce some consequences of Theorem 2.3.

Theorem 2.8. *A polynomial $p \in P_{3,2}$ is uniquely determined by its ten values*

$$D^\alpha p(P_j) \quad |\alpha| \le 1, \quad (j = 1, 2, 3); \quad \frac{\partial p}{\partial n_a}(Q_1)\,, \tag{2.21}$$

where the meaning of the symbols P_i, Q_1 and n_a is the same as in Theorem 2.3.

Proof. It is sufficient to prove the uniqueness. Let us assume that the values (2.21) are equal to zero. Setting $\varphi(x_1, x_2) = p(x_1, x_2)$ in Theorem 2.3 we have $M_4 = 0$ and estimate (2.4) implies $p(x_1, x_2) \equiv 0$. □

Theorem 2.9. *Let $u \in C^1(\overline{T})$ and let*

$$|D^\alpha u(P)| \le M_4 \quad \forall |\alpha| = 4, \quad \forall P \in T.$$

Let $p \in P_{3,2}$ satisfies the relations

$$\begin{aligned} D^\alpha p(P_j) = D^\alpha u(P_j), \quad |\alpha| \le 1 \quad (j = 1, 2, 3), \\ \frac{\partial p}{\partial n_a}(Q_1) = \frac{\partial u}{\partial n_a}(Q_1)\,. \end{aligned} \tag{2.22}$$

Then the function

$$\varphi(x_1, x_2) \equiv u(x_1, x_2) - p(x_1, x_2) \tag{2.23}$$

satisfies relations (2.4) and (2.5).

Proof. It follows from the assumptions of Theorem 2.9 that function (2.23) satisfies all conditions of Theorem 2.3. □

Remark 2.10. We return to the first part of the proof or Theorem 2.1: If the right-hand sides of (2.1) are equal to zero, then also $(\partial \Pi u/\partial n_a)(Q_1) = 0$ and $(\Pi u)(x, y) \equiv 0$, according to Theorem 2.8. □

It follows from Theorem 2.9 that triangular finite elements with polynomials $p \in P_{3,2}$ uniquely determined by parameters (2.21) can be used in triangulations satisfying the maximum angle condition: Estimate (2.5) requires the next-to-smallest angles of all triangles to be bounded away from zero. This requirement (we call it *the second angle condition*) is equivalent with the maximum angle condition.

Some triangular finite elements of the Hermite type are sketched in Fig. 2. The black circle denotes the function value, the arrows and double arrows denote the first and second normal derivatives, respectively, and the circled integers k denote the values $D^\alpha p(P_i)$, $|\alpha| \le k$, where P_i is the centre of the circle.

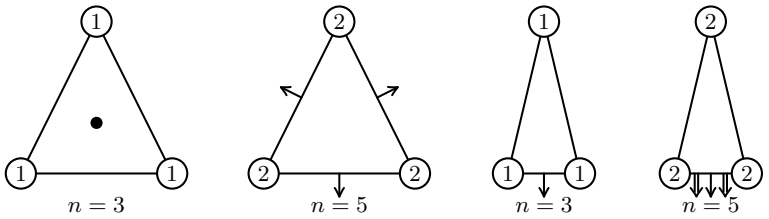


Fig. 2. Triangular finite elements of the Hermite type.

Remark 2.11. The method of the proof of Theorem 2.3 does not work successfully in the case of the classical Hermite triangular finite element of third degree where the last condition (2.3) is substituted by $\varphi(P_0) = 0$, P_0 being the center of gravity of $\overline{T}$, because we obtain only

$$|(\partial^2\varphi/\partial a\partial n_a|_{P_2P_3})| \le KM_4l^3/a \quad (l = \operatorname{dist}(P_1Q_1))$$

and $l/a \to \infty$ with $a \to 0$.

The hypothesis (H.2) is not also satisfied. This can be proved by the following example: Let $u(x, y) = y^4$ and let the triangle $\overline{T}$ have the vertices $P_1(0, 0)$, $P_2(1, 0)$, $P_3(0, 1)$. Then the polynomial of third degree satisfying the first nine conditions (2.22) and condition $p(P_0) = u(P_0)$, where P_0 is the center of gravity of $\overline{T}$, has the form

$$p(x, y) = \frac{4}{3}\left(xy - \frac{3}{4}y^2 - x^2y - xy^2 + \frac{3}{2}y^3\right).$$

We see that $\partial u/\partial x \equiv 0$ while $\partial p/\partial x \neq 0$ in T. Thus hypothesis (H.2) is not satisfied and we cannot apply Jamet's theory on this finite element.

Remark 2.12. In [2, p. 222] the parameters

$$D^\alpha p(P_j) \quad |\alpha| \le 1 \quad (j = 1, 2, 3); \qquad \iint_T \frac{\partial^2 p}{\partial x\partial y}dxdy \tag{2.24}$$

were considered in connection with the maximum angle condition for a cubic triangular finite element on a right triangle with the sides P_1P_2 and P_2P_3 lying on the axes x and y, respectively. However, parameters (2.24) do not determine in a general case a polynomial $p \in P_{3,2}$ uniquely. To prove it let us consider a triangle with vertices $P_i(x_i, y_i)$ $(i = 1, 2, 3)$ and let T_0 be the triangle lying in the ξ, η-plane with vertices $P_1^*(0, 0)$, $P_2^*(1, 0)$, $P_3^*(0, 1)$. The transformation

$$x = x(\xi, \eta) \equiv x_1 + \overline{x}_2\xi + \overline{x}_3\eta, \quad y = y(\xi, \eta) \equiv y_1 + \overline{y}_2\xi + \overline{x}_3\eta, \tag{2.25}$$

where

$$\overline{x}_j = x_j - x_1, \quad \overline{y}_j = y_j - y_1 \quad (j = 2, 3), \tag{2.26}$$

maps the triangle $\overline{T}_0$ one-to-one onto $\overline{T}$. Let us set

$$p^*(\xi,\eta) = p(x(\xi,\eta), y(\xi,\eta)). \tag{2.27}$$

If all ten parameters (2.24) are equal to zero then

$$D^\alpha p^*(P_j^*) = 0 \quad |\alpha| \le 1 \quad (j = 1,2,3), \tag{2.28}$$

$$\iint_{T_0} \left\{ -\overline{x}_3\overline{y}_3 \frac{\partial^2 p^*}{\partial\xi^2} + (\overline{x}_2\overline{y}_3 + \overline{x}_3\overline{y}_2) \frac{\partial^2 p^*}{\partial\xi\partial\eta} - \overline{x}_2\overline{y}_2 \frac{\partial^2 p^*}{\partial\eta^2} \right\} d\xi d\eta = 0. \tag{2.29}$$

Relations (2.28) imply

$$p^*(\xi,\eta) = K\xi\eta(1-\xi-\eta). \tag{2.30}$$

Inserting (2.30) into (2.29) we obtain

$$K\{2(\overline{x}_2\overline{y}_2 + \overline{x}_3\overline{y}_3) - (\overline{x}_2\overline{y}_3 + \overline{x}_3\overline{y}_2)\} = 0. \tag{2.31}$$

If the difference standing in braces is different from zero then (2.31) implies $K = 0$ and parameters (2.24) determine uniquely $p \in P_{3,2}$. However, if

$$2(\overline{x}_2\overline{y}_2 + \overline{x}_3\overline{y}_3) = \overline{x}_2\overline{y}_3 + \overline{x}_3\overline{y}_2\,, \tag{2.32}$$

then (2.31) is satisfied with $K \neq 0$ and $p(x,y) \neq 0$, according to (2.30) and (2.27).

Let us describe these situations. It cannot be simultaneously $\overline{x}_2 = \overline{x}_3 = 0$ (and similarly $\overline{y}_2 = \overline{y}_3 = 0$). Let $\overline{x}_2 \neq 0$. If $\overline{y}_2 = 0$ then (2.32) gives $\overline{x}_3 = \overline{x}_2/2$ with arbitrary $\overline{y}_3 \neq 0$. Conversely, if $\overline{x}_3 = \overline{x}_2/2$ then (2.32) implies $\overline{y}_2 = 0$. In other cases

$$\overline{y}_3 = \frac{(2\overline{x}_2 - \overline{x}_3)\overline{y}_2}{\overline{x}_2 - 2\overline{x}_3} \quad (\overline{y}_2 \neq 0,\ \overline{x}_2 \neq 2\overline{x}_3).$$

The situation $\overline{x}_3 \neq 0$ can be treated similarly with the same results. □

Now we mention briefly some higher-degree polynomials. We shall modify the family of triangular finite elements introduced by Koukal in [7] and [8].

Theorem 2.13. *Let $u \in C^k(\overline{T})$ $(k \ge 1)$. A polynomial $p \in P_{2k+1,2}$ is uniquely determined by conditions*

$$D^\alpha p(P_j) = D^\alpha u(P_j), \quad |\alpha| \le k \quad (j = 1,2,3), \tag{2.33}$$

$$\frac{\partial^r p}{\partial n_a^r}(Q_j^{(r)}) = \frac{\partial^r u}{\partial n_a^r}(Q_j^{(r)}) \quad (j = 1,\dots,r;\ r = 1,\dots,k)\,, \tag{2.34}$$

where the symbol $\partial/\partial n_a$ has the meaning as in Theorem 2.3 and $Q_1^{(r)},\dots,Q_r^{(r)}$ $(1 \le r \le k)$ are the points dividing the side P_2P_3 into $r+1$ parts of the same length.

Theorem 2.14. *Let $u \in C^k(\overline{T})$ $(k \geq 1)$. A polynomial $\Pi u \in P_{2k+1,2}$ is uniquely determined by the conditions*

$$D^\alpha(\Pi u)(P_j) = D^\alpha u(P_j), \quad |\alpha| \leq 1 \quad (j = 1, 2, 3), \tag{2.35}$$

$$\frac{\partial^r(\Pi u)}{\partial s_2^r}(Q_j^{(r)}) = \frac{\partial^r u}{\partial s_2^r}(Q_j^{(r)}) \quad (j = 1, \ldots, r;\ r = 1, \ldots, k)\,, \tag{2.36}$$

where $\partial/\partial s_2$ denotes the derivative in the direction of the side P_3P_1.

For $k = 1$ the assertions of both theorems are contained in Theorems 2.1 and 2.8. In the case $k \geq 2$ the proof is a modification of the proof of [18, Theorem 17.1].

Generalizing a little the preceding considerations we can prove:

Theorem 2.15. *Let $u \in W^{2k+2,p}(T)$, where $k \geq 1$ and $p \in [1, \infty]$, and let the operator Π be defined by (2.35), (2.36). Then we have for $m = 0, 1$*

$$|u - \Pi u|_{m,p,T} \leq C \frac{h_T^{2k+1}}{\cos(\gamma_T/2)} |u|_{2k+2,p,T}. \tag{2.37}$$

Remark 2.16. A generalization of Theorem 2.3 to the case of interpolation polynomials introduced in Theorem 2.13 is possible. Instead of special Lemmas 2.5–2.7 we can use [16, Theorem 2]. We obtain the estimates

$$|\varphi(P)| \leq C\, M_{2k+2} c^{2k+2}, \quad \left|\frac{\partial \varphi}{\partial x_j}(P)\right| \leq \frac{C}{\sin\beta}\, M_{2k+2} c^{2k+1},$$

where $P \in \overline{T}$ and $j = 1, 2$.

Remark 2.17. The construction of finite elements introduced in Theorem 2.13 implies the following conjecture: *It is impossible to construct a triangular finite C^1-element which satisfies the maximum angle condition.*

3 Variational crimes and semiregular finite elements in the case of smooth solutions

3.A Formulation of the problem

We shall consider the boundary value problem

$$-\sum_{i=1}^{2} \frac{\partial}{\partial x_i}\left(k_i(x)\frac{\partial u}{\partial x_i}\right) = f(x), \quad x \in \Omega, \tag{3.1}$$

$$u = 0 \quad \text{on } \Gamma_1, \tag{3.2}$$

$$\sum_{i=1}^{2} k_i \frac{\partial u}{\partial x_i} n_i(\Omega) = q \quad \text{on } \Gamma_2, \tag{3.3}$$

where Ω is a two-dimensional bounded domain with the boundary $\partial\Omega = \Gamma_1 \cup \Gamma_2$, Γ_1 and Γ_2 being the circles with radii R_1 and $R_2 = R_1 + \varrho$, respectively. We assume that the circles Γ_1, Γ_2 have the same center S_0 and that

$$R_1 \gg \varrho. \tag{3.4}$$

The symbols $n_i(G)$ $(i = 1, 2)$ denote the components of the unit outward normal to ∂G.

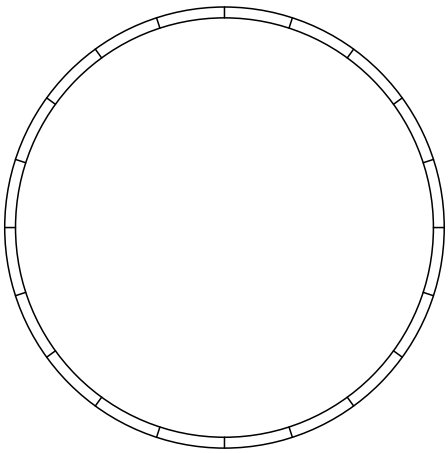


Fig. 3.

A weak solution of problem (3.1)–(3.3) is a solution of the following variational problem (which can be obtained from (3.1)–(3.3) by means of Green's theorem in a standard way).

Problem 3.1. Let Ω be a bounded domain with a Lipschitz continuous boundary $\partial\Omega = \Gamma_1 \cup \Gamma_2$. Let

$$V = \{v \in H^1(\Omega) : v = 0 \quad \text{on } \Gamma_1\}, \tag{3.5}$$

$$a(w, v) = \sum_{i=1}^{2} \iint_{\Omega} k_i(x) \frac{\partial w}{\partial x_i} \frac{\partial v}{\partial x_i} \, \mathrm{d}x_1 \mathrm{d}x_2, \tag{3.6}$$

$$L(v) = L^{\Omega}(v) + L^{\Gamma}(v) = \iint_{\Omega} vf \, \mathrm{d}x_1 \mathrm{d}x_2 + \int_{\Gamma_2} vq \, \mathrm{d}s \,, \tag{3.7}$$

where

$$\begin{gathered} k_i \in W^{1,\infty}(\Omega), \quad f \in W^{1,\infty}(\Omega), \\ q = Q \,\big|_{\Gamma_2}, \quad Q \in C^2(\overline{U}), \end{gathered} \tag{3.8}$$

$$k_i(x) \geq \mu_0 > 0, \tag{3.9}$$

U being a neighbourhood of Γ_2 (i.e., a domain containing Γ_2). Find $u \in V$ such that

$$a(u, v) = L(v) \quad \forall v \in V. \tag{3.10}$$

Assumptions (3.8)–(3.9) guarantee that the symmetric bilinear form (3.6) is bounded and strongly coercive and that the linear form (3.7) is continuous. (Of course, this also holds when $f \in L_2(\Omega)$ and $q \in L_2(\Gamma_2)$. We assume (3.8) because of numerical integration.)

Lemma 3.2. *Let a solution $u \in V$ of Problem 3.1 satisfy $u \in H^2(\Omega)$. Then relation (3.1) holds almost everywhere in Ω and relation (3.3) holds almost everywhere on Γ_2.*

The proof is omitted. Also the following lemma is well-known:

Lemma 3.3. *If (3.9) holds then Problem 3.1 has a unique solution.*

We shall solve Problem 3.1 approximately by the finite element method. To this end let us approximate Γ_2 by a regular polygon Γ_{2h} with vertices $Q_1, \ldots, Q_n$ such that every segment Q_iQ_{i+1} has no common point with Γ_1. Let the vertices $P_1, \ldots, P_n$ of the polygon Γ_{1h} approximating Γ_1 be obtained in the following way: P_i is the intersection of the segment S_0Q_i with Γ_1. The symbol Ω_h will denote the polygonal domain with the boundary $\partial\Omega_h = \Gamma_{1h} \cup \Gamma_{2h}$.

We divide each segment P_iQ_i by the points $A_1^i, A_2^i, \ldots, A_{m-1}^i$ into m parts of the same length in such a way that we have formally $A_0^i = P_i$, $A_m^i = Q_i$. The points A_j^i are the vertices of quadrilaterals into which the domain Ω_h is divided. Such a division of Ω_h will be denoted $\mathcal{D}_h^K$. If we divide each quadrilateral of $\mathcal{D}_h^K$ into two triangles we obtain a division $\mathcal{D}_h^T$ (see Fig. 4). We shall also consider an auxiliary division $\mathcal{D}_h^A$ which will be constructed from $\mathcal{D}_h^K$ by dividing each quadrilateral $A_{m-1}^i A_{m-1}^{i+1} Q_i Q_{i+1}$ into two triangles.

We admit to use narrow quadrilaterals and narrow triangles. This means that we shall have

$$\frac{\varrho}{m} \ll h \tag{3.11}$$

in our considerations, where h is the length of the greatest segment in the division of Ω_h.

We shall assume that $k_i \in W^{1,\infty}(\widetilde{\Omega})$, $f \in W^{1,\infty}(\widetilde{\Omega})$, where $\widetilde{\Omega}$ is such that $\Omega_h \subset \widetilde{\Omega}$ for sufficiently small h. When we consider the functions k_i and f in Ω_h we shall use symbols $\widetilde{k}_i$ and $\widetilde{f}$. In the opposite case the original symbols k_i and f will be used.

The discrete problem is now formulated in an almost standard way. (The expression "almost" concerns the approximation of the term $L^\Gamma(v)$ which needs some space.) Let $\mathcal{D}_h$ denote one of the three divisions $\mathcal{D}_h^K$, $\mathcal{D}_h^T$, $\mathcal{D}_h^A$. We define

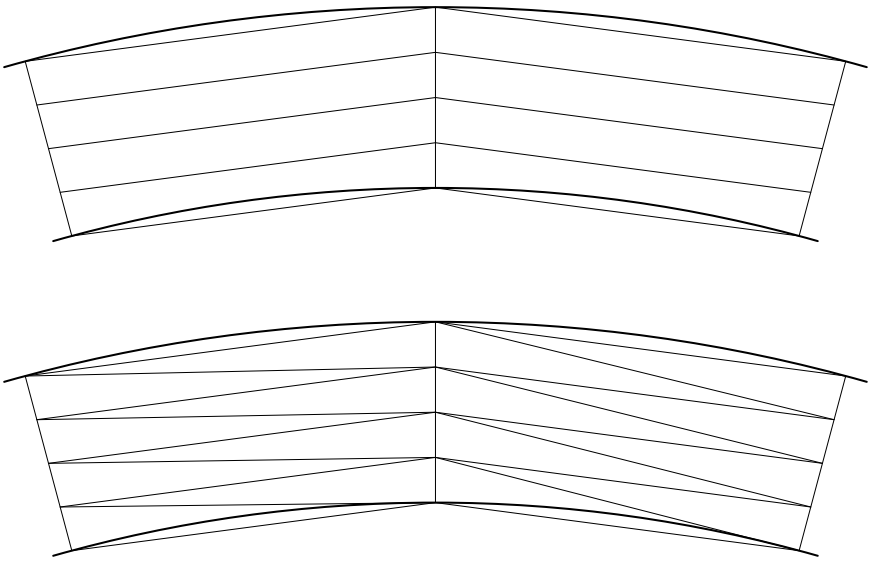


Fig. 4.

spaces

$$X_h = \{v \in C(\overline{\Omega}_h) : v\,\big|_{\,K} = \text{ a four-node isoparametric function} \quad \forall \overline{K} \in \mathcal{D}_h, \\ v\,\big|_{\,T} = \text{ a linear polynomial} \quad \forall \overline{T} \in \mathcal{D}_h\} \tag{3.12}$$

and

$$V_h = \{v \in X_h : v = 0 \text{ on } \Gamma_{1h}\}. \tag{3.13}$$

We set for all $v, w \in H^1(\Omega_h)$

$$\widetilde{a}_h(v,w) = \sum_{i=1}^{2} \iint_{\Omega_h} \widetilde{k}_i \frac{\partial v}{\partial x_i} \frac{\partial w}{\partial x_i} \,\mathrm{d}x_1 \mathrm{d}x_2 \tag{3.14}$$

and

$$\widetilde{L}_h^{\Omega}(v) = \iint_{\Omega_h} v\widetilde{f} \,\mathrm{d}x_1 \mathrm{d}x_2 \quad \forall v \in X_h. \tag{3.15}$$

To define $\widetilde{L}_h^{\Gamma}(v)$ is more complicated. Therefore, we omit it and refer only to [21].

The symbols $a_h(v,w)$, $L_h^\Omega(v)$ and $L_h^\Gamma(v)$, where $v, w \in X_h$, will denote the approximations of $\widetilde{a}_h(v,w)$, $\widetilde{L}_h^\Omega(v)$ and $\widetilde{L}_h^\Gamma(v)$, respectively, when using numerical integration. For example, in the case of $\mathcal{D}_h^T$ we have for all $v, w \in X_h$

$$a_h(v,w) = \sum_{T\in\mathcal{D}_h^T} \sum_{i=1}^{2} \sum_{j=1}^{N_T} 2\omega_{T_0,j} \widetilde{k}_i(x_{T,j}) \left.\frac{\partial v}{\partial x_i}\right|_T \left.\frac{\partial w}{\partial x_i}\right|_T \operatorname{mes}_2 T,$$

where $x_{T,j}$ are the integration points on a triangle T and $\omega_{T_0,j}$ the corresponding coefficients of the given integration formulas (prescribed on the reference triangle $\overline{T}_0$).

Now we can define the approximate problem:

Problem 3.4. Find $u_h \in V_h$ such that

$$a_h(u_h, v) = L_h(v) \quad \forall v \in V_h. \tag{3.16}$$

3.B An abstract error estimate

Definition 3.5. Let $u \in H^2(\Omega)$. We define $Q_h u \in X_h$ by

$$Q_h u \,|_{\overline{K}\in\mathcal{D}_h} = Q_K u = \text{the four-node isoparametric interpolant of } u,$$
$$Q_h u \,|_{\overline{T}\in\mathcal{D}_h} = I_T u = \text{the linear interpolant of } u,$$

where $\mathcal{D}_h$ is one of the divisions $\mathcal{D}_h^K$, $\mathcal{D}_h^T$, $\mathcal{D}_h^A$.

Lemma 3.6. *Let Γ_0 be the circle with a center S_0 and radius $R_0 = R_1 - \varrho$. Let $\widetilde{\Omega}$ be a bounded domain such that $\partial\widetilde{\Omega} = \Gamma_0 \cup \Gamma_2$. There exists a linear and bounded extension operator $E : H^k(\Omega) \to H^k(\widetilde{\Omega})$ such that the constant C appearing in the inequality*

$$\|E(v)\|_{k,\widetilde{\Omega}} \le C\|v\|_{k,\Omega} \quad \forall v \in H^k(\Omega)$$

does not depend on R_1/ϱ and v. The operator E is also a linear and bounded extension operator from $H^{k-i}(\Omega)$ into $H^{k-i}(\widetilde{\Omega})$ $(1 \le i \le k)$.

Lemma 3.6 follows from the considerations introduced in [13, pp. 20–22].

Theorem 3.7. *Let $u \in H^2(\Omega)$, $\widetilde{u} := E(u)$ and let the condition*

$$\|v\|^2_{1,\Omega_h} \le C a_h(v,v) \quad \forall v \in V_h, \quad \forall h \in (0, h_0) \tag{3.17}$$

be satisfied, where the constant C does not depend on v and h and where h_0 is sufficiently small. Then Problem 3.4 has a unique solution $u_h \in V_h$ and there

exists a positive constant C_0 *independent of* $u \in H^2(\Omega)$ *and* $w \in V_h$ *such that*

$$C_0^{-1}\|\widetilde{u} - u_h\|_{1,\Omega_h} \le \|Q_h u - \widetilde{u}\|_{1,\Omega_h} + \sup_{\substack{w\in V_h \\ w\neq 0}} \frac{|a_h(Q_h u, w) - \widetilde{a}_h(Q_h u, w)|}{\|w\|_{1,\Omega_h}} +$$

$$+ \sup_{\substack{w\in V_h \\ w\neq 0}} \frac{|\widetilde{L}_h^{\Omega}(w) - L_h^{\Omega}(w)|}{\|w\|_{1,\Omega_h}} + \sup_{\substack{w\in V_h \\ w\neq 0}} \frac{|\widetilde{L}_h^{\Gamma}(w) - L_h^{\Gamma}(w)|}{\|w\|_{1,\Omega_h}} + \sup_{\substack{w\in V_h \\ w\neq 0}} \frac{|\widetilde{a}_h(\widetilde{u}, w) - \widetilde{L}_h(w)|}{\|w\|_{1,\Omega_h}}. \tag{3.18}$$

Theorem 3.7 is proved in [21]. Our first aim is to prove that condition (3.17) is satisfied. This will be done in subsection 3.D, where we also give estimates of the second, third and fourth terms appearing on the right-hand side of (3.18). These terms express the error of numerical integration.

The estimate of the first term, which expresses the interpolation error, is introduced in subsection 3.C. This estimate follows from the known interpolation theorems. The fifth term, which expresses the error due to the approximation of the boundary, will be estimated in subsection 3.E.

3.C The interpolation error

The estimate of the first term appearing on the right-hand side of (3.18) follows from Theorems 1.3 and 1.4:

Theorem 3.8. *We have*

$$\|Q_h u - \widetilde{u}\|_{1,\Omega_h} \le Ch\|u\|_{2,\Omega},$$

where the constant C *is independent of* h, u *and the division* $\mathcal{D}_h$.

3.D The effect of numerical integration

The effect of numerical integration must be analyzed more carefully than in the case of regular elements. In the case of triangles the result is that the numerical integration does not depend on the geometry of triangles and that the degrees of precision of quadrature formulas sufficient for the rate of convergence $O(h)$ are the same as in the regular case (except for the integration along the boundary Γ_{2h} – see Theorem 3.18). The proofs of the assertions presented in this subsection can be found in [21].

First we mention the analysis of the numerical integration on quadrilaterals. Let $\overline{K}$ be a quadrilateral whose greatest side lies on the axis x_1 and let it have the vertices

$$P_1(h, 0),\ P_2(0, 0),\ P_3(\delta\cos\beta, \delta\sin\beta),\ P_4(h - \varepsilon\cos\alpha, \varepsilon\sin\alpha)$$

where $\varepsilon = \operatorname{dist}(P_1, P_4)$, $\delta = \operatorname{dist}(P_2, P_3)$ and α and β are the angles at P_1 and P_2, respectively. As each quadrilateral belonging to $\mathcal{D}_h$ has parallel long sides we have

$$b := \frac{\varrho}{m} = \varepsilon\sin\alpha = \delta\sin\beta.$$

Let $\overline{K}_0$ be the reference square lying in the coordinate system ξ_1, ξ_2 and having the vertices $P_1^*(1,0)$, $P_2^*(0,0)$, $P_3^*(0,1)$, $P_4^*(1,1)$. If we denote

$$\varepsilon_3 = \delta \cos\beta, \quad \varepsilon_4 = \varepsilon \cos\alpha, \quad \varepsilon^* = \varepsilon_3 + \varepsilon_4,$$

then the one-to-one mapping of $\overline{K}_0$ onto $\overline{K}$ has the form

$$x_1 = h\xi_1 + \varepsilon_3\xi_2 - \varepsilon^*\xi_1\xi_2, \quad x_2 = b\xi_2. \tag{3.19}$$

If the side P_1P_2 makes an angle φ with the axis x_1 and the vertex P_2 has coordinates x_{10}, x_{20} then (3.19) is substituted by the mapping

$$\begin{aligned} x_1 &= x_1^K(\xi_1,\xi_2) \equiv x_{10} + (h\xi_1 + \varepsilon_3\xi_2 - \varepsilon^*\xi_1\xi_2)\cos\varphi - b\xi_2\sin\varphi, \\ x_2 &= x_2^K(\xi_1,\xi_2) \equiv x_{20} + (h\xi_1 + \varepsilon_3\xi_2 - \varepsilon^*\xi_1\xi_2)\sin\varphi + b\xi_2\cos\varphi. \end{aligned} \tag{3.20}$$

Both transformations (3.19) and (3.20) have the same Jacobian

$$J_K = (h - \varepsilon^*\xi_2)b.$$

It should be noted that for $n \gg 1$ we have

$$\varepsilon_i \approx \frac{1}{2n}(2\pi(R_1 + \Delta + \frac{\varrho}{m}) - 2\pi(R_1 + \Delta)) = \frac{\pi\varrho}{nm} \quad (i = 3,4;\ 0 \le \Delta \le \varrho(1 - 1/m)).$$

Further

$$h \approx \frac{2\pi R_1}{n}.$$

The last two relations imply in this case

$$\varepsilon_i = \sigma_i b, \quad \sigma_i \le Ch \quad (i = 3,4). \tag{3.21}$$

Let us denote

$$(1) := 2, \quad (2) := 1, \quad \kappa_{ij} = (-1)^{i+j}. \tag{3.22}$$

Then we can write (omitting the subscript K at J)

$$\frac{\partial\xi_i}{\partial x_j} = \kappa_{ij}\frac{1}{J}\frac{\partial x_{(j)}}{\partial\xi_{(i)}} \quad (i,j = 1,2) \tag{3.23}$$

and the theorem on transformation of an integral yields

$$E_K\left(\sum_{i=1}^{2}\widetilde{k}_i\frac{\partial v}{\partial x_i}\frac{\partial w}{\partial x_i}\right) = E_{K_0}\left(\sum_{i,r,s=1}^{2}\widetilde{k}_i^*\chi_{irs}\frac{\partial v^*}{\partial\xi_r}\frac{\partial w^*}{\partial\xi_s}\right) \tag{3.24}$$

where

$$E_K(F) := \iint_K F(x_1, x_2)\,\mathrm{d}x_1\mathrm{d}x_2 - \sum_{j=1}^{N_K} \omega_{K_0,j} F(x_{K,j}) |J_K(\xi_{1j}, \xi_{2j})|, \tag{3.25}$$

$$F^*(\xi_1, \xi_2) := F(x_1(\xi_1, \xi_2), x_2(\xi_1, \xi_2)),$$

$$E_{K_0}(F) := \iint_{K_0} F^*(\xi_1, \xi_2)\mathrm{d}\xi_1\mathrm{d}\xi_2 - \sum_{j=1}^{N_K} \omega_{K_0,j} F^*(\xi_{1j}, \xi_{2j}), \tag{3.26}$$

$$\chi_{irs} = \kappa_{ir}\kappa_{is}\frac{1}{J}\frac{\partial x_{(i)}}{\partial \xi_{(r)}}\frac{\partial x_{(i)}}{\partial \xi_{(s)}}$$

with $[\xi_{1j}, \xi_{2j}]$ the integration points on $\overline{K}_0$.

Theorem 3.9. *Let*

$$E_{K_0}(p) = 0 \quad \forall p \in \mathcal{P}_2,$$

where $\mathcal{P}_k$ denotes the set of polynomials in two variables of degree not greater than k. Then we have

$$\left| E_K \left(\sum_{i=1}^{2} \widetilde{k}_i \frac{\partial v}{\partial x_i}\frac{\partial w}{\partial x_i} \right) \right| \leq Ch \max_{i=1,2} \|\widetilde{k}_i\|_{1,\infty,K} |v|_{1,K} |w|_{1,K} \quad \forall v,\, w \in X_h.$$

As the Jacobian J of both transformations (3.19) and (3.20) is the same the proof in both cases is very similar.

Remark 3.10. In the cases when relation (3.21) is not satisfied (however, the long sides are parallel) the assertion of Theorem 3.9 can be proved provided

$$E_{K_0}(p) = 0 \quad \forall p \in \mathcal{P}_4.$$

Remark 3.11. The case of a quadrilateral K with parallel long sides is a special case of quadrilaterals K satisfying the condition

$$|\varepsilon \sin\alpha - \delta \sin\beta| \leq Cbh. \tag{3.27}$$

It can be proved that the results of Theorem 3.9 and Remark 3.10 can be extended to the case (3.27).

The effect of numerical integration in the case of narrow triangles must be analyzed more carefully than in the case of regular triangles. Let $\overline{T}$ be an arbitrary triangle lying in the plane x_1, x_2 and let $\overline{T}_0$ be the reference triangle with vertices $(0,0)$, $(1,0)$, $(0,1)$ lying in the plane ξ_1, ξ_2. Let

$$x_1 = x_1(\xi_1, \xi_2), \quad x_2 = x_2(\xi_1, \xi_2) \tag{3.28}$$

be the linear transformation which maps $\overline{T}_0$ one-to-one onto $\overline{T}$ (for its form see, for example, (2.25), (2.26)) and let $\xi_1 = \xi_1(x_1, x_2)$, $\xi_2 = \xi_2(x_1, x_2)$ be its inverse.

Lemma 3.12. *Let $v \in C^1(\overline{T})$ and let*

$$v^*(\xi_1, \xi_2) = v(x_1(\xi_1, \xi_2), x_2(\xi_1, \xi_2)).$$

Then we have

$$\left\| \sum_{r=1}^{2} \frac{\partial v^*}{\partial \xi_r} \frac{\partial \xi_r}{\partial x_i} \right\|_{0,T_0} \le C|J|^{-1/2}|v|_{1,T},$$

where J is the Jacobian of (3.28).

The error functionals E_T and E_{T_0} on a triangle $\overline{T}$ and the reference triangle $\overline{T}_0$, respectively, are defined in a similar way as E_K and E_{K_0} (see (3.25) and (3.26)), their expression is only simpler. Using Lemma 3.12 we can prove the following theorem.

Theorem 3.13. *Let $\overline{T}$ be an arbitrary triangle (not necessarily satisfying the maximum angle condition). Let*

$$E_{T_0}(p) = 0 \quad \forall p \in \mathcal{P}_0.$$

Then we have

$$\left| E_T\left(\sum_{i=1}^{2} \widetilde{k}_i \frac{\partial v}{\partial x_i} \frac{\partial w}{\partial x_i} \right) \right| \le Ch \max_{i=1,2} |\widetilde{k}_i|_{1,\infty,T} |v|_{1,T} |w|_{1,T} \quad \forall v,\, w \in X_h.$$

For $v, w \in V_h$ we have

$$a_h(v, w) = \widetilde{a}_h(v, w) - \{\widetilde{a}_h(v, w) - a_h(v, w)\},$$

$$\widetilde{a}_h(v, w) - a_h(v, w) = \sum_{\overline{K} \in \mathcal{D}_h} E_K\left(\sum_{i=1}^{2} \widetilde{k}_i \frac{\partial v}{\partial x_i} \frac{\partial w}{\partial x_i} \right) + \sum_{\overline{T} \in \mathcal{D}_h} E_T\left(\sum_{i=1}^{2} \widetilde{k}_i \frac{\partial v}{\partial x_i} \frac{\partial w}{\partial x_i} \right).$$

Using these relations we obtain from Theorems 3.9 and 3.13 (details are similar as in the proof of [18, Theorem 11.8]; we use in addition the discrete Friedrichs' inequality of the type [18, (29.1)] (for its proof see Appendix) which together with (3.9) implies $\|v\|^2_{1,\Omega_h} \le C\widetilde{a}_h(v, v)\ \forall v \in V_h$):

Corollary 3.14. *If the forms $a_h(v, w)$, where $v, w \in X_h$, are computed from $\widetilde{a}_h(v, w)$ by means of quadrature formulas required in Theorems 3.9 and 3.13, then condition (3.17) is satisfied.*

Theorem 3.15. *Let*

$$E_{K_0}(p) = 0 \quad \forall p \in \mathcal{P}_2, \qquad E_{T_0}(p) = 0 \quad \forall p \in \mathcal{P}_0.$$

Then we have for $u \in H^2(\Omega)$

$$\sup_{\substack{w \in V_h \\ w \neq 0}} \frac{|a_h(Q_h u, w) - \widetilde{a}_h(Q_h u, w)|}{\|w\|_{1,\Omega_h}} \le Ch \max_{i=1,2} \|\widetilde{k}_i\|_{1,\infty,\widetilde{\Omega}} \|u\|_{2,\Omega}\,, \tag{3.29}$$

where the constant C does not depend on u, $\widetilde{k}_i$, and h.

Proof. Relation (3.29) follows from Theorems 3.9, 3.13 and 1.3, 1.4. Details are the same as in the proof of [18, Theorem 11.12]. □

Theorem 3.16. *Let*

$$E_{K_0}(p) = 0 \quad \forall p \in \mathcal{P}_2 \ (\text{or } \forall p \in \mathcal{Q}_1),$$
$$E_{T_0}(p) = 0 \quad \forall p \in \mathcal{P}_0\,,$$

where $\mathcal{Q}_1$ is the set of all bilinear polynomials. Then we have

$$\sup_{\substack{w \in V_h \\ w \neq 0}} \frac{|\widetilde{L}_h^{\Omega}(w) - L_h^{\Omega}(w)|}{\|w\|_{1,\Omega_h}} \leq Ch \|\widetilde{f}\|_{1,\infty,\widetilde{\Omega}} \sqrt{\mathrm{mes}_2 \Omega},$$

where the constant C does not depend on $\widetilde{f}$ and h.

In order to estimate the effect of numerical integration along Γ_2 we introduce the following error functionals:

$$E_r(F) := \int_0^{l_r} F(\xi_r)\mathrm{d}\xi_r - \sum_{j=1}^{N_r} l_r \beta_{r,j} F(s_{r,j}),$$
$$E_0(F^*) := \int_0^1 F^*(t)\mathrm{d}t - \sum_{j=1}^{N_r} \beta_{r,j} F^*(t_j),$$

where $s_{r,j}$ are integration points on $[0, l_r]$, $\beta_{r,j}$ the corresponding coefficients of the given integration formula and

$$F^*(t) := F(l_r t), \quad t \in I \equiv [0,1].$$

Hence

$$E_r(F) = l_r E_0(F^*).$$

When considering the line integrals we need also the trace inequalities which are introduced in the following lemma.

Lemma 3.17. *We have*

$$\|v\|_{0,\partial\Omega} \leq \frac{C}{\sqrt{\varrho}} \|v\|_{1,\Omega} \quad \forall v \in H^1(\Omega), \tag{3.30}$$
$$\|v\|_{0,\partial\Omega_h} \leq \frac{C}{\sqrt{\varrho}} \|v\|_{1,\Omega_h} \quad \forall v \in H^1(\Omega_h), \tag{3.31}$$

where the constant C does not depend on v, h and ϱ.

The proofs of (3.30) and (3.31) are similar to [12, pp. 15–16]).

Theorem 3.18. *Let*

$$E_0(p) = 0 \quad \forall p \in \mathcal{P}_2.$$

Then we have

$$\sup_{\substack{w \in V_h \\ w \neq 0}} \frac{|\widetilde{L}_h^{\Gamma}(w) - L_h^{\Gamma}(w)|}{\|w\|_{1,\Omega_h}} \leq \frac{C}{\sqrt{\varrho}} h^2 M_2(q) \sqrt{\mathrm{mes}_1 \Gamma_2},$$

where the constant C does not depend on q, ϱ and h and where $M_2(q)$ depends on the first and second derivatives of the function Q at the points of Γ_2 (as to the relation between q and Q see (3.8)).

3.E The error of the approximation of the boundary

The estimate of the last term in (3.18) will be divided into several lemmas.

Notation 3.19. We denote

$$\tau_h = \Omega_h - \overline{\Omega}, \quad \omega_h = \Omega - \overline{\Omega}_h. \tag{3.32}$$

Further, let $w \in X_h$. The symbol $\overline{w}$ is called the natural extension of w and denotes the function $\overline{w} : \overline{\Omega}_h \cup \overline{\Omega} \to R^1$ such that $\overline{w} = w$ on Ω_h and

$$\overline{w}\,\big|_{\overline{T}^{\,\mathrm{id}} - \overline{T}} = p\,\big|_{\overline{T}^{\,\mathrm{id}} - \overline{T}},$$

where $p \in \mathcal{P}_1$ satisfies $p\,\big|_{\overline{T}} = w\,\big|_{\overline{T}}$. ($\overline{T}^{\,\mathrm{id}} \subset \Omega$ is the curved triangle which is approximated by $\overline{T}$.)

Lemma 3.20. *Let $u \in H^2(\Omega)$. Then we have for $w \in V_h$*

$$\begin{aligned}
|\widetilde{a}_h(\widetilde{u}, w) - \widetilde{L}_h(w)| \leq |L^{\Gamma}(\overline{w}) - \widetilde{L}_h^{\Gamma}(w)| + \\
+ \left| \iint_{\omega_h} \sum_{i=1}^{2} \frac{\partial}{\partial x_i} \left(k_i \frac{\partial u}{\partial x_i} \right) \overline{w} \, \mathrm{d}x_1 \mathrm{d}x_2 \right| + \\
+ \left| \iint_{\omega_h} \sum_{i=1}^{2} k_i \frac{\partial u}{\partial x_i} \frac{\partial \overline{w}}{\partial x_i} \, \mathrm{d}x_1 \mathrm{d}x_2 \right| + \\
+ \left| \iint_{\tau_h} \left(\sum_{i=1}^{2} \frac{\partial}{\partial x_i} \left(\widetilde{k}_i \frac{\partial \widetilde{u}}{\partial x_i} \right) + \widetilde{f} \right) w \, \mathrm{d}x_1 \mathrm{d}x_2 \right|.
\end{aligned} \tag{3.33}$$

Proof. Using the definitions of $\widetilde{a}_h(\widetilde{u}, w)$, $\widetilde{L}_h(w)$ and Green's theorem we obtain

$$\begin{aligned}
\widetilde{a}_h(\widetilde{u}, w) - \widetilde{L}_h(w) = \iint_{\Omega_h} \sum_{i=1}^{2} \widetilde{k}_i \frac{\partial \widetilde{u}}{\partial x_i} \frac{\partial w}{\partial x_i} \, \mathrm{d}x_1 \mathrm{d}x_2 - \\
- \widetilde{L}_h^{\Omega}(w) - \widetilde{L}_h^{\Gamma}(w) = \int_{\Gamma_{2h}} \sum_{i=1}^{2} \widetilde{k}_i \frac{\partial \widetilde{u}}{\partial x_i} n_i(\Omega_h) w \, \mathrm{d}s - \\
- \iint_{\Omega_h} \left(\sum_{i=1}^{2} \frac{\partial}{\partial x_i} \left(\widetilde{k}_i \frac{\partial \widetilde{u}}{\partial x_i} \right) + \widetilde{f} \right) w \, \mathrm{d}x_1 \mathrm{d}x_2 - \widetilde{L}_h^{\Gamma}(w).
\end{aligned}$$

To the right-hand side let us add zero in the form

$$-\int_{\Gamma_2}\sum_{i=1}^{2}k_i\frac{\partial u}{\partial x_i}n_i(\Omega)\overline{w}\,\mathrm{d}s+L^{\Gamma}(\overline{w})=0.$$

If we denote $\Delta=\overline{T}^{\,\mathrm{id}}-T$ and use Lemma 3.2 then we can write

$$\widetilde{a}_h(\widetilde{u},w)-\widetilde{L}_h(w)=-\sum_{\Delta\subset\omega_h}\int_{\partial\Delta}\sum_{i=1}^{2}k_i\frac{\partial u}{\partial x_i}n_i(\Delta)\overline{w}\,\mathrm{d}s-$$
$$-\iint_{\tau_h}\left(\sum_{i=1}^{2}\frac{\partial}{\partial x_i}\left(\widetilde{k}_i\frac{\partial\widetilde{u}}{\partial x_i}\right)+\widetilde{f}\right)w\,\mathrm{d}x_1\mathrm{d}x_2+L^{\Gamma}(\overline{w})-\widetilde{L}_h^{\Gamma}(w).$$

Transforming the first term on the right-hand side by means of Green's theorem we obtain (3.33). □

The third term on the right-hand side is most disagreeable. It is estimated in the following lemma:

Lemma 3.21. *Let $u\in H^2(\Omega)$ and $\widetilde{k}_i\in W^{1,\infty}(\Omega)$ $(i=1,2)$. Then*

$$\left|\iint_{\omega_h}\sum_{i=1}^{2}k_i\frac{\partial u}{\partial x_i}\frac{\partial\overline{w}}{\partial x_i}\,\mathrm{d}x_1\mathrm{d}x_2\right|\le Ch^2\frac{\sqrt{m}}{\varrho}\max_{i=1,2}\|k_i\|_{1,\infty,\Omega}\|u\|_{2,\Omega}\|w\|_{1,\Omega_h}.\quad(3.34)$$

If in addition

$$u\in H^2(\Omega)\cap W^{1,\infty}(\Omega),\quad(3.35)$$

then

$$\left|\iint_{\omega_h}\sum_{i=1}^{2}k_i\frac{\partial u}{\partial x_i}\frac{\partial\overline{w}}{\partial x_i}\,\mathrm{d}x_1\mathrm{d}x_2\right|\le Ch^2\sqrt{\frac{m}{\varrho}}\max_{i=1,2}\|k_i\|_{1,\infty,\Omega}\|u\|_{1,\infty,\Omega}\|w\|_{1,\Omega_h}.\quad(3.36)$$

Proof. We have

$$\left|\iint_{\omega_h}\sum_{i=1}^{2}k_i\frac{\partial u}{\partial x_i}\frac{\partial\overline{w}}{\partial x_i}\,\mathrm{d}x_1\mathrm{d}x_2\right|\le\max_{i=1,2}\|k_i\|_{1,\infty,\Omega}|u|_{1,\omega_h}|\overline{w}|_{1,\omega_h}.\quad(3.37)$$

Assumption (3.35) gives

$$|u|_{1,\omega_h}\le Ch|u|_{1,\infty,\Omega}.\quad(3.38)$$

Let us denote $\Delta=T^{\mathrm{id}}-\overline{T}$. Then

$$|\overline{w}|^2_{1,\omega_h}=\sum_{\Delta\subset\omega_h}\mathrm{mes}_2\Delta|(\nabla w\,|_T)|^2\le C\sum_{\Delta\subset\omega_h}h_T^3|(\nabla w\,|_T)|^2=$$
$$=C\frac{m}{\varrho}\sum_{\Delta\subset\omega_h}h_T^3\frac{\varrho}{m}|(\nabla w\,|_T)|^2\le C\frac{m}{\varrho}h^2\sum_{\Delta\subset\omega_h}|w|^2_{1,T}\le C\frac{m}{\varrho}h^2|w|^2_{1,\Omega_h}$$

because

$$\frac{\varrho}{m} h_T |(\nabla w \,|_T)|^2 \le C|w|^2_{1,T}.$$

Hence

$$|w|_{1,\omega_h} \le Ch\sqrt{\frac{m}{\varrho}}|w|_{1,\Omega_h}. \tag{3.39}$$

Combining (3.37)–(3.39) we obtain (3.36). For the proof of (3.34) see [21]. □

Estimate (3.36) cannot be improved. Thus, if we want to obtain the rate of convergence $O(h)$ we must assume that

$$C_1 h^2 \le \frac{\varrho}{m} \quad (C_1 > 0). \tag{3.40}$$

Assumption (3.40) is also necessary in estimating the first term on the right-hand side of (3.33) if we want to obtain in it the rate of convergence $O(h)$ (see [21]).

3.F The final result

All preceding results yield the following theorem:

Theorem 3.22. *Let us consider a division $\mathcal{D}_h^T$ (or $\mathcal{D}_h^A$). Let $u \in H^2(\Omega)$, $\widetilde{f} \in W^{1,\infty}(\widetilde{\Omega})$, $\widetilde{k}_i \in W^{1,\infty}(\widetilde{\Omega})$ $(i = 1, 2)$. Let assumptions $(3.8)_{3,4}$, (3.9), (3.40) and assumptions concerning the degrees of precision of the quadrature formulas (see Theorems 3.9, 3.13, 3.15, 3.16 and 3.18) be satisfied. Then*

$$\|\widetilde{u} - u_h\|_{1,\Omega_h} \le \frac{C}{\sqrt{\varrho}} h\,, \tag{3.41}$$

where the constant C does not depend on u, ϱ, m, h and the division $\mathcal{D}_h^T$ (or $\mathcal{D}_h^A$).

If in addition $u \in W^{1,\infty}(\Omega)$ (see (3.35)) then

$$\|\widetilde{u} - u_h\|_{1,\Omega_h} \le Ch\,, \tag{3.42}$$

where again the constant C does not depend on u, ϱ, m, h and the division $\mathcal{D}_h^T$ (or $\mathcal{D}_h^A$).

Theorem 3.23. *If we use divisions $\mathcal{D}_h^K$ for the definition of the spaces X_h then the assertions of Theorem 3.22 remain without changes.*

For the proof see [21, pp. 390–392].

Now we mention results in the case of the boundary value problem of equation (3.1) with boundary conditions opposite to conditions (3.2) and (3.3):

$$u = 0 \quad \text{on } \Gamma_2, \tag{3.43}$$

$$\sum_{i=1}^{2} k_i \frac{\partial u}{\partial x_i} n_i(\Omega) = q \quad \text{on } \Gamma_1. \tag{3.44}$$

In this case Problem 3.4 and all results up to relation (3.32) inclusive remain without changes, except for Lemma 3.2, where (3.3) is replaced by (3.44), and except for the definition of $\mathcal{D}_h^A$: we divide into two triangles each quadrilateral $P_iP_{i+1}A_1^iA_1^{i+1}$. Doing some additional considerations (see [21, pp. 393–397]) we obtain the following theorems:

Theorem 3.24. *Let the assumptions of Theorem 3.22 be satisfied except for the additional assumption $u \in W^{1,\infty}(\Omega)$ which is substituted by $\widetilde{u} \in W^{1,\infty}(\widetilde{\Omega})$. Then estimates (3.41) and (3.42) are again valid.*

Theorem 3.25. *If we use divisions $\mathcal{D}_h^K$ for the definition of the spaces X_h then the assertions of Theorem 3.24 remain without changes.*

Remark 3.26. Modifying considerations of [12, Chapter 4] we can prove the following regularity results: Let $j \geq 1$. If $k_i \in C^{j-1,1}(\overline{\Omega})$, $f \in W_2^{j-1}(\Omega)$, $q \in C^{j-1,1}(\Gamma_r)$ ($r = 1$ or 2) then $u \in H^{j+1}(\Omega)$. This means that the assumption guaranteeing (3.42) can be satisfied.

4 Composite domains in magnetostatical problems

In this section we restrict ourselves for a greater simplicity to triangular elements. We shall study the situation indicated in Fig. 5, where the circle consists of three subdomains, the middle one being very narrow. We shall see that in such a case requirement (3.40) can be omitted.

Problem 4.1. Let Ω be a simply connected domain with a Lipschitz continuous boundary $\partial\Omega$ such that

$$\overline{\Omega} = \overline{\Omega}^R \cup \overline{\Omega}^A \cup \overline{\Omega}^S$$

where R, S and A stand for rotor, stator and air, respectively, and Ω^R, Ω^S and Ω^A are domains with Lipschitz continuous boundaries. Let

$$V = \{v \in H^1(\Omega) : v = 0 \quad \text{on } \Gamma_1\}, \tag{4.1}$$

$$\left.\begin{aligned} a(w,v) &= \sum_{i=1}^{2} \iint_\Omega \nu(|\nabla w|^2) \frac{\partial w}{\partial x_i} \frac{\partial v}{\partial x_i} \,\mathrm{d}x_1\mathrm{d}x_2, \\ \nu \equiv \nu_0 \text{ in } \Omega^A, &\quad \nu \equiv \nu_0\nu_r^R \text{ in } \Omega^R, \quad \nu \equiv \nu_0\nu_r^S \text{ in } \Omega^S, \end{aligned}\right\} \tag{4.2}$$

$$L(v) = L^\Omega(v) + L^\Gamma(v) = \iint_\Omega vf \,\mathrm{d}x_1\mathrm{d}x_2 + \int_{\Gamma_2} vq \,\mathrm{d}s\,, \tag{4.3}$$

where $f \in L_2(\Omega)$, $q \in L_2(\Gamma_2)$. Find $u \in H^1(\Omega)$ such that

$$u - z \in V, \tag{4.4}$$

$$a(u,v) = L(v) \quad \forall v \in V, \tag{4.5}$$

where $z \in W^{1,p}(\Omega)$ ($p > 2$) satisfies $\operatorname{tr} z = \overline{u}$ on Γ_1. (We note that as usual $\partial\Omega = \overline{\Gamma}_1 \cup \overline{\Gamma}_2$, $\Gamma_1 \cap \Gamma_2 = \emptyset$, $\operatorname{mes}_1\Gamma_1 > 0$.) □

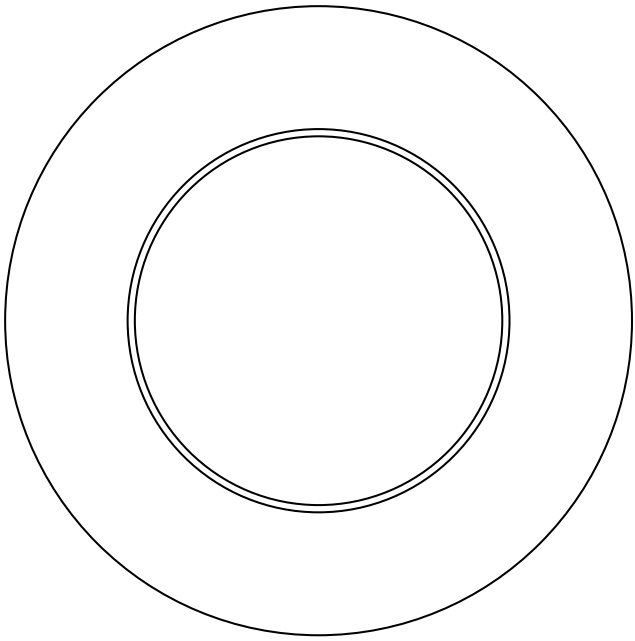


Fig. 5.

Problem 4.1 corresponds to a two-dimensional magnetostatical problem; its connection with Maxwell's equations is explained, for example, in [20] — here we only note that $u = u(x, y)$ has the physical meaning of the z-component of the magnetic potential vector $\overrightarrow{A} = (0, 0, u)$, the positive function $\nu = \nu(s)$ is the magnetic reluctivity, $f = f(x, y)$ is the z-component of the external current density vector $\overrightarrow{J_e} = (0, 0, f)$ and $\overline{u}$ and q are functions appearing on the right-hand sides of the Dirichlet and Neumann boundary conditions, respectively.

We have $\nu_r^M \in C^\infty([0, \infty))$. Using the expression for ν_r^M, which is introduced, e.g., in [10], [11], we can prove (similarly as in [18, Example 33.3]) that there exist positive constants β_1^M, β_2^M $(M = R, S)$ such that

$$\beta_1^M \le \frac{\mathrm{d}}{\mathrm{d}s}(s\nu_r^M(s^2)) \le \beta_2^M \quad \forall s \in [0, \infty), \quad M = R,\, S. \tag{4.6}$$

Property (4.6) has an important consequence: if we integrate (4.6) in $[0, t]$ $(t > 0)$ then we obtain

$$\beta_1^M \le \nu_r^M(t^2) \le \beta_2^M \quad \forall t \in (0, \infty).$$

This result and the continuity of ν_r^M give

$$\beta_1^M \le \nu_r^M(s^2) \le \beta_2^M \quad \forall s \in [0, \infty). \tag{4.7}$$

Making use of (4.6), (4.7) we can prove that Problem 4.1 has a unique solution $u \in H^1(\Omega)$ (see [20, Lemma 2 and Theorem 3]).

In order to obtain a discrete solution of Problem 4.1 by the finite element method we triangulate the closed domain $\overline{\Omega}$ in such a way that the triangulation $\mathcal{T}_h$ of $\overline{\Omega}$ is a union of triangulations $\mathcal{T}_h^R$, $\mathcal{T}_h^S$ and $\mathcal{T}_h^A$ of $\overline{\Omega}^R$, $\overline{\Omega}^S$ and $\overline{\Omega}^A$, respectively. On the contrary to the standard theories we assume that the minimum angle condition

$$\vartheta_h^M := \min_{\overline{T} \in \mathcal{T}_h^M} \vartheta_T \geq \vartheta_0 > 0 \quad \forall h \in (0, h_0), \tag{4.8}$$

where ϑ_T is the magnitude of the minimum angle of $\overline{T}$, is satisfied only for $M = R, S$. As the domain Ω^A is very narrow the triangulations $\mathcal{T}_h^A$ are supposed to satisfy the *maximum angle condition*

$$\gamma_T \leq \gamma_0 < \pi \quad \forall \overline{T} \in \mathcal{T}_h^A, \quad \forall h \in (0, h_0), \tag{4.9}$$

where γ_T is the magnitude of the maximum angle of $\overline{T}$.

Assumption 4.2. In order to simplify our considerations we shall assume that Ω^S, Ω^A and Ω^R are such that $\partial\Omega^S = \partial K_1 \cup \partial K_2$, $\partial\Omega^A = \partial K_2 \cup \partial K_3$ and $\partial\Omega^R = \partial K_3$, where ∂K_1, ∂K_2 and ∂K_3 are circles with the same center S_0 and radii R_1, R_2 and R_3, respectively, which satisfy the relations

$$R_1 > R_2 > R_3 > 0, \quad R_3 = R_2 - \varrho, \quad R_1 - R_2 \gg \varrho, \quad R_3 \gg \varrho$$

where $\varrho > 0$ is fixed (see Fig. 5). □

The discrete problem is formulated in a standard way. We define the spaces

$$X_h = \{v \in C(\overline{\Omega}_h) : v\,|_T = \text{ a linear polynomial} \quad \forall T \in \mathcal{T}_h\}, \tag{4.10}$$

$$V_h = \{v \in X_h : v = 0 \text{ on } \overline{\Gamma}_{1h}\} \tag{4.11}$$

and the set

$$W_h = \{v \in X_h : v(P_i) = \overline{u}(P_i) \quad \forall P_i \in \sigma_h \cap \overline{\Gamma}_1\}, \tag{4.12}$$

where $\overline{\Omega}_h$ is the union of the closed triangles $\overline{T} \in \mathcal{T}_h$, $\overline{\Gamma}_{1h}$ is the part of $\partial\Omega_h$ approximating $\overline{\Gamma}_1$ and σ_h is the set of all nodes of $\mathcal{T}_h$. Further we set

$$a_h(v, w) = \sum_{M=R,A,S} \sum_{i=1}^{2} \iint_{\Omega_h^M} \nu^M(|\nabla v|^2) \frac{\partial v}{\partial x_i} \frac{\partial w}{\partial x_i} \, \mathrm{d}x_1 \mathrm{d}x_2 \quad \forall v, w \in H^1(\Omega_h), \tag{4.13}$$

which gives

$$a_h(v, w) = \sum_{M=R,S} \sum_{\overline{T} \in \mathcal{T}_h^M} \sum_{i=1}^{2} \nu_0 \nu_r^M(|\nabla(v\,|_T)|^2) \left.\frac{\partial v}{\partial x_i}\right|_T \left.\frac{\partial w}{\partial x_i}\right|_T \mathrm{mes}_2 T +$$

$$+ \sum_{\overline{T} \in \mathcal{T}_h^A} \sum_{i=1}^{2} \nu_0 \left.\frac{\partial v}{\partial x_i}\right|_T \left.\frac{\partial w}{\partial x_i}\right|_T \mathrm{mes}_2 T \quad \forall v, w \in X_h. \tag{4.14}$$

Finally, we set

$$L_h(v) = L_h^{\Omega}(v) + L_h^{\Gamma}(v) \quad \forall v \in X_h\,, \tag{4.15}$$

where $L_h^{\Omega}(v)$ and $L_h^{\Gamma}(v)$ are the approximations of the forms

$$\widetilde{L}_h^{\Omega}(v) = \iint_{\Omega_h} v f \,\mathrm{d}x_1 \mathrm{d}x_2\,, \qquad \widetilde{L}_h^{\Gamma}(v) = \int_{\Gamma_{2h}} q_h v \,\mathrm{d}s \tag{4.16}$$

by means of quadrature formulas of first degree of precision. (Details and the definition of the function q_h are introduced in [4], [18] and [21].) Using (4.10)–(4.15) we define:

Problem 4.3. Find $u_h \in W_h$ such that

$$a_h(u_h, v) = L_h(v) \quad \forall v \in V_h. \tag{4.17}$$

It can be proved similarly as in [4], [17] or [18] that every discrete problem has a unique solution u_h. The main result of this section is the following theorem.

Theorem 4.4. *Let the solution $u \in H^1(\Omega)$ of Problem 4.1 satisfy*

$$u_M \in H^2(\Omega^M) \quad (M = R,\, S,\, A), \tag{4.18}$$

where $u_M := u\,\big|_{\Omega^M}$. Let $f \in W^{1,\infty}(\Omega)$ and $q \in C^1(\overline{\Gamma}_2)$. Then we have for all $h \in (0, h_0)$

$$\|u_h - u\|_{1,\Omega_h} \le C\,h\,, \tag{4.19}$$

where $u \in H^1(\Omega)$ is the solution of Problem 4.1, $\|\cdot\|_{1,\Omega_h}$ is the norm in the space $H^1(\Omega_h)$ and C is a constant independent of $h := \max\limits_{T \in \mathcal{T}_h} h_T$ and ϱ.

Assumption (4.18) is guaranteed if $\Gamma_2 = \emptyset$ and $\overline{u}$ is sufficiently smooth.

The proof of Theorem 4.4 is based on the following abstract error estimate which can be proved in the same way as [4, Theorem 3.3.1] or [18, Theorem 38.5]:

$$\|u - u_h\|_{1,\Omega_h} \le C\left\{\inf_{v \in W_h} \|u - v\|_{1,\Omega_h} + \sup_{\substack{w \in V_h \\ w \neq 0}} \frac{|a_h(u, w) - L_h(w)|}{\|w\|_{1,\Omega_h}}\right\}, \tag{4.20}$$

where the constant C does not depend on h and ϱ. The two terms on the right-hand side of (4.20) will be estimated in Theorems 4.8 and 4.13.

The following lemma is a reformulation of Lemma 3.6:

Lemma 4.5. *Let ∂K_4 be the circle with the center S_0 and radius $R_4 = R_2 - 2\varrho$, where ϱ is the same as in Assumption 4.2. Let $\widetilde{\Omega}^S$, $\widetilde{\Omega}^A$ be bounded domains such that $\partial\widetilde{\Omega}^S = \partial K_1 \cup \partial K_4$, $\partial\widetilde{\Omega}^A = \partial K_2 \cup \partial K_4$. There exist linear and bounded extension operators $E_M : H^2(\Omega^M) \to H^2(\widetilde{\Omega}^M)$ $(M = S, A)$ such that the constant C_M appearing in the inequality*

$$\|E_M(v)\|_{2,\widetilde{\Omega}^M} \le C_M \|v\|_{2,\Omega^M} \quad \forall v \in H^2(\Omega^M)$$

does not depend on R_2/ϱ and v.

Remark 4.6. As Lemma 4.5 is used in the proof of Theorem 4.8 the polygonal domains Ω_h^A must be situated between the circles ∂K_2 and ∂K_4. We derive now the expression for the minimum number of vertices of such a polygonal domain in the case $\varrho/R_2 < 10^{-1}$.

Let A_1 be an arbitrary point of the circle ∂K_2 and let t be one of the two tangents to the circle ∂K_3 which pass through the point A_1. Let $B = t \cap \partial K_3$, $A_2 = \{t \cap \partial K_2\} - \{A_1\}$. If $\varrho/R_2 < 10^{-1}$ then we can neglect the terms depending on ϱ^3 and find

$$d_1 = \operatorname{dist}(A_1, B) = (2\varrho R_2 - \varrho^2)^{1/2}, \quad d_2 = \operatorname{dist}(A_1, A_2) = 2d_1.$$

Let us approximate ∂K_2 by a regular polygon with vertices $P_1, \dots, P_n$ where

$$n = n_2 = \left[\frac{2\pi R_2}{d_2}\right] + 1 = \left[\frac{\pi R_2}{(2\varrho R_2 - \varrho^2)^{1/2}}\right] + 1.$$

Let the vertices $Q_1, \dots, Q_n$ of the polygon ∂K_3^h approximating ∂K_3 be obtained in the following way: Q_i is the intersection of the segment S_0P_i with ∂K_3.

For example, if $\varrho = 1\,\text{mm}$ and $R_2 = 50\,\text{mm}$ then $n_2 = 16$. This is a surprisingly small number. Of course, it is better to use the relation

$$n = n_1 = \left[\frac{2\pi R_2}{d_1}\right] + 1.$$

In the case $\varrho = 1\,\text{mm}$, $R_2 = 50\,\text{mm}$ we have $n_1 = 32$.

If we divide every quadrilateral $P_iP_{i+1}Q_iQ_{i+1}$ into two triangles we obtain a triangulation which satisfies (from a practical point of view) the maximum angle condition only: For $n = n_1$ the minimum angle is less than 6 degrees and for $n = n_2$ less than 3 degrees. □

Lemma 4.7. *If the solution $u \in H^1(\Omega)$ of Problem 4.1 satisfies assumption (4.18) then*

$$N_i^M(u) := \nu^M(|\nabla u_M|^2)\frac{\partial u_M}{\partial x_i} \in H^1(\Omega^M) \quad (M = R, S, A). \tag{4.21}$$

Consequently,

$$\sum_{i=1}^{2} \frac{\partial}{\partial x_i}\left(\nu^M(|\nabla u_M|^2)\frac{\partial u_M}{\partial x_i}\right) + f_M = 0 \quad \text{a.e. in } \Omega^M \quad (M = R, S, A), \tag{4.22}$$

$$q = \sum_{i=1}^{2} \nu^S(|\nabla u_S|^2)\frac{\partial u_S}{\partial x_i} n_i(\Omega^S) \quad \text{a.e. on } \Gamma_2, \tag{4.23}$$

where $f_M = f|_{\Omega^M}$ *and the symbols* $n_i(G)$ $(i = 1, 2)$ *denote the components of the unit outward normal to* ∂G. *Finally,*

$$\nu_r^M \frac{\partial u_M}{\partial n}\bigg|_{\partial K_j} = \frac{\partial u_A}{\partial n}\bigg|_{\partial K_j} \quad \text{a.e. on } \partial K_j \quad (M = R, S), \tag{4.24}$$

where $j = 2$ *for* $M = S$ *and* $j = 3$ *for* $M = R$ *and* $\partial/\partial n$ *is the normal derivative (the orientation of* n *can be chosen arbitrarily).*

Theorem 4.8. *Under the assumptions of Theorem 4.4 we have*

$$\inf_{v \in W_h} \|u - v\|_{1,\Omega_h} \le C\, h \left\{ \sum_{M=R,S} (1 + \sup|\nu_r^M|)\|u_M\|_{2,\Omega^M} + \|u_A\|_{2,\Omega^A} \right\}, \tag{4.25}$$

where the constant C *does not depend on both* h *and* ϱ.

For the proofs of Lemma 4.7 and Theorem 4.8 see [20, Lemma 12 and Theorem 13].

Notation 4.9. a) We denote

$$\omega_h^M := \Omega^M - \overline{\Omega}_h^M, \quad \tau_h^M := \Omega_h^M - \overline{\Omega}^M.$$

b) The natural extension $\overline{w}_M$ of $w_M := w\,\big|_{\Omega_h^M}$ from $\overline{\Omega}_h^M$ onto $\overline{\Omega}_h^M \cup \overline{\Omega}^M$ is the function $\overline{w}_M : \overline{\Omega}_h^M \cup \overline{\Omega}^M \to R^1$ satisfying $\overline{w}_M = w_M$ on $\overline{\Omega}_h^M$ and

$$\overline{w}_M\,\big|_{T^{\mathrm{id}}} = p\,\big|_{T^{\mathrm{id}}} \quad \text{on} \quad T^{\mathrm{id}} \supset T,$$

where p is the polynomial of first degree satisfying $p\,\big|_T = w\,\big|_T$ and T^{id} is the ideal curved triangle associated with T (it is also called the exact curved triangle). (For more detail see [18] or [4].)

c) The natural extension $\overline{w}$ of $w \in X_h$ is the function $\overline{w} : \Omega \to R^1$ such that $\overline{w} = w$ on $\overline{\Omega}_h$ and $\overline{w} = \overline{w}_S$ on ω_h^S.

Lemma 4.10. *We have*

$$\|v\|_{0,\tau_h^M} \le C(h\|v\|_{0,\partial K_{i+1}} + h^2|v|_{1,\tau_h^M}) \quad \forall v \in H^1(\tau_h^M) \quad (M = S,\, A),$$

where $i = 1$ *and* $i = 2$ *for* $M = S$ *and* $M = A$, *respectively, and where the constant* C *does not depend on both* h *and* ϱ.

Lemma 4.10 follows from the proof of [18, Lemma 28.3].

Lemma 4.11. *We have for all* $w \in X_h$

$$\|\overline{w}_M\|_{0,\varepsilon_h^M} \le C\,h\|w\|_{1,\Omega_h} \quad (\varepsilon = \tau,\omega;\ M = R,S), \tag{4.26}$$

$$|\overline{w}_M|_{1,\varepsilon_h^M} \le C\,h^{1/2}|w|_{1,\Omega_h} \quad (\varepsilon = \tau,\omega;\ M = R,S). \tag{4.27}$$

Proof. As $\mathcal{T}_h^R$, $\mathcal{T}_h^S$ satisfy the minimum angle condition estimates (4.26), (4.27) follow from [4, Lemma 3.3.12].

Lemma 4.12. *We have for all* $w \in X_h$

$$|L_h^{\Omega}(w) - \widetilde{L}_h^{\Omega}(w)| \le C\,h\|f\|_{1,\infty,\Omega}\|w\|_{1,\Omega_h}, \tag{4.28}$$

$$|L_h^{\Gamma}(w) - \widetilde{L}_h^{\Gamma}(w)| \le C\,h(\mathrm{mes}_1\Gamma_2)^{1/2}|q|_{1,\infty,\Gamma_2}\|w\|_{1,\Omega_h}, \tag{4.29}$$

$$|\widetilde{L}_h^{\Gamma}(w) - L^{\Gamma}(\overline{w})| \le C\,h^{3/2}\|q\|_{0,\Gamma_2}\|w\|_{1,\Omega_h}. \tag{4.30}$$

For the proof of (4.28), (4.29) and (4.30) see, for example, [3, Theorem 4.5.1], [18, Lemma 30.1] and [4, Lemma 3.3.13], respectively.

Theorem 4.13. *Under the assumptions of Theorem 4.4 we have for all* $w \in V_h$

$$\begin{aligned} |a_h(u,w) - L_h(w)| \le C\,h\{&\|f\|_{1,\infty,\Omega} + (\mathrm{mes}_1\Gamma_2)^{1/2}\|q\|_{1,\infty,\Gamma_2} + \\ &+ (1+\sup|\nu_r^S|)\|u_S\|_{2,\Omega^S} + \sum_{M=A,R}\|u_M\|_{2,\Omega^M} + \\ &+ \sum_{i=1}^{2}\left\|\partial N_i^S(u)/\partial x_i\right\|_{0,\Omega^S}\}\|w\|_{1,\Omega_h}, \end{aligned} \tag{4.31}$$

where $N_i^S(u)$ *is defined in* (4.21).

Proof. Instead of S, A and R we shall write 1, 2 and 3, respectively. We have

$$|a_h(u,w) - L_h(w)| \le |a_h(u,w) - \widetilde{L}_h(w)| + |\widetilde{L}_h(w) - L_h(w)|, \tag{4.32}$$

where

$$\widetilde{L}_h(w) = \widetilde{L}_h^{\Omega}(w) + \widetilde{L}_h^{\Gamma}(w). \tag{4.33}$$

After a longer computation we obtain (see [20, pp. 413–415])

$$a_h(u,w) - \widetilde{L}_h(w) = D_1 + \sum_{j=1}^{2}(D_2^{(j,j)} - D_2^{(j+1,j)}) - D_3 - D_4\,, \tag{4.34}$$

where

$$D_1 = L^{\Gamma}(\overline{w}) - \widetilde{L}_h^{\Gamma}(w),$$

$$D_2^{(k,j)} = \sum_{i=1}^{2} \iint_{\tau_h^j} \nu^k(|\nabla u_{j+1}|^2) \frac{\partial u_{j+1}}{\partial x_i} \frac{\partial w}{\partial x_i} \,\mathrm{d}x_1 \mathrm{d}x_2,$$

$$D_3 = \sum_{i=1}^{2} \iint_{\omega_h^1(2)} \overline{w} \frac{\partial}{\partial x_i} \left(\nu^1(|\nabla u_1|^2) \frac{\partial u_1}{\partial x_i} \right) \mathrm{d}x_1 \mathrm{d}x_2,$$

$$D_4 = \sum_{i=1}^{2} \iint_{\omega_h^1(2)} \nu^1(|\nabla u_1|^2) \frac{\partial u_1}{\partial x_i} \frac{\partial \overline{w}}{\partial x_i} \,\mathrm{d}x_1 \mathrm{d}x_2,$$

where $\omega_h^1(2)$ denotes the part of ω_h^1 which is adjacent to Γ_2.

The estimate of $|D_1|$ is given in (4.30). The term $D_2^{(k,j)}$ is of the same type as the term appearing in Lemma 3.21. However, the presence of the domains $\Omega^R \equiv \Omega^3$, $\Omega^S \equiv \Omega^1$ enable us to avoid requirement (3.40). It follows from (4.7) that

$$|D_2^{(k,j)}| \le K |u_{j+1}|_{1,\tau_h^j} |w|_{1,\tau_h^j}. \tag{4.35}$$

As $u_{j+1} \in H^2(\Omega^{j+1})$ we have by Lemma 4.10

$$|u_{j+1}|_{1,\tau_h^j} \le C \sum_{i=1}^{2} \left(h \left\| \frac{\partial u_{j+1}}{\partial x_i} \right\|_{0,\partial K_{j+1}} + h^2 \left| \frac{\partial u_{j+1}}{\partial x_i} \right|_{1,\tau_h^j} \right). \tag{4.36}$$

The trace theorem yields

$$\left\| \frac{\partial u_3}{\partial x_i} \right\|_{0,\partial K_3} \le C \|u_3\|_{2,\Omega^3}. \tag{4.37}$$

Owing to the fact that $u \in C(\overline{\Omega})$ we have

$$u_1 \big|_{\partial K_2} = u_2 \big|_{\partial K_2}.$$

This relation implies that

$$\frac{\partial u_1}{\partial t} = \frac{\partial u_2}{\partial t} \quad \text{a.e. on } \partial K_2,$$

where $\partial/\partial t$ is the tangential derivative. Combining this result with (4.24) (where $j = 2$) and using the trace theorem on Ω^1 we derive

$$\left\| \frac{\partial u_2}{\partial x_i} \right\|_{0,\partial K_2} \le C(1 + \sup |\nu_r^1|) \|u_1\|_{2,\Omega^1}. \tag{4.38}$$

Estimates (4.35)–(4.38) give

$$\sum_{j=1}^{2} (|D_2^{(j,j)}| + |D_2^{(j+1,j)}|) \le Ch \left\{ (1 + \sup |\nu_r^1|) \|u_1\|_{2,\Omega^1} + \sum_{j=2}^{3} \|u_j\|_{2,\Omega^j} \right\} |w|_{1,\Omega_h}. \tag{4.39}$$

Relation (4.21), the Schwarz inequality and Lemma 4.11 imply

$$|D_3| \le C\,h\left(\sum_{i=1}^{2}\left\|\frac{\partial}{\partial x_i}\left(\nu^1(|\nabla u_1|^2)\frac{\partial u_1}{\partial x_i}\right)\right\|_{0,\Omega^1}\right)\|w\|_{1,\Omega_h}. \tag{4.40}$$

Finally, as $\mathcal{T}_h^1$ satisfies the minimum angle condition and $u_1 \in H^2(\Omega^1)$ (see (4.18)) we have by (4.7), (4.27), Lemma 4.10 (which holds also for ω_h^1 with ∂K_1 instead of ∂K_{i+1}) and the trace inequality

$$|D_4| \le K|u_1|_{1,\omega_h^1}|\overline{w}|_{1,\omega_h^1} \le C\,h^{3/2}\|u_1\|_{2,\Omega^1}\|w\|_{1,\Omega_h}. \tag{4.41}$$

Relations (4.34), (4.30), (4.39)–(4.41) give the bound of the first term on the right-hand side of (4.32). The estimate of the second term on the right-hand side of (4.32) follows from Lemma 4.12. Hence we obtain (4.31). □

Theorem 4.4 follows now from (4.20) and Theorems 4.8 and 4.13.

5 General convergence theorem

On the contrary to Section 3 we shall assume $u \in H^1(\Omega)$ only and we shall prove the convergence (without any rate of convergence) under a stronger assumption than (3.40):

$$C_1 h^{2-\delta} \le \frac{\varrho}{m} \le C_2 h^{2-\delta}\,, \tag{5.1}$$

where

$$0 < \delta < 1 \tag{5.2}$$

is a given number which can be arbitrarily small and $C_1[\mathrm{m}^{1-\delta}]$, $C_2[\mathrm{m}^{1-\delta}]$ are positive constants. The abstract error estimate has in the case $u \in H^1(\Omega)$ the form:

Theorem 5.1. *Let condition (3.17) be satisfied. Then Problem 3.4 has a unique solution $u_h \in V_h$ and we have*

$$C_0^{-1}\|\widetilde{u} - u_h\|_{1,\Omega_h} \le \inf_{v\in V_h}\left(\|v-\widetilde{u}\|_{1,\Omega_h} + \sup_{\substack{w\in V_h\\ w\neq 0}}\frac{|a_h(v,w)-\widetilde{a}_h(v,w)|}{\|w\|_{1,\Omega_h}}\right) +$$
$$+\sup_{\substack{w\in V_h\\ w\neq 0}}\frac{|\widetilde{L}_h^{\Omega}(w)-L_h^{\Omega}(w)|}{\|w\|_{1,\Omega_h}} + \sup_{\substack{w\in V_h\\ w\neq 0}}\frac{|\widetilde{L}_h^{\Gamma}(w)-L_h^{\Gamma}(w)|}{\|w\|_{1,\Omega_h}} + \sup_{\substack{w\in V_h\\ w\neq 0}}\frac{|\widetilde{a}_h(\widetilde{u},w)-\widetilde{L}_h(w)|}{\|w\|_{1,\Omega_h}}, \tag{5.3}$$

where C_0 is a positive constant, $u \in H^1(\Omega)$ is the solution of Problem 3.1 and $\widetilde{u} = E(u)$ with $E : H^1(\Omega) \to H^1(\widetilde{\Omega})$ (see Lemma 3.6 where $k = 1$).

In what follows we restrict ourselves to the case of triangular elements with linear polynomials. First we generalize interpolation results for Zlámal's simplest ideal triangular finite element (see [25] and also [18]).

Let $\overline{T} \in \mathcal{D}_h^T$ be an arbitrary triangle with two vertices lying on $\partial\Omega$. We shall denote them by $P_2(x_1^{(2)}, x_2^{(2)})$, $P_3(x_1^{(3)}, x_2^{(3)})$ in such a way that

$$\operatorname{dist}(P_1, P_2) = \frac{\varrho}{m}, \tag{5.4}$$

$P_1(x_1^{(1)}, x_2^{(1)})$ being the vertex lying in Ω. Thus the smallest angle α_T of $\overline{T}$, which tends to zero with $h \to 0$, lies at P_3. The angles lying at P_1 and P_2 will be denoted by β_T and γ_T, respectively. Both these angles tend to $\pi/2$ with $h \to 0$.

Setting

$$\overline{x}_2 = x_1^{(2)} - x_1^{(1)},\ \overline{x}_3 = x_1^{(3)} - x_1^{(1)},\ \overline{y}_2 = x_2^{(2)} - x_2^{(1)},\ \overline{y}_3 = x_2^{(3)} - x_2^{(1)}$$

we can write the transformation, which maps the triangle $\overline{T}_0$ with vertices $R_1(0,0)$, $R_2(1,0)$ and $R_3(0,1)$ one-to-one onto $\overline{T}$, in the form

$$\begin{aligned} x_1 &= x_1^{(0)}(\xi_1, \xi_2) \equiv x_1^{(1)} + \overline{x}_2\xi_1 + \overline{x}_3\xi_2, \\ x_2 &= x_2^{(0)}(\xi_1, \xi_2) \equiv x_2^{(1)} + \overline{y}_2\xi_1 + \overline{y}_3\xi_2. \end{aligned} \tag{5.5}$$

We have for the triangles lying along $\partial\Omega$

$$2\operatorname{mes}_2 T = \operatorname{dist}(P_1, P_2)\operatorname{dist}(P_2, P_3)\sin\gamma_T.$$

From here, from (5.1), (5.4) and from the maximum angle condition we easily obtain

$$C_3 h_T^{3-\delta} \le \operatorname{mes}_2 T \le C_4 h_T^{3-\delta}, \tag{5.6}$$

h_T being the length of the greatest side of $\overline{T}$ and C_3, C_4 positive constants.

Now we remind some results introduced in [18, Section 22]. Let λ_h and λ be the segment P_2P_3 and the part of $\partial\Omega$ approximated by P_2P_3, respectively. Let

$$x_1 = \varphi_\lambda(\xi_2), \quad x_2 = \psi_\lambda(\xi_2), \quad \xi_2 \in [0,1], \tag{5.7}$$

be a parametric representation of λ defined on $[0,1]$ with the property

$$\varphi_\lambda(0) = x_1^{(2)},\ \varphi_\lambda(1) = x_1^{(3)},\ \psi_\lambda(0) = x_2^{(2)},\ \psi_\lambda(1) = x_2^{(3)}.$$

We define the functions $\Phi_\lambda(\xi_2)$, $\Psi(\xi_2)$ on $[0,1]$ by

$$\begin{aligned} \Phi_\lambda(\xi_2) &= [\varphi_\lambda(\xi_2) - x_1^{(2)} - \overline{x}_{32}\xi_2]/(1-\xi_2), \quad \xi_2 \in [0,1), \\ \Phi_\lambda(1) &= -\varphi_\lambda'(1) + \overline{x}_{32}, \quad \Phi_\lambda^{(j)}(1) = -\frac{1}{j+1}\varphi_\lambda^{(j+1)}(1), \\ \Psi_\lambda(\xi_2) &= [\psi_\lambda(\xi_2) - x_2^{(2)} - \overline{y}_{32}\xi_2]/(1-\xi_2), \quad \xi_2 \in [0,1), \\ \Psi_\lambda(1) &= -\psi_\lambda'(1) + \overline{y}_{32}, \quad \Psi_\lambda^{(j)}(1) = -\frac{1}{j+1}\psi_\lambda^{(j+1)}(1), \end{aligned}$$

where $\overline{x}_{32} = x_1^{(3)} - x_1^{(2)}$, $\overline{y}_{32} = x_2^{(3)} - x_2^{(2)}$. If $\varphi_\lambda, \psi_\lambda \in C^{(n+1)}([0,1])$ then, according to [18, Section 22], $\Phi_\lambda, \Psi_\lambda \in C^n([0,1])$ and

$$\begin{aligned} &\Phi_\lambda(\xi_2) = O(h_T^2), \quad \Phi_\lambda^{(j)}(\xi_2) = O(h_T^{j+1}), \quad \xi_2 \in [0,1], \\ &\Psi_\lambda(\xi_2) = O(h_T^2), \quad \Psi_\lambda^{(j)}(\xi_2) = O(h_T^{j+1}), \quad \xi_2 \in [0,1], \end{aligned} \tag{5.8}$$

where $j = 1, \ldots, n$. The symbol $\overline{T}_\lambda^{\mathrm{id}}$ will denote the curved triangle with two straight sides P_1P_2, P_1P_3 and the curved side λ.

Theorem 5.2. *Let the boundary $\partial\Omega$ of the domain Ω be piecewise of class C^{k+1}. Then for $h \in (0, h_0)$, where h_0 is sufficiently small, we have:*

a) The transformation

$$\begin{aligned} x_1 = x_1^\lambda(\xi_1,\xi_2) &\equiv x_1^{(1)} + \overline{x}_2\xi_1 + \overline{x}_3\xi_2 + \xi_1\Phi_\lambda(\xi_2), \\ x_2 = x_2^\lambda(\xi_1,\xi_2) &\equiv x_2^{(1)} + \overline{y}_2\xi_1 + \overline{y}_3\xi_2 + \xi_1\Psi_\lambda(\xi_2) \end{aligned} \tag{5.9}$$

maps one-to-one the reference triangle $\overline{T}_0$, which lies in the ξ_1,ξ_2-plane and has the vertices $R_1(0,0)$, $R_2(1,0)$, $R_3(0,1)$, onto the ideal triangle $\overline{T}_\lambda^{\mathrm{id}}$ with vertices $P_i(x_1^{(i)}, x_2^{(i)})$ ($i = 1,2,3$ - a local notation) and curved side λ, which has parametric equations (5.7), in such a way that

$$R_i \leftrightarrow P_i \quad (i = 1,2,3), \quad R_1R_j \leftrightarrow P_1P_j \quad (j = 2,3), \quad R_2R_3 \leftrightarrow \lambda \tag{5.10}$$

and $T_0 \equiv \operatorname{int}\overline{T}_0 \leftrightarrow \operatorname{int}\overline{T}_\lambda^{\mathrm{id}} \equiv T_\lambda^{\mathrm{id}}$.

b) The Jacobian $J_\lambda(\xi_1,\xi_2)$ of transformation (5.9) is different from zero on $\overline{T}_0$ and it holds for $(\xi_1,\xi_2) \in \overline{T}_0$:

$$C_5 h_T^{3-\delta} \le |J_\lambda(\xi_1,\xi_2)| \le C_6 h_T^{3-\delta} \quad (C_i = \text{const} > 0). \tag{5.11}$$

c) Both mapping (5.9) and its inverse mapping are of class C^k and for $(\xi_1,\xi_2) \in \overline{T}_0$ we have

$$\frac{\partial x_i^\lambda}{\partial \xi_1} = O(h_T^{2-\delta}), \quad \frac{\partial x_i^\lambda}{\partial \xi_2} = O(h_T) \quad (i = 1,2), \tag{5.12}$$

$$\frac{\partial^2 x_i^\lambda}{\partial \xi_j \partial \xi_k} = O(h_T^2) \quad (i,j,k = 1,2), \tag{5.13}$$

$$\frac{\partial \xi_1^\lambda}{\partial x_i} = O(h_T^{-2+\delta}), \quad \frac{\partial \xi_2^\lambda}{\partial x_i} = O(h_T^{-1}) \quad (i = 1,2), \tag{5.14}$$

where

$$\xi_1 = \xi_1^\lambda(x_1,x_2), \quad \xi_2 = \xi_2^\lambda(x_1,x_2) \tag{5.15}$$

is the inverse mapping to mapping (5.9).

d) Let $\widetilde{S}_1$, $\widetilde{S}_2$ be arbitrary points of $\overline{T}_0$ and S_1, S_2 their images in transformation (5.9). Let ε be the distance between $\widetilde{S}_1$, $\widetilde{S}_2$ and let η be the distance between S_1, S_2. Then

$$C_7 \varepsilon h_T^{2-\delta} \le \eta \le C_8 \varepsilon h_T\,, \tag{5.16}$$

where C_7, C_8 are positive constants independent of ε and h_T.

Proof. A) First we prove assertions concerning $J(\xi_1, \xi_2)$. Using the relations

$$|\overline{x}_2| = O(h_T^{2-\delta}),\ |\overline{y}_2| = O(h_T^{2-\delta}),\ |\overline{x}_3| = O(h_T),\ |\overline{y}_3| = O(h_T), \tag{5.17}$$

we obtain from (5.9) and (5.8)

$$J_\lambda(\xi_1, \xi_2) = [\overline{x}_2 + \Phi_\lambda(\xi_2)][\overline{y}_3 + \xi_1 \Psi'_\lambda(\xi_2)] - \\ - [\overline{x}_3 + \xi_1 \Phi'_\lambda(\xi_2)][\overline{y}_2 + \Psi_\lambda(\xi_2)] = 2\mathrm{mes}_2 T + O(h_T^3).$$

This result together with (5.6) imply both $J_\lambda(\xi_1, \xi_2) \neq 0$ on $\overline{T}_0$ and estimates (5.11).

B) The proof of inequalities (5.16) follows the same lines as part (c) of the proof of [18, Theorem 22.4]. Instead of [18, Lemma 22.2] we use the fact that at least one of the estimates

$$|\alpha_1 \overline{x}_2 + \alpha_2 \overline{x}_3| \geq C h_T^{2-\delta}, \quad |\alpha_1 \overline{y}_2 + \alpha_2 \overline{x}_3| \geq C h_T^{2-\delta} \tag{5.18}$$

holds, where α_1, α_2 are real numbers satisfying

$$\alpha_1^2 + \alpha_2^2 = 1. \tag{5.19}$$

If $\alpha_1 = 0$ or $\alpha_2 = 0$ then assertion (5.18) is evident. Let $\alpha_1 \neq 0$, $\alpha_2 \neq 0$. First we consider the case

$$\operatorname{sign} \alpha_1 = \operatorname{sign} \alpha_2. \tag{5.20}$$

Then the expression

$$V_1 = \frac{1}{|\alpha_1 + \alpha_2|}[(\alpha_1 \overline{x}_2 + \alpha_2 \overline{x}_3)^2 + (\alpha_1 \overline{y}_2 + \alpha_2 \overline{y}_3)^2]^{1/2}$$

is the length of the segment $P_1 P_{23}$, where

$$P_{23} = \big((|\alpha_1| x_1^{(2)} + |\alpha_2| x_1^{(3)})/|\alpha_1 + \alpha_2|, (|\alpha_1| x_2^{(2)} + |\alpha_2| x_2^{(3)})/|\alpha_1 + \alpha_2|\big)$$

is a point of the segment $P_2 P_3$. If $\beta_T \leq \pi/2$ then $V_1 > P_1 P_2$. As $P_1 P_2 \geq C h_T^{2-\delta}$, according to (5.1), assertion (5.18) follows because by (5.19) and (5.20) we have $|\alpha_1 + \alpha_2| > 1$.

If $\beta_T > \pi/2$ then $\beta_T = \omega_T$ where ω_T is the maximum angle of T. We have $V_1 \geq d$ where d is the distance of the vertex P_1 from the segment $P_2 P_3$. As α_T is small the angle made by $P_1 P_2$ and the segment of the length d is less than $\omega_T/2$. Hence $d > P_1 P_2 \cos(\omega_T/2)$ and assertion (5.18) follows, according to the maximum angle condition.

Now let

$$\operatorname{sign} \alpha_1 = -\operatorname{sign} \alpha_2 \tag{5.21}$$

and let the point P^* be such that $P_1 = \frac{1}{2}(P_3 + P^*)$. This gives $P^* = (x_1^*, x_2^*) = 2P_1 - P_3 = (2x_1^{(1)} - x_1^{(3)}, 2x_2^{(1)} - x_2^{(3)})$ and

$$V_2 = \frac{1}{|\alpha_1| + |\alpha_2|}[(|\alpha_1|\overline{x}_2 - |\alpha_2|\overline{x}_3)^2 + (|\alpha_1|\overline{y}_2 - |\alpha_2|\overline{y}_3)^2]^{1/2}$$

is the length of the segment $P_1P_{23}^*$, where

$$P_{23}^* = \big((|\alpha_1|x_1^{(2)} + |\alpha_2|x_1^*)/(|\alpha_1| + |\alpha_2|), (|\alpha_1|x_2^{(2)} + |\alpha_2|x_2^*)/(|\alpha_1| + |\alpha_2|)\big)$$

is a point of the segment P_2P^*. Let T^* be the triangle with vertices P_1, P_2, P^*. In T^* the angle at P_1 is equal to $\pi - \beta_T$. If $\pi - \beta_T \le \pi/2$, then $V_2 \ge P_1P_2 \ge Ch_T^{2-\delta}$.

If $\pi - \beta_T > \pi/2$ then $\pi - \beta_T = \omega_T + \alpha_T$, where $\omega_T = \gamma_T$. We have $V_2 \ge d^*$ with d^* the distance of the vertex P_1 from the segment P_2P^*. As the angle α_T^* at P^* is small, we have $d^* > P_1P_2\cos(\omega_T/2 + \alpha_T/2)$ and assertion (5.18) follows, according to the maximum angle condition, because α_T is small and β_T is not small.

C) Setting $\xi_2 = 0$ in (5.9) we obtain a parametric representation of P_1P_2:

$$x_1 = x_1^{(1)} + \overline{x}_2\xi_1, \quad x_2 = x_2^{(1)} + \overline{y}_2\xi_1, \quad \xi_1 \in [0,1].$$

Setting $\xi_1 = 0$ in (5.9) we obtain a parametric representation of P_1P_3:

$$x_1 = x_1^{(1)} + \overline{x}_3\xi_2, \quad x_2 = x_2^{(1)} + \overline{y}_3\xi_2, \quad \xi_2 \in [0,1].$$

Thus segments P_1P_2 and P_1P_3 are images of segments R_1R_2 and R_1R_3, respectively, in transformation (5.9).

Relations $\xi_1 = 1 - t$, $\xi_2 = t$ $(t \in [0,1])$ form a parametric representation of the segment R_2R_3. In this case we obtain from (5.9) and the definitions of the functions Φ_λ, Ψ_λ:

$$x_1 = x_1^\lambda(1-t,t) = \varphi(t), \quad x_2 = x_2^\lambda(1-t,t) = \psi(t), \quad t \in [0,1].$$

This means that the arc λ is the image of the segment R_2R_3 in transformation (5.9).

Consequently, the Jordan curve $\partial T_\lambda^{\mathrm{id}}$ is the image of the Jordan curve ∂T_0 in transformation (5.9).

Owing to inequalities (5.16) mapping (5.9) is injective. As (5.9) is also continuous on $\overline{T}_0$ it is a homeomorphism. A homeomorphism maps the interior of the Jordan curve onto the interior of its image.

If f is a homeomorphism then f is bijective and f^{-1} is continuous. Thus relations (5.10) and $\operatorname{int}\overline{T}_0 \leftrightarrow \operatorname{int}\overline{T}_\lambda^{\mathrm{id}}$ hold and mapping (5.15) is continuous.

D) Owing to [18, Lemma 22.1] mapping (5.9) is of class C^k. The validity of relations (5.12), (5.13) follows immediately from (5.9), (5.8) and (5.17).

It remains to prove the assertions concerning the inverse mapping (5.15). In part C we proved that $\xi_i^\lambda(x_1,x_2)$ are continuous on $\overline{T}_\lambda^{\mathrm{id}}$.

Using (3.22), (3.23) together with (5.11) and (5.12) we obtain (5.14) and the continuity of the first derivatives. The continuity of higher derivatives can be proved similarly as in [18, p. 184]. □

Theorem 5.3. *Let the boundary $\partial\Omega$ be piecewise of class C^3. Let the polynomial $w^*(\xi_1, \xi_2)$ of degree not greater than one be uniquely determined by the conditions*

$$w^*(R_i) = g_i \quad (i = 1, 2, 3).$$

Then the function $\widetilde{w} : \overline{T}_\lambda^{\mathrm{id}} \to R^1$ defined by the relations

$$\widetilde{w}(x_1, x_2) := w^*(\xi_1^\lambda(x_1, x_2), \xi_2^\lambda(x_1, x_2)), \quad (x_1, x_2) \in \overline{T}_\lambda^{\mathrm{id}},$$

where $\xi_i^\lambda(x_1, x_2)$ are the functions from (5.15), has the following properties:

a) it satisfies the relation

$$w^*(\xi_1, \xi_2) = \widetilde{w}(x_1^\lambda(\xi_1, \xi_2), x_2^\lambda(\xi_1, \xi_2)), \quad (\xi_1, \xi_2) \in \overline{T}_0$$

and is uniquely determined by the conditions

$$\widetilde{w}(P_i) = g_i \quad (i = 1, 2, 3); \tag{5.22}$$

b) $\widetilde{w} \in C^2(\overline{T}_\lambda^{\mathrm{id}})$;

c) the function values on both straight sides P_1P_j are polynomials in one variable of degree not greater than one uniquely determined by the parameters g_1 and g_j prescribed at P_1 and P_j, respectively;

d) if both parameters g_2, g_3 prescribed at P_2, $P_3 \in \lambda$ are equal to zero then $\widetilde{w}(x_1, x_2) = 0$ for all $(x_1, x_2) \in \lambda$.

The proof is the same as the proof of [18, Theorem 23.1].

Definition 5.4. The function $\widetilde{w} : \overline{T}_\lambda^{\mathrm{id}} \to R^1$ from Theorem 5.3 is called the ideal triangular finite C^0-element of the type $(L, 1)$ (where L stands for Lagrange) belonging to $\overline{T}_\lambda^{\mathrm{id}}$ and is uniquely determined by conditions (5.22). The set of all such finite elements is briefly denoted by $(\overline{T}_\lambda^{\mathrm{id}}, L, 1)$.

Theorem 5.5. *Let the boundary $\partial\Omega$ be piecewise of class C^3. Let $u \in H^2(T_\lambda^{\mathrm{id}})$, where the curved side λ of $\overline{T}_\lambda^{\mathrm{id}}$ is not approximated by the shortest side of $\overline{T}$, and let $u_I \in (\overline{T}_\lambda^{\mathrm{id}}, L, 1)$ be the ideal triangular finite C^0-element uniquely determined by the conditions*

$$u_I(P_j) = u(P_j) \quad (j = 1, 2, 3). \tag{5.23}$$

Then

$$\|u_I - u\|_{0, T_\lambda^{\mathrm{id}}} \le Ch^2 \|u\|_{0, T_\lambda^{\mathrm{id}}}, \quad |u_I - u|_{1, T_\lambda^{\mathrm{id}}} \le Ch_T^\delta \|u\|_{2, T_\lambda^{\mathrm{id}}}, \tag{5.24}$$

where C is a constant independent of h_T, $\overline{T}_\lambda^{\mathrm{id}}$ and u.

Proof. We have, according to the theorem on transformation of an integral and Theorem 5.2,

$$\|u - u_I\|^2_{0, T_\lambda^{\mathrm{id}}} \le Ch_T^{3-\delta} \|u^* - u_I^*\|^2_{0, T_0}. \tag{5.25}$$

Considering in the same way as in the proof of [18, Theorem 10.5] we obtain (cf. [18, (10.12)])

$$\|u^* - u_I^*\|_{0,T_0}^2 \le |u^*|_{2,T_0}^2. \tag{5.26}$$

Using again Theorem 5.2 and the theorem on transformation of an integral we find that

$$\left|\frac{\partial u^*}{\partial \xi_i}\right|_{1,T_0}^2 \le \frac{C}{h_T^{3-\delta}} h_T^4 \|u\|_{2,T_\lambda^{\mathrm{id}}}^2 \quad (i=1,2). \tag{5.27}$$

Combining (5.25)–(5.27) we obtain $(5.24)_1$.

Further,

$$|u_I - u|_{1,T_\lambda^{\mathrm{id}}}^2 = \iint_{T_\lambda^{\mathrm{id}}} \left\{ \left(\frac{\partial}{\partial x_1}(u_I - u)\right)^2 + \left(\frac{\partial}{\partial x_2}(u_I - u)\right)^2 \right\} \mathrm{d}x_1 \mathrm{d}x_2 \le$$

$$\le C h_T^{3-\delta} \left(h_T^{-4+2\delta} \left\| \frac{\partial}{\partial \xi_1}(u_I^* - u^*) \right\|_{0,T_0}^2 + h_T^{-2} \left\| \frac{\partial}{\partial \xi_2}(u_I^* - u^*) \right\|_{0,T_0}^2 \right). \tag{5.28}$$

Similarly as in [6]

$$\left\| \frac{\partial}{\partial \xi_i}(u_I^* - u^*) \right\|_{0,T_0}^2 \le C \left| \frac{\partial u^*}{\partial \xi_i} \right|_{1,T_0}^2 \quad (i=1,2). \tag{5.29}$$

Combining (5.28), (5.29) and (5.27) we obtain $(5.24)_2$. □

Remark 5.6. In the case of the minimum angle condition we have $\delta = 1$ and Theorem 5.5 is identical with [18, Theorem 25.3] where $n = 1$.

Remark 5.7. If the curved side λ of $\overline{T}_\lambda^{\mathrm{id}}$ is approximated by the shortest side of $\overline{T}$ then h_T^δ, which appears on the right-hand side of $(5.24)_2$, is substituted by h_T.

Definition 5.8. a) Let $\mathcal{T}_h^{\mathrm{id}}$ be the ideal triangulation of $\overline{\Omega}$ corresponding to the triangulation $\mathcal{D}_h^T$. (We obtain $\mathcal{T}_h^{\mathrm{id}}$ by replacing the triangles $\overline{T} \in \mathcal{D}_h^T$ lying along $\partial\Omega$ by corresponding ideal triangles.) The symbol M_h denotes the set of ideal triangles $\overline{T}_\lambda^{\mathrm{id}} \in \mathcal{T}_h^{\mathrm{id}}$ lying along the part of $\partial\Omega$ where the homogeneous Dirichlet condition is prescribed.

b) The function $\widehat{w} \in H^1(\Omega)$ is said to be associated with a given function $w \in X_h$ if:

(i) $\widehat{w} \in C(\overline{\Omega})$;

(ii) $\widehat{w}(P_i) = w(P_i)$ at all nodal points P_i of $\mathcal{D}_h^T$;

(iii) $\widehat{w}$ is linear on each triangle $\overline{T} \in \mathcal{D}_h^T \cap \mathcal{T}_h^{\mathrm{id}}$ and on each ideal triangle $\overline{T}_\lambda^{\mathrm{id}} \notin M_h$;

(iv) if $\overline{T}_\lambda^{\mathrm{id}} \in M_h$, then

$$\widehat{w}\,|_{\overline{T}_\lambda^{\mathrm{id}}} = \widetilde{w}\,|_{\overline{T}_\lambda^{\mathrm{id}}},$$

where $\widetilde{w}$ is defined in Definition 5.4.

Now we are prepared to estimate the fifth term appearing on the right-hand side of (5.3) in the case when $u \in H^1(\Omega)$ only.

Lemma 5.9. *For all $w \in V_h$ and $U \in H^1(\widetilde{\Omega})$ satisfying $U = u$ in Ω we have*

$$|\widetilde{L}_h(w) - \widetilde{a}_h(U,w)| \le |\widetilde{L}_h^{\Gamma}(w) - L^{\Gamma}(\overline{w})| + \\ + \sum_{\overline{T}_\lambda^{\mathrm{id}} \in M_h} \left| \iint_{\overline{T}_\lambda^{\mathrm{id}}} \left\{ (\overline{w} - \widehat{w}) f + \sum_{i=1}^{2} k_i \frac{\partial u}{\partial x_i} \frac{\partial (\widehat{w} - \overline{w})}{\partial x_i} \right\} \mathrm{d}x_1 \mathrm{d}x_2 \right| + \\ + \left| \iint_{\tau_h} \left\{ - \sum_{i=1}^{2} \widetilde{k}_i \frac{\partial U}{\partial x_i} \frac{\partial w}{\partial x_i} + w \widetilde{f} \right\} \mathrm{d}x_1 \mathrm{d}x_2 \right| + \\ + \left| \iint_{\omega_h} \left\{ \sum_{i=1}^{2} k_i \frac{\partial u}{\partial x_i} \frac{\partial \overline{w}}{\partial x_i} - \overline{w} f \right\} \mathrm{d}x_1 \mathrm{d}x_2 \right|. \quad (5.30)$$

Proof. We have

$$\widetilde{L}_h(w) = (\widetilde{L}_h^{\Omega}(w) - L^{\Omega}(\widehat{w})) + (\widetilde{L}_h^{\Gamma}(w) - L^{\Gamma}(\widehat{w})) + L(\widehat{w}),$$

where $\widehat{w} \in V$ is associated with $w \in V_h$ in the sense of Definition 5.8. It holds $a(u, \widehat{w}) = L(\widehat{w})$. Hence

$$-\widetilde{a}_h(U,w) = (a(u,\widehat{w}) - \widetilde{a}_h(U,w)) - L(\widehat{w}).$$

The rest of the proof is straightforward (see, for example, the proof of [18, Theorem 38.9]). □

Theorem 5.10. *We have*

$$|\widetilde{L}_h(w) - \widetilde{a}_h(\tilde{u}, w)| \le C h^{\delta/2} \|w\|_{1,\Omega_h} \quad \forall w \in V_h\,, \quad (5.31)$$

where the constant C does not depend on h and w and where the extension $\tilde{u}$ of u has the same meaning as in Theorem 5.1.

Proof. A) Let us denote the terms appearing on the right-hand side of (5.30) by $D_1, \ldots, D_4$. By [21, Lemmas 29, 37] and assumption (5.1) we have

$$D_1 \le Ch \|q\|_{0,\Gamma_1} \|w\|_{1,\Omega_h}. \quad (5.32)$$

Now we estimate D_2. Let B_h be the union of triangles of $\mathcal{D}_h^T$ lying along the part Γ_j of $\partial\Omega$ on which the homogeneous Dirichlet boundary condition is prescribed. Using this notation we have in the case $j = 1$, according to the Cauchy inequality,

$$D_2 \le \left(\|f\|_{0,B_h - \tau_h} + \max_{i=1,2} \|\widetilde{k}_i\|_{0,\infty,\widetilde{\Omega}} |u|_{1,B_h - \tau_h} \right) \left(\sum_{T_\lambda^{\mathrm{id}} \in M_h} \|\widehat{w} - \overline{w}\|^2_{1,T_\lambda^{\mathrm{id}}} \right)^{1/2} \quad (5.33)$$

and in the case $j = 2$

$$D_2 \le \left(\|f\|_{0,B_h \cup \omega_h} + \max_{i=1,2} \|\widetilde{k}_i\|_{0,\infty,\widetilde{\Omega}} |u|_{1,B_h \cup \omega_h} \right) \left(\sum_{T_\lambda^{\mathrm{id}} \in M_h} \|\widehat{w} - \overline{w}\|^2_{1,T_\lambda^{\mathrm{id}}} \right)^{1/2}. \quad (5.34)$$

The function $\widehat{w}\,|_{\overline{T}_\lambda^{\mathrm{id}}}$, where $\overline{T}_\lambda^{\mathrm{id}} \in M_h$, interpolates the function $\overline{w}\,|_{\overline{T}_\lambda^{\mathrm{id}}}$ on $\overline{T}_\lambda^{\mathrm{id}}$. Thus Theorem 5.5 and the linearity of $\overline{w}\,|_{\overline{T}_\lambda^{\mathrm{id}}}$ give

$$\|\widehat{w} - \overline{w}\|_{1,T_\lambda^{\mathrm{id}}} \le Ch_T^{\delta} \|\overline{w}\|_{2,T_\lambda^{\mathrm{id}}} = Ch_T^{\delta} \|\overline{w}\|_{1,T_\lambda^{\mathrm{id}}}.$$

Hence in the case $j = 2$ (i.e., in the case $u = 0$ on Γ_2)

$$\sum_{T_\lambda^{\mathrm{id}} \in M_h} \|\widehat{w} - \overline{w}\|^2_{1,T_\lambda^{\mathrm{id}}} \le Ch^{2\delta} \sum_{T_\lambda^{\mathrm{id}} \in M_h} \|\overline{w}\|^2_{1,T_\lambda^{\mathrm{id}}} \le Ch^{2\delta} \|\overline{w}\|^2_{1,\Omega} \le$$

$$\le Ch^{2\delta} \{\|\overline{w}\|^2_{1,\Omega_h} + \|\overline{w}\|^2_{1,\omega_h}\} \quad (5.35)$$

and in the case $j = 1$

$$\sum_{T_\lambda^{\mathrm{id}} \in M_h} \|\widehat{w} - \overline{w}\|^2_{1,T_\lambda^{\mathrm{id}}} \le Ch^{2\delta} \|\overline{w}\|^2_{1,\Omega_h}. \quad (5.36)$$

If $j = 2$, then relations [21, (74), (75)] and $w = 0$ on Γ_{2h} yield

$$\|\overline{w}\|_{1,\omega_h} \le Ch \sqrt{\frac{m}{\varrho}} |w|_{1,\Omega_h}.$$

Using (5.1) we obtain

$$\sqrt{\frac{m}{\varrho}} \le Ch^{\delta/2-1}. \quad (5.37)$$

Hence

$$\|\overline{w}\|^2_{1,\omega_h} \le Ch^{\delta} |w|^2_{1,\Omega_h}$$

and (5.35) implies that also in the case $j = 2$ estimate (5.36) holds. Thus for $j = 1, 2$, according to (5.33), (5.34),

$$D_2 \le Ch^{\delta} \|w\|_{1,\Omega_h}, \quad (5.38)$$

where

$$C \le \|f\|_{0,\Omega} + \max_{i=1,2} \|\widetilde{k}_i\|_{0,\infty,\widetilde{\Omega}} |u|_{1,\Omega}.$$

As to the estimate of D_3 we start from the expression, which follows from the third term on the right-hand side of (5.30) with $U = \tilde{u}$:

$$D_3 \le \max_{i=1,2} \|\widetilde{k}_i\|_{0,\infty,\widetilde{\Omega}} |\tilde{u}|_{1,\tau_h} |w|_{1,\tau_h} + \|\tilde{f}\|_{0,\tau_h} \|w\|_{0,\tau_h}. \quad (5.39)$$

Using (5.37) and considering similarly as in part B of the proof of [21, Lemma 25] we can derive

$$|w|_{1,\tau_h} \le Ch^{\delta/2} |w|_{1,\Omega_h}. \quad (5.40)$$

Further

$$\|w\|^2_{0,\tau_h} \le Ch^2 \left(\|w\|^2_{0,\Gamma_{1h}} + Ch^2|w|^2_{1,\tau_h}\right) \le C\frac{h^2}{\varrho}\|w\|^2_{1,\Omega_h}. \tag{5.41}$$

The first inequality follows from the proof of [18, Lemma 28.3] and the second from (3.31) and (5.40). Finally,

$$\|\widetilde{f}\|_{0,\tau_h} \le \|\widetilde{f}\|_{0,\infty,\widetilde{\Omega}}\sqrt{\mathrm{mes}_2\tau_h} \le Ch\|\widetilde{f}\|_{0,\infty,\widetilde{\Omega}}\sqrt{\mathrm{mes}_1\Gamma_1}. \tag{5.42}$$

Combining (5.39)–(5.42) we find that

$$D_3 \le Ch^{\delta/2}\|w\|_{1,\Omega_h}, \tag{5.43}$$

where the constant C does not depend on h and w. Similarly,

$$D_4 \le Ch^{\delta/2}\|w\|_{1,\Omega_h}. \tag{5.44}$$

Relations (5.32), (5.38), (5.43), (5.44) together with Lemma 5.9 yield estimate (5.31). □

Now we shall analyze the first term on the right-hand side of (5.3). We start with the following finite element density theorem.

Lemma 5.11. *Let $V = \{w \in H^1(\Omega) : \mathrm{tr}\, w = 0$ on $\Gamma_j\}$. For every pair $\varepsilon > 0$, $w \in V$ we can find $w_\varepsilon \in C^\infty(\overline{\Omega}) \cap V$ and $h_{\varepsilon,w} > 0$ such that for all $h \in (0, h)_{\varepsilon,w}$ we have*

$$\|\widetilde{w} - I_h w_\varepsilon\|_{1,\Omega_h} < \varepsilon \tag{5.45}$$

where $\widetilde{v} \in H^k(\widetilde{\Omega})$ is the extension of $v \in H^k(\Omega)$ according to Lemma 3.6 and $I_h v \in X_h \equiv \{w \in C(\overline{\Omega}_h) : w\big|_T \in (T, L, 1)\ \forall T \in \mathcal{T}_h\}$ is the interpolant of $v \in C(\overline{\Omega})$ defined by $(I_h v)(P_i) = v(P_i)\ \forall P_i$.

Proof. By [18, Theorem $\mathcal{P}$.92] the set $C^\infty(\overline{\Omega}) \cap V$ is dense in V. Hence, there exists a function $w_\varepsilon \in C^\infty(\overline{\Omega}) \cap V$ such that

$$\|w - w_\varepsilon\|_{1,\Omega} < \varepsilon/(2C_1) \tag{5.46}$$

where C_1 is the constant from the inequality

$$\|\widetilde{v}\|_{1,\widetilde{\Omega}} \le C_1\|v\|_{1,\Omega} \quad \forall v \in H^1(\Omega). \tag{5.47}$$

We shall consider $\widetilde{w}$ in $H^1(\widetilde{\Omega})$ and $\widetilde{w_\varepsilon}$ in $H^2(\widetilde{\Omega})$. As the extension $\widetilde{w_\varepsilon}$ is equal to the extension of w_ε from $H^1(\Omega)$ (see Lemma 3.6), we have, according to the linearity of extension operators, $\widetilde{w} - \widetilde{w_\varepsilon} = (w - w_\varepsilon)^\sim$; thus (5.46) and (5.47) yield

$$\|\widetilde{w} - \widetilde{w_\varepsilon}\|_{1,\widetilde{\Omega}} < \varepsilon/2. \tag{5.48}$$

The triangular inequality gives

$$\|\widetilde{w} - I_h w_\varepsilon\|_{1,\Omega_h} \le \|\widetilde{w} - \widetilde{w_\varepsilon}\|_{1,\Omega_h} + \|\widetilde{w_\varepsilon} - I_h w_\varepsilon\|_{1,\Omega_h}. \tag{5.49}$$

Now we estimate the terms on the right-hand side of (5.49). By (5.48) we have

$$\|\widetilde{w} - \widetilde{w_\varepsilon}\|_{1,\Omega_h} < \varepsilon/2. \tag{5.50}$$

As to the second term, we have

$$I_h w_\varepsilon = I_h \widetilde{w_\varepsilon}$$

because $\Omega \subset \widetilde{\Omega}$. This fact, the interpolation theorem for semiregular triangular linear elements (see Theorem 1.3) and the extension theorem (see Lemma 3.6) yield

$$\|\widetilde{w_\varepsilon} - I_h w_\varepsilon\|_{1,\Omega_h} \le Ch\|\widetilde{w_\varepsilon}\|_{2,\Omega_h} \le C_2 Ch\|w_\varepsilon\|_{2,\Omega}.$$

Thus there exists such an $h_{\varepsilon,w}$ that

$$\|\widetilde{w_\varepsilon} - I_h w_\varepsilon\|_{1,\Omega_h} < \varepsilon/2 \quad \forall h \in (0, h_{\varepsilon,w}). \tag{5.51}$$

Combining relations (5.49)–(5.51) we obtain (5.45). □

Theorem 5.12. *We have*

$$\lim_{h\to 0} \left\{ \inf_{v\in V_h} \|v - \widetilde{u}\|_{1,\Omega_h} \right\} = 0. \tag{5.52}$$

Proof. By Lemma 5.11, for a given $\varepsilon > 0$ we can find $u_\varepsilon \in C^\infty(\Omega) \cap V$ and $h_{\varepsilon,u} > 0$ such that

$$\|\widetilde{u} - I_h u_\varepsilon\|_{1,\Omega_h} < \varepsilon \quad \forall h \in (0, h_{\varepsilon,u}).$$

As $I_h u_\varepsilon \in V_h$ we have

$$\inf_{v\in V_h} \|v - \widetilde{u}\|_{1,\Omega_h} \le \|\widetilde{u} - I_h u_\varepsilon\|_{1,\Omega_h}.$$

Both inequalities imply (5.52). □

Theorem 5.13. *We have for all $h \in (0, h_0)$*

$$IS := \inf_{v\in V_h} \sup_{\substack{w\in V_h \\ w\neq 0}} \frac{|a_h(v,w) - \widetilde{a}_h(v,w)|}{\|w\|_{1,\Omega_h}} \le Ch(1 + \|u\|_{1,\Omega}),$$

where $u \in H^1(\Omega)$ is the solution of the continuous variational problem and the constant C does not depend on h and u.

Proof. Let $\varepsilon = 1$ and let us set

$$v = I_h u_\varepsilon \in V_h\,, \tag{5.53}$$

where, according to Lemma 5.11,

$$\|\widetilde{u} - I_h u_\varepsilon\|_{1,\Omega_h} < \varepsilon = 1 \quad \forall h \in (0, h_{\varepsilon,u}). \tag{5.54}$$

Using (5.53) and Theorem 3.13 we find

$$IS \le Ch\|I_h u_\varepsilon\|_{1,\Omega_h}. \tag{5.55}$$

Triangular inequality, extension theorem and relation (5.54) imply

$$\|I_h u_\varepsilon\|_{1,\Omega_h} \le \|\widetilde{u}\|_{1,\Omega_h} + \|\widetilde{u} - I_h u_\varepsilon\|_{1,\Omega_h} \le \|\widetilde{u}\|_{1,\widetilde{\Omega}} + 1 \le C\|u\|_{1,\Omega} + 1.$$

Combining this result with (5.55) we obtain the assertion of Theorem 5.13. □

The third and fourth terms appearing on the right-hand side of (5.3) are estimated in Theorems 3.16 and 3.18, respectively. Thus using the preceding results we obtain

Theorem 5.14. *Let us consider the set of divisions* $\{\mathcal{D}_h^T\}$ $(h \in (0, h_0))$ *introduced in Section 3. Let assumptions of Problem 3.1 and assumptions concerning the degrees of precision of quadrature formulas on a triangle and its side (see Theorems 3.13 and 3.18) be satisfied. If inequalities (5.1) hold then*

$$\lim_{h\to 0} \|\widetilde{u} - u_h\|_{1,\Omega_h} = 0$$

where u_h *is the solution of Problem 3.4 belonging to* $\mathcal{D}_h^T$, $u \in H^1(\Omega)$ *is the solution of Problem 3.1 and* $\widetilde{u} = E(u) \in H^1(\widetilde{\Omega})$ *its extension in the sense of Lemma 3.6 with* $k = 1$.

6 Appendix: Discrete Friedrichs' inequality

In [21] the inequality

$$\|v\|_{1,\Omega_h} \le C|v|_{1,\Omega_h} \quad \forall v \in V_h \quad \forall h < h_0 \tag{6.1}$$

was used without proof. As the proof differs from the proof, which was presented in [18] in the case of regular finite elements, we introduce the following lemma which is sufficient for the considerations in [21] and this paper.

Lemma 6.1. *Let* Ω *be a domain considered in Sections 3 and 5 and let (3.40) be satisfied, i.e. let*

$$C_1 h^2 \le \frac{\varrho}{m} \quad (C_1 > 0).$$

Then inequality (6.1) holds.

Proof. a) The case of the Dirichlet boundary condition (3.2). In this case

$$V_h = \{v \in X_h : \ v = 0 \text{ on } \Gamma_{1h}\}.$$

Let $\overline{v}$ be the natural extension of v and let $\widetilde{\Omega}$ be the bounded domain with boundary $\partial\widetilde{\Omega} = \Gamma_2 \cup \Gamma_3$ where Γ_3 is the circle with the centre S_0 and radius $R_3 < R_1$. We set $\overline{v} \equiv 0$ in the bounded set U_h with the boundary $\partial U_h = \Gamma_3 \cup \Gamma_{1h}$. According to the Friedrichs inequality

$$\|\overline{v}\|^2_{0,\widetilde{\Omega}} \le C|\overline{v}|^2_{1,\widetilde{\Omega}}. \tag{6.2}$$

As $\Omega_h \in \widetilde{\Omega}$ we have

$$\|v\|^2_{0,\Omega_h} \le \|\overline{v}\|^2_{0,\widetilde{\Omega}}. \tag{6.3}$$

It remains to prove

$$|\overline{v}|^2_{1,\widetilde{\Omega}} \le C|v|^2_{1,\Omega_h}. \tag{6.4}$$

We have

$$|\overline{v}|^2_{1,\widetilde{\Omega}} = |v|^2_{1,\Omega_h} + |\overline{v}|^2_{1,\omega_h}. \tag{6.5}$$

First we consider the case of the division $\mathcal{D}_h^T$. (For the definition of $\mathcal{D}_h^T$ and other types of divisions see the text following Lemma 3.3.) Let $\lambda_h \subset \Gamma_{2h}$ be the segment Q_jQ_{j+1} which approximates the arc $\lambda \subset \Gamma_2$. Similarly as in the proof of [21, Lemma 33] we can prove that

$$\operatorname{dist}(Q_j^*, \Gamma_2) \le \frac{1}{8}\frac{\varrho}{m} \equiv \frac{1}{8}b,$$

where Q_j^* is the mid-point of λ_h. Thus

$$\operatorname{mes}_2 \mathcal{P}_h \le \frac{1}{4}\operatorname{mes}_2 T,$$

where $\mathcal{P}_h$ is the bounded domain with the boundary $\partial\mathcal{P}_h = \lambda \cup \lambda_h$ and T the triangle adjacent to $\mathcal{P}_h$. As v is piecewise linear we have

$$|\overline{v}|^2_{1,\mathcal{P}_h} \le \frac{1}{4}|v|^2_{1,T}.$$

Hence

$$|\overline{v}|^2_{1,\omega_h} \le \frac{1}{4}|v|^2_{1,\Omega_h}.$$

Inserting this result into (6.5) we obtain estimate (6.4) with $C = 5/4$. The same result can be obtained in the case of the division $\mathcal{D}_h^A$.

In the case of the division $\mathcal{D}_h^K$ we use the result for $\mathcal{D}_h^A$ and estimate [21, (91)].

Combining (6.2)–(6.4) we arrive at

$$\|v\|^2_{0,\Omega_h} \le C|v|^2_{1,\Omega_h} \quad \forall v \in V_h.$$

Hence (6.1) follows.

b) The case of the Dirichlet boundary condition $v = 0$ on Γ_2. In this case

$$V_h = \{v \in X_h : \ v = 0 \text{ on } \Gamma_{2h}\}$$

and we define the quasinatural extension $\overline{v}$ of $v \in V_h$ by

$$\overline{v} = v \text{ on } \Omega_h, \quad \overline{v} = 0 \text{ on } \omega_h. \tag{6.6}$$

The Friedrichs inequality gives

$$\|\overline{v}\|^2_{0,\Omega} \le C|\overline{v}|^2_{1,\Omega}. \tag{6.7}$$

Relations (6.6) imply

$$|\overline{v}|^2_{1,\Omega} \le |v|^2_{1,\Omega_h}. \tag{6.8}$$

If we prove

$$\|\overline{v}\|^2_{0,\Omega} \ge C\|v\|^2_{0,\Omega_h} \quad (C > 0), \tag{6.9}$$

then (6.1) follows from (6.7)–(6.9).

Let us consider the case of $\mathcal{D}_h^K$. Transformation (3.20) maps one-to-one the reference square $\overline{K}_0$ with vertices $P_1^*(1,0)$, $P_2^*(0,0)$, $P_3^*(0,1)$, $P_4^*(1,1)$ onto the quadrilateral $\overline{K}$ with vertices P_1, P_2, P_3, P_4 where P_1, P_2 lie on Γ_1 and P_3P_4 is parallel to P_1P_2. Let $S_1 \in P_1P_4$, $S_2 \in P_2P_3$, let S_1S_2 be parallel to P_1P_2 and let

$$\operatorname{dist}(P_1P_2, S_1S_2) = \frac{1}{8}b.$$

Then, according to [21, Lemma 33], the arc $\lambda \subset \Gamma_1$ which is approximated by $\lambda_h = P_1P_2$ lies in Δ, where Δ denotes the quadrilateral with vertices P_1, P_2, S_2, S_1. Let us assume that we proved

$$\|v\|^2_{0,\Delta} \le \frac{3}{4}\|v\|^2_{0,K}. \tag{6.10}$$

Then

$$\|v\|^2_{0,K-\mathcal{P}_h} \ge \|v\|^2_{0,K-\Delta} = \|v\|^2_{0,K} - \|v\|^2_{0,\Delta} = \frac{1}{4}\|v\|^2_{0,K},$$

where $\mathcal{P}_h$ is the bounded domain with the boundary $\partial\mathcal{P}_h = \lambda \cup \lambda_h$. Hence (6.9) follows with $C = \frac{1}{4}$.

Let us prove (6.10). According to the definition, the function $v(x,y)$ is on every quadrilateral $\overline{K}$ such that

$$\widetilde{v}(\xi,\eta) \equiv v(x^K(\xi,\eta), y^K(\xi,\eta)) = \sum_{i=1}^{4} B_i p_i(\xi,\eta),$$

where

$$p_1 = \xi(1-\eta),\ p_2 = (\xi-1)(\eta-1),\ p_3 = (1-\xi)\eta,\ p_4 = \xi\eta$$

and $B_i = v(P_i)$ $(i = 1, \dots, 4)$. The functions $x^K(\xi,\eta),\ y^K(\xi,\eta)$ are the right-hand sides of transformation (3.20).

The quadrilateral Δ is the image of the rectangle Δ_0 with vertices P_1^*, P_2^*, S_2^*, S_1^* in transformation (3.20), where $S_1^* = [1, \frac{1}{8}]$, $S_2^* = [0, \frac{1}{8}]$. First we prove

$$\iint_{\Delta_0} [\widetilde{v}(\xi,\eta)]^2 \,\mathrm{d}\xi\mathrm{d}\eta \le \frac{1}{2} \iint_{K_0} [\widetilde{v}(\xi,\eta)]^2 \,\mathrm{d}\xi\mathrm{d}\eta. \tag{6.11}$$

Let us express the integrals

$$J_1 = \iint_{K_0} [\widetilde{v}(\xi,\eta)]^2 \,\mathrm{d}\xi\mathrm{d}\eta = \int_0^1 \left\{ \int_0^1 \left(\sum_{i=1}^4 B_i p_i(\xi,\eta) \right)^2 \mathrm{d}\eta \right\} \mathrm{d}\xi,$$

$$J_2 = \iint_{\Delta_0} [\widetilde{v}(\xi,\eta)]^2 \,\mathrm{d}\xi\mathrm{d}\eta = \int_0^1 \left\{ \int_0^{1/8} \left(\sum_{i=1}^4 B_i p_i(\xi,\eta) \right)^2 \mathrm{d}\eta \right\} \mathrm{d}\xi$$

as the quadratic forms of $B_1, \dots, B_4$. Let us denote $A = B_2$, $B = B_1$, $C = B_4$, $D = B_3$. Then

$$\begin{aligned} &4608(J_1 - 2J_2) = \\ &\quad = (174A + 87B + 117C + 134D)^2/174 + (130,5B + 175,5C)^2/130,5 + \\ &\qquad + (195,31035C + 97,655175D)^2/195,31035 + 146,48277D^2, \end{aligned}$$

from which estimate (6.11) follows.

The Jacobian J of transformation (3.20) is of the form

$$J = (h - \varepsilon^*\eta)b,$$

where, according to (3.21) and (3.40), $b = O(h^2)$, $\varepsilon^* = O(h^3)$. Thus using (6.11) and the relation

$$\iint_{K_0} [\widetilde{v}(\xi,\eta)]^2 \eta \,\mathrm{d}\xi\mathrm{d}\eta = \eta_0 \iint_{K_0} [\widetilde{v}(\xi,\eta)]^2 \,\mathrm{d}\xi\mathrm{d}\eta \quad (0 < \eta_0 < 1),$$

which is a consequence of the mean-value theorem, we obtain

$$\begin{aligned} \|v\|_{0,\Delta}^2 &= \iint_{\Delta_0} [\widetilde{v}(\xi,\eta)]^2 (h - \varepsilon^*\eta) b \,\mathrm{d}\xi\mathrm{d}\eta \le \\ &\le \iint_{\Delta_0} [\widetilde{v}(\xi,\eta)]^2 hb \,\mathrm{d}\xi\mathrm{d}\eta \le \frac{1}{2} \iint_{K_0} [\widetilde{v}(\xi,\eta)]^2 hb \,\mathrm{d}\xi\mathrm{d}\eta \le \\ &\le \frac{3}{4} \left\{ \iint_{K_0} [\widetilde{v}(\xi,\eta)]^2 hb \,\mathrm{d}\xi\mathrm{d}\eta - \iint_{K_0} [\widetilde{v}(\xi,\eta)]^2 \varepsilon^* \eta_0 b \,\mathrm{d}\xi\mathrm{d}\eta \right\} = \\ &= \frac{3}{4} \iint_{K_0} [\widetilde{v}(\xi,\eta)]^2 (h - \varepsilon^*\eta) b \,\mathrm{d}\xi\mathrm{d}\eta = \frac{3}{4} \|v\|_{0,K}^2, \end{aligned}$$

which proves (6.10).

In the case of division $\mathcal{D}_h^T$ the proof of (6.9) is similar but simpler: Let T be a triangle with vertices P_1, P_2 lying on Γ_1 and let Q_1 and Q_2 be the mid-points of the sides P_1P_3 and P_2P_3, respectively. Let T^* denote the triangle with vertices Q_1, Q_2, P_3. Then

$$\|v\|^2_{0,T-\mathcal{P}_h} \geq \|v\|^2_{0,T^*}$$

and it is relatively easy to compute that

$$\|v\|^2_{0,T^*} \geq \frac{1}{64}\|v\|^2_{0,T} \ .$$

The last two inequalities imply (6.9) with $C = 1/64$. □

Acknowledgement. The work was supported by the grant No. 201/97/0153 of the Grant Agency of the Czech Republic. This support is gratefully acknowledged.

References

1. T. Apel, and M. Dobrowolski, Anisotropic interpolation with applications to the finite element method, *Computing*, **47** (1992), 277–293
2. I. Babuška and A. K. Aziz, On the angle condition in the finite element method, *SIAM J. Numer. Anal.*, **13** (1976), 214–226
3. P. G. Ciarlet *The Finite Element Method for Elliptic Problems*, North-Holland, Amsterdam 1978
4. M. Feistauer and A. Ženíšek, Finite element solution of nonlinear elliptic problems, *Numer. Math.*, **50** (1987), 451–475
5. P. Jamet, Estimations d'erreur pour des éléments finis presque dégénérés, *RAIRO Anal. Numér.*, **10** (1976), 43–61
6. M. Křížek, On semiregular families of triangulations and linear interpolation, *Appl. Math.*, **36** (1991), 223–232
7. S. Koukal, Piecewise polynomial interpolations and their applications in partial differential equations, *Sborník VAAZ Brno* (1970), 29–30 (in Czech)
8. S. Koukal, Piecewise polynomial interpolations in the finite element method, *Apl. Mat.*, **18** (1973), 146–160
9. A. Kufner, O. John and S. Fučík, *Function Spaces*, Academia, Prague 1977
10. A. Marrocco, Analyse numérique de problémes d'électrotechnique, *Ann. Sci. Math. Québec*, **1** (1977) 271–296
11. F. Melkes and A. Ženíšek, On a certain two-sided symmetric condition in magnetic field analysis and computations, *Appl. Math.* (1997), 147–159
12. J. Nečas, *Les Méthodes Directes en Théorie des Equations Elliptiques*, Academia/Masson, Prague/Paris 1967
13. L. A. Oganesian and L. A. Rukhovec, *Variational-Difference Methods for the Solution of Elliptic Problems*, Izd. Akad. Nauk ArSSR, Jerevan 1979 (in Russian)
14. J. L. Synge, *The Hypercircle in Mathematical Physics*, Cambridge Univ. Press, London 1957
15. A. Ženíšek, The convergence of the finite element method for boundary value problems of a system of elliptic equations, *Apl. Mat.*, **14** (1969), 355–377 (in Czech)

16. A. Ženíšek, Interpolation polynomials on the triangle, *Numer. Math.*, **15** (1970), 283–296
17. A. Ženíšek, The finite element method for nonlinear elliptic equations with discontinuous coefficients, *Numer. Math.*, **58** (1990), 51–57
18. A. Ženíšek, *Nonlinear Elliptic and Evolution Problems and Their Finite Element Approximations*, Academic Press, London 1990
19. A. Ženíšek, Maximum-angle condition and triangular finite elements of Hermite type, *Math. Comp.*, **64** (1995), 929–941
20. A. Ženíšek, The maximum angle condition in the finite element method for monotone problems with applications in magnetostatics, *Numer. Math.*, **71** (1995), 399–417
21. A. Ženíšek, Finite element variational crimes in the case of semiregular elements, *Appl. Math.*, **41** (1996), 367–398
22. A. Ženíšek and M. Vanmaele, The interpolation theorem for narrow quadrilateral isoparametric finite elements, *Numer. Math.*, **72** (1995), 123–141
23. A. Ženíšek and M. Vanmaele, Applicability of the Bramble-Hilbert lemma in interpolation problems of narrow quadrilateral isoparametric finite elements, *J. Comp. Appl. Math.*, **63** (1995), 109–122
24. M. Zlámal, On the finite element method, *Numer. Math.*, **12** (1968), 394–409
25. M. Zlámal, Curved elements in the finite element method I., *SIAM J. Numer. Anal.*, **10** (1973), 229–240

Author Index

Subject Index